The Adversarial Joint Source-Channel Problem

Yuval Kochman, Arya Mazumdar and Yury Polyanskiy

Abstract—This paper introduces the problem of joint source-channel coding in the setup where channel errors are adversarial and the distortion is worst case. Unlike the situation in the case of stochastic source-channel model, the separation principle does not hold in adversarial setup. This surprising observation demonstrates that designing good distortion-correcting codes cannot be done by serially concatenating good covering codes with good error-correcting codes. The problem of the joint code design is addressed and some initial results are offered.

I. INTRODUCTION

One of the great contributions of Shannon [1] was creation of tractable and highly descriptive stochastic models for the signal sources and communication systems. Shortly after, his work was followed up by Hamming [2], who proposed a combinatorial variation of the channel coding part. This combinatorial formulation has become universally accepted in the coding-theoretic community. Similarly, for the case of lossy compression Shannon [3] proposed a stochastic model and the rate-distortion formula, while shortly after Kolmogorov followed up with a non-stochastic definition of the ϵ-entropy [4]. The research that followed demonstrated how both ways of thinking, stochastic and combinatorial, naturally complement each other, reinforcing intuition and yielding new results.

To the best of our knowledge, in the setup of joint source-channel coding, however, only the stochastic approach has been investigated so far, starting with [1], [3]. This paper aims to fill in this omission.

In Section II we define the adversarial separate source and channel coding problems and present known results about them. Then, we build on these definitions to define the adversarial *joint source channel coding* (JSCC) problem. Next, in Section III we prove asymptotic bounds on the performance limits of adversarial JSSC codes. It turns out that the celebrated separation principle [1], [3] does not hold in the adversarial model. Therefore, the problem of constructing asymptotically optimal adversarial JSSC codes requires a joint approach and cannot be solved by combining good compressors with good error-correcting codes. In Section IV we focus on the binary case and propose methods for designing such codes and analyzing their performance.

II. PRELIMINARIES

A. Source coding

A source problem is specified by a source and reproduction alphabets $\mathcal{S}$, $\hat{\mathcal{S}}$, a distribution P on $\mathcal{S}$ and a distortion metric

Yuval Kochman is with the School of Computer Science and Engineering, the Hebrew University of Jerusalem, Jerusalem, Israel. e-mail: `yuvalko@cs.huji.ac.il`. Arya Mazumdar and Yury Polyanskiy are with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139, USA. e-mail: {`aryam`,`yp`}`@mit.edu`.

$d : \mathcal{S} \times \hat{\mathcal{S}} \to \mathbb{R}_+$. The distortion between a source string s^k and a reproduction $\hat{s}^k$ is given by:

$$d(s^k, \hat{s}^k) \triangleq \frac{1}{k} \sum_{j=1}^{k} d(s_j, \hat{s}_j). \qquad (1)$$

In the stochastic setting, an (k, M_k, D)-source code is specified by a surjective map $\phi : \mathcal{S}^k \to \mathcal{C}$ for some $\mathcal{C} \subseteq \hat{\mathcal{S}}^k$ such that $|\mathcal{C}| = M_k$ and the expected distortion is at most D, where the mean is taken with $S^k \sim P^k$ (memoryless source). The *rate* of the source code is defined by $1/k \cdot \log M_k$ and asymptotically, the best possible rate for the distortion D is given by [3]:

$$R(P, D) \triangleq \min_{P_{\hat{S}|S} : \mathbb{E}[d(S, \hat{S})] \leq D} I(S; \hat{S}).$$

In the adversarial setting, a *source set* $\mathcal{F} \subseteq \mathcal{S}^k$ is selected and then the smallest cardinality of a covering of $\mathcal{F}$ upto distortion D is sought; cf. [4]. Here we restrict ourselves to the case of $\mathcal{F}$ being the set of all source sequences that are strongly typical[1] with respect to the source distribution P.

The adversarial (k, M_k, D) source code is defined by a collection of M_k points $\mathcal{C} \subset \hat{\mathcal{S}}^k$ such that for any P-typical source sequence s^k there exists a point $\hat{s}^k$ in $\mathcal{C}$ such that $d(s^k, \hat{s}^k) \leq D$. The asymptotic fundamental limit of adversarial source coding is defined to be

$$R_{ad}(P, D) \triangleq \lim_{k \to \infty} \frac{1}{k} \log \max\{M_k : \exists (k, M_k, D) \\ \text{-adversarial source code}\}.$$

Not only does this limit exist, but remarkably it coincides with $R(P, D)$:

Theorem 1 (Berger's type covering [6]):

$$R_{ad}(P, D) = R(P, D).$$

As an example, take $\mathcal{S} = \hat{\mathcal{S}} = \mathbb{F}_2$ and P is the uniform distribution, with the Hamming distortion measure. It is known that

$$R_{ad}(P, D) = R(P, D) = 1 - h_2(D),$$

where $h_2(x) = -x \log x - (1 - x) \log(1 - x)$ is the binary entropy function. Indeed the same rate is achievable even if the source set is entire $\mathbb{F}_2^k$ [7].

---

[1]Here and in the sequel, strong typicality is in the sense of [5, Chapter 2].

B. Channel coding

A channel problem is specified by input[2] and output alphabets $\mathcal{X}$, $\mathcal{Y}$, and a conditional distribution $W : \mathcal{X} \to \mathcal{Y}$.

In the stochastic setting, an (n, M, ϵ)-channel code is specified by a pair of maps $f : \{1, \ldots, M\} \to \mathcal{X}^n$ and $g : \mathcal{Y}^n \to \{1, \ldots, M\}$ such that

$$\mathbb{P}[g(Y^n) = i | X^n = f(i)] \geq 1 - \epsilon, \qquad i = 1, \ldots, M,$$

where $P_{Y^n | X^n} = W^n$ (a memoryless channel). The *rate* of the code is defined as $1/n \cdot \log M_n$ and asymptotically the largest achievable rate is given by Shannon capacity [1]:

$$C(W) = \max_{P_X} I(X; Y).$$

In the adversarial setting, for each input sequence $x^n \in \mathcal{X}^n$ the channel output may be arbitrary within a subset of $\mathcal{Y}^n$. We choose this set to be $\mathcal{A}(x^n) \subseteq \mathcal{S}^n$, the set of strongly typical sequences y^n given x^n with respect to W. The adversarial (n, M_n) channel code is defined as a collection of M_n points $\mathcal{C} \subset \mathcal{X}^n$ such that for any pair of different points $x^n, z^n \in \mathcal{C}$, $\mathcal{A}(x^n) \cap \mathcal{A}(z^n) = \emptyset$. The asymptotic fundamental limits of adversarial channel coding are defined to be

$$\overline{C}_{ad}(W) \triangleq \limsup_{n \to \infty} \frac{1}{n} \log \max\{M_n : \exists (n, M_n)$$
$$\text{-adversarial channel code}\} \tag{2}$$

$$\underline{C}_{ad}(W) \triangleq \liminf_{n \to \infty} \frac{1}{n} \log \max\{M_n : \exists (n, M_n)$$
$$\text{-adversarial channel code}\}. \tag{3}$$

Note that because the limits are not known to coincide for most channels of interest, we have to define both upper and lower limits.

It is known that $\overline{C}_{ad}(W) \leq C(W)$. Furthermore, in contrast to source coding, this inequality is known to be strict in the next example.

The most studied case of the adversarial channel coding is that of a binary symmetric channel with crossover probability δ, BSC(δ). Let $A(n, d)$ be the cardinality of a largest set in $\mathbb{F}_2^n$ with minimal Hamming distance between any pair of elements not smaller than d. We have[3]:

$$\overline{C}_{ad} = \limsup_{n \to \infty} \frac{1}{n} \log A(n, 2n\delta + 1),$$

and similarly for $\underline{C}_{ad}$. Therefore, by the classical results on $A(n, d)$:

$$R_{GV}(\delta) \leq \underline{C}_{ad}(\delta) \leq \overline{C}_{ad}(\delta) \leq R_{MRRW}(\delta) < C(\delta), \tag{4}$$

where the MRRW II bound [8] is

$$R_{MRRW}(\delta) = \min_{0 < \alpha \leq 1 - 4\delta} 1 + \hat{h}(\alpha^2) - \hat{h}(\alpha^2 + 4\delta\alpha + 4\delta), \tag{5}$$

[2]A cost function on $\mathcal{X}$ may also be present. We omit it to save space.

[3]Traditionally, one considers the case when the adversary is free to choose noise vectors e satisfying $\text{wt}(e) \leq \delta n$, whereas in our setting the typicality constrains $\text{wt}(e) = \delta n \pm o(n)$. This is asymptotically immaterial, since in Hamming space two spheres of the same radius are disjoint if and only if the corresponding balls are.

with $\hat{h}(x) = h_2(1/2 - 1/2\sqrt{1-x})$, and the Gilbert-Varshamov bound [9] is

$$R_{GV}(\delta) = 1 - h_2(2\delta). \tag{6}$$

C. Adversarial JSCC

A JSCC problem is specified by:
- Adversarial source: $\mathcal{S}$, $\hat{\mathcal{S}}$, P_S, $d(\cdot, \cdot)$
- Adversarial channel: $\mathcal{X}$, $\mathcal{Y}$, $W_{Y|X}$.

At source and channel blocklengths (k, n), a JSCC scheme is specified by:
- an encoder map $\mathcal{S}^k \to \mathcal{X}^n$ from the source to channel input: $x^n = f(s^k)$.
- a decoder map $\mathcal{Y}^n \to \hat{\mathcal{S}}^k$ from the channel output to reconstruction: $\hat{s}^k = g(y^n)$.

We say that a JSSC scheme is (k, n, D) adversarial if for all P-typical source sequence s^k and corresponding channel outputs $y^n \in \mathcal{A}(f(s^k))$, $d(s^k, g(y^n)) \leq D$.

The asymptotically optimal tradeoff between the achievable distortion and the bandwidth expansion factor $\rho = \frac{n}{k}$ is given by

$$\overline{D}^*_{ad}(\rho) = \limsup_{k \to \infty} \inf\{D : \exists (k, \lfloor \rho k \rfloor, D)$$
$$\text{adversarial JSCC}\}, \tag{7}$$

$$\underline{D}^*_{ad}(\rho) = \liminf_{k \to \infty} \inf\{D : \exists (k, \lfloor \rho k \rfloor, D)$$
$$\text{adversarial JSCC}\}. \tag{8}$$

As in the source and channel cases, we use the stochastic setting performance as a benchmark. In this setting, the source and channel are i.i.d. according to $P = P_S$ and $W = W_{Y|X}$, and the requirement is for expected distortion to be at most D. It is well known that any k-to-n stochastic JSCC must satisfy [3],

$$k \cdot R(P, D) \leq n \cdot C(W). \tag{9}$$

In the asymptotic limit this can be approached, yielding the asymptotic fundamental limit:

$$D^*(\rho) = \inf\{D : R(P, D) \leq \rho C(W)\}.$$

D. The separation principle

We say that an (k, n) JSCC scheme is separation-based if for some space $\mathcal{M}$ ("the message space") the encoder consists of a source encoder $f_S : \mathcal{S}^k \to \mathcal{M}$ and a channel encoder $f_C : \mathcal{M} \to \mathcal{X}^n$. The decoder consists of a channel decoder $g_C : \mathcal{Y}^n \to \mathcal{M}$ and a source decoder $\mathcal{M} \to \hat{\mathcal{S}}^k$. Furthermore, following e.g. [10] we introduce a bijection $\sigma : \mathcal{M} \to \mathcal{M}$ that is applied at the encoder and reversed in the decoder, which is meant to ensure that there the mapping of source messages to channel ones is arbitrary. The encoder and decoder are thus given by

$$f = f_S \circ \sigma \circ f_C \, ; g = g_C \circ \sigma^{-1} \circ g_S \tag{10}$$

where performance is required to hold for *any* bijection σ.

The asymptotic performance limits of the separation schemes are denoted as $\overline{D}^*_{ad, sep}(\rho)$ and $\underline{D}^*_{ad, sep}(\rho)$ and defined

in complete analogy with (7) and (8). In the stochastic setting, the asymptotic performance of the optimal separation scheme coincides with $D^*(\rho)$ and thus does not need a special notation.

III. BOUNDS ON ADVERSARIAL JSCC

We start this section with an immediate lower bound on the fundamental limit of adversarial asymptotic distortion.

Theorem 2 (Converse):

$$\underline{D}_{ad}^*(\rho) \geq D^*(\rho).$$

Proof: Any adversarial JSCC can be used as a usual (probabilistic) JSCC, in which case by typicality arguments it will achieve (maximal) distortion D with vanishing excess probability (namely, we assume excess distortion whenever the source or channel behavior are not strongly typical). Thus D must not be smaller than $D^*(\rho)$. ∎

A. Separated schemes

Theorem 3 (Separated schemes): If $R(P, D) > \rho\overline{C}_{ad}(W)$ then

$$\underline{D}_{ad,sep}^*(\rho) \geq D. \tag{11}$$

If $R(P, D) \leq \rho\underline{C}_{ad}(W)$ then

$$\overline{D}_{ad,sep}^*(\rho) \leq D. \tag{12}$$

We will show shortly, that (11) demonstrates (in special cases) that $D_{ad,sep}^* > D_{ad}^*$.

B. Single-letter schemes

Another special class of JSCC schemes is single-letter codes. In that case, the mappings $f(\cdot)$ and $g(\cdot)$ are scalar, and when applied to a block they are computed in parallel for each entry. Some examples where single-letter schemes yield the optimum D^* have been known for a long time, and Gastpar et al. [11] give the sufficient and necessary conditions for that to hold.

Theorem 4: If in the stochastic setting a single-letter scheme achieves some D_{sl}, then

$$\overline{D}_{ad}^*(1) \leq D_{sl}. \tag{13}$$

We omit a simple proof of this result, but its essence will be clear from the example in the next section.

Corollary 5: Whenever single-letter codes are optimal in the stochastic setting, i.e., $D_{sl} = D^*(1)$ we have

$$\overline{D}_{ad}^*(1) = \underline{D}_{ad}^*(1) = D^*(1).$$

Using Theorems 3 and 4, one may find examples in which single-letter schemes achieve D^* while separation-based scheme do not, leading to the surprising conclusion that separation is not optimal in the adversarial setting.

C. Binary example

We now combine the binary examples presented in sections II-A and II-B: the source is binary symmetric with Hamming distortion, and the channel is BSC(δ). The information-theoretic optimum $D^*(\rho)$ is given by the solution D to:

$$1 - h_2(D) = \rho(1 - h_2(\delta))$$

whenever the r.h.s. is lower than one, zero otherwise. Bounds on the performance of separation-based schemes are given by the solutions to:

$$1 - h_2(D) = \rho \cdot R_{MRRW}(\delta)$$
$$1 - h_2(D) = \rho \cdot R_{GV}(\delta),$$

where again the bounds are zero for r.h.s. above one. Since $R_{MRRW} < 1 - h_2(\delta)$ for all $\delta > 0$, it follows that $\underline{D}_{ad,sep}^*(\rho) > D^*(\rho)$ strictly whenever $\rho R_{MRRW}(\delta) < 1$.

For $\rho = 1$ the optimum $D^*(1)$ is achievable by a trivial single-letter scheme (namely, the identity encoder and decoder). Therefore, for $\rho = 1$ and any $\delta > 0$,

$$D^*(1) = D_{ad}^*(1) < \underline{D}_{ad,sep}^*(\rho).$$

For other values of ρ, separation may also be suboptimal:

Proposition 6: For any positive integer ρ, repetition coding (i.e., x^n is constructed by ρ repetitions of s^k) achieves asymptotically:

$$D_{rep}(\rho) = \frac{2\rho\delta}{1 + \rho} \tag{14}$$

By (4) and Theorem 3, it is easy to see that $\overline{D}_{ad,sep}^*(\rho) = \underline{D}_{ad,sep}^*(\rho) = 1/2$ whenever $\delta = 1/4$. Thus, comparing with (14) and by continuity for any positive integer ρ there is an interval of δ for which simple repetition coding outperforms any separation-based scheme.

IV. BINARY SYMMETRIC SOURCE-CHANNEL (BSSC)

In this section we slightly change the problem definition, in order to make it closer in spirit to that of traditional approach taken in the coding-theoretic literature for the Hamming space. Namely, we drop the strong typicality constraints on the source and the channel. Instead, we let the source outputs be any binary sequences in $\mathbb{F}_2^k$, while the (adversarial) channel is allowed to flip up to δn bits.

Definition 1: A (k, n, D) adversarial JSSC code for the $BSSC(\delta)$ is a pair of maps $f : \mathbb{F}_2^k \to \mathbb{F}_2^n$, $g : \mathbb{F}_2^n \to \mathbb{F}_2^k$ such that

$$\text{wt}(x + g(f(x) + e)) \leq kD,$$

for all $x \in \mathbb{F}_2^k$ and all $\text{wt}(e) \leq \delta n$. The asymptotic fundamental limits $\overline{D}_{ad}^*(\rho)$ and $\underline{D}_{ad}^*(\rho)$ are defined as in (7)-(8).

Note that while in channel coding the two definitions lead to similar results (recall Footnote 3), it is not clear whether the same holds for JSCC. For example, in Proposition 6, for even ρ the decoding relies on the fact that the adversary *must* flip approximately δn bits, and if this assumption does not hold, repetition with even expansion ρ is equivalent to repetition with expansion $\rho - 1$ followed by channel uses that can be ignored.

A. Information theoretic converse

Note that by Theorem 2, we have that any asymptotically achievable distortion D over $BSSC(\delta)$ satisfies

$$1 - h_2(D) \leq \rho(1 - h_2(\delta)). \tag{15}$$

In fact, if there exists a JSCC that achieves distortion D, then any ball of radius δn in $\mathbb{F}_2^n$ must not contain more than T_{Dk}^k codewords, where T_r^m is the volume of a ball of radius r in $\mathbb{F}_2^m$. However there exists a ball of radius δn in $\mathbb{F}_2^n$ that contains at least $2^{k-n}T_{\delta n}^n$ codewords. Hence D must satisfy

$$2^k T_{\delta n}^n \leq 2^n T_{Dk}^k. \tag{16}$$

Asymptotically (16) coincides with (15), but otherwise is tighter.

B. New coding converse

The above lower bound on achievable distortion D can be improved for a region of δ if we consider the fact that any JSCC also gives rise to an error-correcting code. Recalling the cardinality $A(n, \delta)$ defined in Section II-B, we have the following.

Theorem 7: If a k-to-n JSCC achieves the distortion D over $BSSC(\delta)$, then

$$A(k, 2Dk + 1) \leq A(n, 2n\delta + 1).$$

Proof: Suppose there is a code $\mathcal{D} \subset \mathbb{F}_2^k$ that corrects up to any Dk errors. Let $\hat{\mathcal{D}}$ be the image of this code in $\mathbb{F}_2^n$ under the JSCC encoding. We claim that $\hat{\mathcal{D}}$ is a code in $\mathbb{F}_2^n$ that corrects any up to δn errors. Indeed, up to δn errors can be reduced to at most Dk errors in $\mathbb{F}_2^k$ with the JSCC decoding. These errors are then correctable with the decoding of $\mathcal{D}$. ∎

Asymptotically, applying (4) to Theorem 7 we obtain:

Corollary 8: For the $BSSC(\delta)$ the distortion $\underline{D}_{ad}^*(\rho)$ satisfies:

$$R_{GV}(\underline{D}_{ad}^*(\rho)) \leq \rho R_{MRRW}(\delta). \tag{17}$$

C. Achievability and converse for separation scheme

As explained in Footnote 3, the limits for channel coding are the same for strongly typical channel and for maximum number of flips. Thus, by Theorem 3, the asymptotic performance of the separation schemes must satisfy

$$\rho R_{GV}(\delta) \leq 1 - h_2(D_{ad,sep}^*(\rho)) \leq \rho R_{MRRW}(\delta). \tag{18}$$

Remark: Note that, although the exact value of $\overline{C}_{ad}$ or $\underline{C}_{ad}$ is unknown, the argument in Theorem 7 demonstrates that in the regime of distortion $D \to 0$, separation yields an optimal (but unknown) performance.

Just as in Section III-C it is clear that in the case $\rho = 1$ separation is strictly suboptimal for all $\delta > 0$. Comparison of the different bounds for this case is shown in Fig. 1. Next, we show examples of codes that beat separation for other $\rho \neq 1$.

D. The optimal decoder for BSSC

Let $\mathcal{B}_n(x, r)$ denote a ball of radius r centered at x in $\mathbb{F}_2^n$. For any set $\mathcal{S} \in \mathbb{F}_2^n$, the radius of the set $\mathrm{rad}(\mathcal{S})$ is defined to be the smallest r such that $\mathcal{S} \subseteq \mathcal{B}_n(x, r)$ for some $x \in \mathbb{F}_2^n$, with the optimal x's called the *Chebyshev center(s)* of $\mathcal{S}$.

Consider some JSSC encoder $f : \mathbb{F}_2^k \to \mathbb{F}_2^n$ for the $BSSC(\delta)$. There exists a decoder achieving distortion D for this if and only if

$$\forall y \in \mathbb{F}_2^n : \mathrm{rad}(f^{-1}\mathcal{B}_n(y, \delta n)) \leq Dk.$$

The optimal decoder is then:

$$g(y) = \text{Chebyshev center of } f^{-1}\mathcal{B}_n(y, \delta n). \tag{19}$$

In other words, the distortion achievable by the encoder f is given by

$$D(f, \delta) = \frac{1}{k} \max_{y \in \mathbb{F}_2^n} \mathrm{rad}(f^{-1}\mathcal{B}_n(y, \delta n)).$$

E. Repetition of a small code

In contrast to channel coding, repetition of a single code of small block length leads to a non-trivial asymptotic performance.

Fix an arbitrary encoder given by the mapping $f : \mathbb{F}_2^u \to \mathbb{F}_2^v$. If there are t errors in the block of length v, $t = 0, \ldots, v$ the performance of the optimal decoder (knowing t) is given by

$$r_0(t) = \max_{y \in \mathbb{F}_2^v} \mathrm{rad}(f^{-1}\mathcal{B}_v(y, t)). \tag{20}$$

Consider also an arbitrary decoder $g : \mathbb{F}_2^v \to \mathbb{F}_2^u$ and its performance curve:

$$r_g(t) = \max_{\mathrm{wt}(e) \leq t} \max_{x \in \mathbb{F}_2^u} \mathrm{wt}(g(f(x) + e) + x).$$

Clearly

$$r_g(t) \geq r_0(t)$$

and the decoder g achieving this bound with equality is called a *universal decoder*. Some trivial properties: $r_0(0) = 0$ if and only if f is injective, $r_g(0) = 0$ if and only if g is a left inverse of f, $r_0(v) = r_g(v) = u$.

Example: Any repetition code $\mathbb{F}_2 \to \mathbb{F}_2^v$ is universally decodable with a majority-vote decoder g (resolving ties arbitrarily):

$$r_g(t) = r_0(t) = \begin{cases} 0, & t < \frac{v}{2}, \\ 1, & t \geq \frac{v}{2}. \end{cases}$$

From a given code f we may construct a longer code by repetition to obtain an $\mathbb{F}_2^k \to \mathbb{F}_2^n$ code as follows, where $Lu = k, Lv = n$:

$$f_L(x_1, \ldots, x_L) = (f(x_1), \ldots, f(x_L)).$$

This yields a sequence of codes with $\rho = n/k = v/u$. We want to find out the achieved distortion $D(\delta)$ as a function of the maximum crossover portion δ of the adversarial channel.
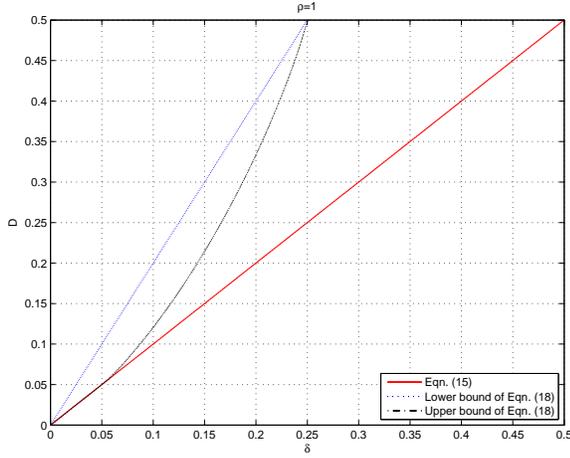
Fig. 1. Trade-off between δ and D in a BSSC(δ) for $\rho = 1$. An identity map (single-letter scheme) is everywhere optimal.
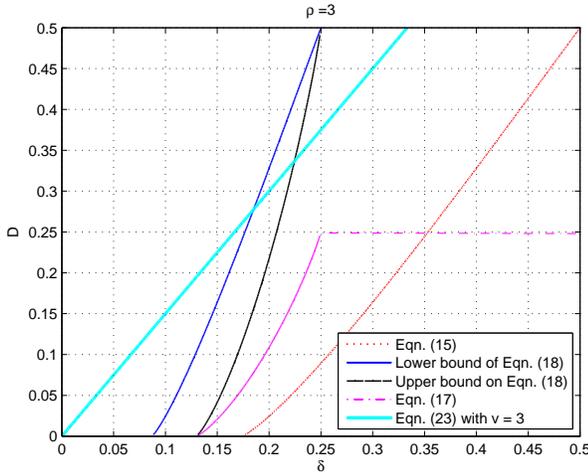

Fig. 2. Trade-off between δ and D in a BSSC(δ) for $\rho = 3$. 3-repetition code achieves better distortion than any separation scheme for $\delta > 0.22$

Theorem 9: The asymptotic distortion achievable by f_L (repetition construction) satisfies

$$\liminf_{L \to \infty} D(f_L, \delta) \geq \frac{1}{u} r_0^{**}(\delta v). \quad (21)$$

A block-by-block decoder g achieves

$$\liminf_{L \to \infty} D_g(f_L, \delta) = \frac{1}{u} r_g^{**}(\delta v), \quad (22)$$

where r_0^{**} and r_g^{**} are upper convex envelopes of r_0 and r_g respectively.

Example: Repetition code: Consider using a $[v, 1, v]$ repetition code. Since for such a code $r_g(t) = r_0(t)$, the upper and lower bounds of Theorem 9 coincide. For odd v we have:

$$D = \frac{2\delta v}{v + 1}. \quad (23)$$

(Compare this with Proposition 6 for the strong-typicality model of Section II-C.) In Fig. 2 the performance of the

3-repetition code is contrasted with that of the separation schemes. In the same plot the converse bounds (17) and (15) are plotted. For $\delta > 0.23$ it is clear that 3-repetition achieves better performance than any separation scheme.

Example: [5,2,3] linear code for $\rho = 5/2$: Consider the linear map $f : \mathbb{F}_2^2 \to \mathbb{F}_2^5$ given by the generator matrix

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

It can be shown that $r_0(t) = \{0, 0, 1, 2, 2, 2\}$ for $t = \{0, 1, 2, 3, 4, 5\}$ and there exists a universal decoder g. Thus by Theorem 9 this code achieves $D = 5\delta/3$. For $\delta > 0.22$, this is better than what any separation scheme can achieve. This example demonstrates that in the JSSC setup one should not always use a simple decoder that maps to the closest codeword. In fact, further analysis demonstrates that perfect codes, Golay and Hamming, are among the worst in terms of distortion tradeoff.

Remark: Note that there exist [12] linear codes of rate ρ^{-1} decodable with finite list size and capable of correcting all errors up to the information theoretic limit $n \, h_2^{-1}(1 - \rho^{-1})$. However, by the converse bound (17) it follows that the radius of the list in $\mathbb{F}_2^k$ must be $\Omega(k)$ regardless of the map between $\mathbb{F}_2^k$ and the codewords. This provides some interesting complement to the study of the properties of lists of codes achieving the information theoretic limit [13], [14].

REFERENCES

[1] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423 and 623–656, July/October 1948.
[2] R.W. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal*, 29(2):147–160, 1950.
[3] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. In *1959 IRE Nat. Conv. Rec.*, volume pt. 4, pages 142–163, 1959.
[4] A. N. Kolmogorov and V. M. Tikhomirov. Theory of transmission of information. *Amer. Math. Soc. Trans.*, 17:277–364, 1961.
[5] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic, New York, 1981.
[6] T. Berger. *Rate-Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, Englewood Cliffs, NJ, 1971.
[7] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein. *Covering Codes*. North-Holland, 1997.
[8] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch. New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE Trans. Inf. Theory*, 23(2):157–166, 1977.
[9] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, 1997.
[10] D. Wang, A. Ingber, and Y. Kochman. The dispersion of joint source-channel coding. In *Proc. 2011 49th Allerton Conference*, volume 49, Allerton Retreat Center, Monticello, IL, USA, September 2011.
[11] M. Gastpar, B. Rimoldi, and M. Vetterli. To code, or not to code: Lossy source-channel communication revisited. *IEEE Trans. Inf. Theory*, 49(5):1147–1158, May 2003.
[12] P. Elias. Error-correcting codes for list decoding. *IEEE Trans. Inf. Theory*, 37(1):5–12, 1991.
[13] V. M. Blinovsky. Bounds for codes in the case of list decoding of finite volume. *Prob. Peredachi Inform.*, 22(1):7–19, 1986.
[14] V. Guruswami and S. Vadhan. A lower bound on list size for list decoding. *IEEE Trans. Inf. Theory*, 56(11):5681–5688, November 2010.