Results on combinatorial joint source-channel coding

Results on Combinatorial Joint Source-Channel Coding

Yuval Kochman, Arya Mazumdar and Yury Polyanskiy

***Abstract*—This paper continues the investigation of the combinatorial formulation of the joint source-channel coding problem. In particular, the connections are drawn to error-reducing codes, isometric embeddings and list-decodable codes. The optimal performance for the repetition construction is derived and is shown to be achievable by low complexity Markov decoders. The compound variation of the problem is proposed and some initial results are put forward.**

I. Introduction

Recently, a combinatorial model of the problem of joint-source channel coding (JSCC) was proposed in [8]. The gist of it for the binary source and symmetric channel (BSSC) can be summarized by the following

Definition 1: A pair of maps $f : \mathbb{F}_2^k \to \mathbb{F}_2^n$ and $g : \mathbb{F}_2^n \to \mathbb{F}_2^k$ is called a (k, n, D, δ) JSCC if

$$|x + g(f(x) + e)| \le kD\,,$$

for all $x \in \mathbb{F}_2^k$ and all $|e| \le \delta n$, where $|\cdot|$ and $d_H(x, y) = |x + y|$ denote the Hamming weight and distance. The asymptotic fundamental limit is:[1]

$$D^*_{ad}(\rho, \delta) = \lim_{k\to\infty} \inf\{D : \exists(k, \lfloor \rho k\rfloor, D, \delta)\text{-JSCC}\}\,. \quad (1)$$

We briefly overview some results from [8]. The performance of any $(k, \lfloor \rho k\rfloor, D, \delta)$ scheme can be bounded by the *information-theoretic converse*:

$$1 - \mathrm{h}_2(D) \le \rho(1 - \mathrm{h}_2(\delta)), \quad (2)$$

where $\mathrm{h}_2(x) = -x \log x - (1-x) \log(1-x)$ is binary entropy. Consequently, $D^*_{ad}(1, \delta) \le \delta$. On the other hand for $n = k$, the simple scheme defined by identity maps yields $D = \delta$. We conclude that $D^*_{ad}(1, \delta) = \delta$.

We now compare this to any separation-based scheme, comprising of source quantization and an error-correcting code. A distortion of:

$$1 - \mathrm{h}_2(D) = \rho R \quad (3)$$

can be achieved if and only if there exists an error correcting code (ECC) of rate R for portion of channel flips δ. By the Plotkin bound, there is no ECC of positive rate for any $\delta \ge 1/4$. Specializing to $\rho = 1$ we see that a separation-based scheme will have $D = 1/2$ for $1/4 \le \delta < 1/2$, cf. the optimal $D = \delta$.

For general values of ρ we do not know the optimal performance, but can still lower- and upper-bound it. First, a stronger upper bound on ECC rates than the Plotkin bound is given by the *MRRW II* bound [10]:

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

with $\hat{h}(x) = \mathrm{h}_2(1/2 - 1/2\sqrt{1-x})$. In conjunction with (3), this gives an upper bound on the performance of any separation-based scheme. On the other hand, the *Gilbert-Varshamov* bound states that an ECC of rate:

$$R_{GV}(\delta) = 1 - \mathrm{h}_2(2\delta) \quad (5)$$

exists, which gives an achievable performance using separation. An upper bound on the performance of any scheme which is stronger than (2) for $\delta < 1/4$ is given by the *coding converse*:

$$R_{GV}(D^*_{ad}(\rho, \delta)) \le \rho R_{MRRW}(\delta)\,. \quad (6)$$

Finally, for integer ρ we can use a ρ-repetition code, with a majority-vote decoder (with ties broken in an arbitrary manner). It can be shown that for odd ρ, this very simple scheme achieves:

$$D = \frac{2\delta\rho}{\rho + 1}\,. \quad (7)$$

Thus, as in the case $\rho = 1$, for any odd ρ there exists an interval of δ where separation is strictly sub-optimal. Figure 1 demonstrates the various bounds for $\rho = 3$.

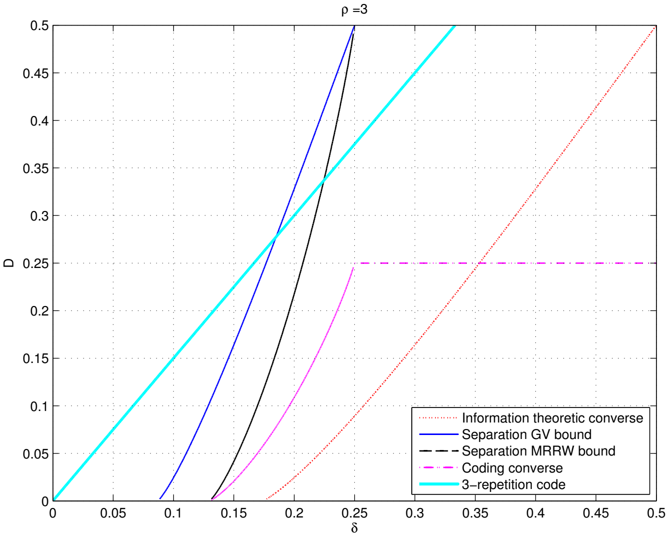

Fig. 1. Trade-off between δ and D in a BSSC(δ) for $\rho = 3$.

The details of this problem can be illustrated by exhibiting the optimal decoder for a given encoding $f : \mathbb{F}_2^k \to \mathbb{F}_2^n$. To that end, for any $S \in \mathbb{F}_2^n$ we define its circumscribed (exterior) radius

$$\mathrm{erad}(S) \quad = \quad \min_{x\in\mathbb{F}_2^n} \max_{y\in S} |y - x| \quad (8)$$

and any x achieving the minimum is called an exterior Chebyshev center of S. Now, let $\mathcal{B}_n(x, r) = \{y : |y - x| \le r\}$

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. This work supported in part by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370.

[1] Assuming that it exists, otherwise limit superior and limit inferior can be defined

be a ball of radius r centered at x in $\mathbb{F}_2^n$. Encoding f is (D,δ)-decodable iff

$$\forall y\,\exists x : f^{-1}B(y,\delta n) \subset B(x, Dk)\,, \tag{9}$$

or, equivalently, the radius of the preimage of every δn-ball does not exceed Dk. The optimal decoder is then:

$$g(y) = \text{exterior center of } f^{-1}\mathcal{B}_n(y,\delta n)\,. \tag{10}$$

The smallest distortion achievable by the encoder f is given by

$$D_{opt}(f,\delta) = \frac{1}{k}\max_{y\in\mathbb{F}_2^n} \mathrm{erad}(f^{-1}\mathcal{B}_n(y,\delta n))\,. \tag{11}$$

Furthermore, an injective f is (D,δ)-decodable iff

$$\forall y\,\exists x :\ B(y,\delta n)\cap \operatorname{Im} f \subset f(B(x,kD))\,,$$

and the optimal distortion-D decoder

$$g(y) = \operatorname*{argmax}_x d_H(y, \mathcal{C}\setminus f(B(x,kD))) \tag{12}$$

searches for the set $f(B(x,kD))$ which contains the largest inscribable sphere centered at y.[2]

The initial results [8] leave many basic questions open. For example, the largest noise level which still leads to non-trivial distortion can be defined as

$$\delta_{max}(\rho) = \sup\left\{\delta : D^*_{ad}(\rho,\delta) < \frac{1}{2}\right\} \tag{13}$$

By the repetition scheme, we know that at least when ρ is an odd integer: $\delta_{max}(\rho) \ge \frac{1}{4}\frac{1+\rho}{\rho}$, while from (3) we know that traditional separated schemes are useless for $\delta > \frac{1}{4}$. What is the actual value of $\delta_{max}(\rho)$? How does it behave as $\rho \to \infty$? In the low distortion regime, separation is optimal [8, Section IV.C]. Does it also achieve the optimal slope of $\delta \mapsto D^*_{ad}(\rho,\delta)$?

In this paper, we discuss relation between the JSCC and other combinatorial problems, such as error-reducing codes (Section II-A), list-decoding of codes and L-multiple sphere packings (Section II-B) and isometric embeddings (Section II-C). Then in Section III we characterize the $D-\delta$ tradeoff of the repetition construction (note: [8] only gave upper and lower bounds). We also show that a certain low-complexity *Markov decoder* is asymptotically optimal.

Generalizing the view beyond the BSSC case is not straightforward. For one, it is not clear what is equivalent of having at most δn flips, since in general there is no channel degradation order. Consequently, [8] proposes an alternative where the adversary is constrained to produce a source and a channel that are both *strongly typical* with respect to some nominal distributions. In Section IV, based upon this model, we further allow the nominal distributions to vary inside some class, thus having a *compound adversary*.

II. Connections to Previous Work

A. Spielman's error-reducing codes

In [11] a closely related concept of an error-reducing code was introduced. An encoder-decoder pair (f,g), $f:\mathbb{F}_2^k\to\mathbb{F}_2^n$ and $g:\mathbb{F}_2^n\to\mathbb{F}_2^k$, is called an error-reducing code of rate $\rho = {}^{n}/_{k}$, error reduction ϵ and distance Δ if

$$\forall x\,\forall |e|\le \Delta n : |x + g(f(x)+e)| < \epsilon\frac{k}{n}|e|\,.$$

In other words, g is a decoder for f that simultaneously achieves all pairs $(D=\epsilon\delta,\delta)$ for all $0\le\delta<\Delta$.

Spielman [11, Section 3] proposed to use a linear f:

$$f(x) = x[I_k\ A]\,, \tag{14}$$

where I_k is a $k\times k$ identity matrix and A is an k by $(n-k)$ adjacency matrix of a bipartite expander graph.

Theorem 1 ([11],[12]): Suppose that the graph $G:[k]\to[m]$ has left degree at most d and is a γ vertex expander for all sets (of left nodes) of size upto αk. Then for (14) there is an $O(k)$-decoder g (see [11]) making (f,g) an error-reducing code with

$$\rho = 1+\frac{m}{k},\ \epsilon = \frac{2\rho}{4\gamma-3d},\ \ \Delta = \frac{\alpha}{\rho(d+1)} \tag{15}$$

Note that on the (D,δ) plane (recall Fig. 1), the performance of Spielman's code always starts at $(D=0,\delta=0)$ with the slope ϵ. The limitation on Δ, however, is quite restrictive. Indeed, even if there existed an expander with $\alpha\approx 1$ and $\gamma\approx d$ we would still only have $\Delta\approx\frac{1}{\rho(d+1)}$, $D_{max} = \epsilon\Delta\approx\frac{2}{d(d+1)}$, demonstrating that Spielman's codes (without additional modifications) are not informative in the regime of $\delta > 1/4$ or $D > 1/4$. As mentioned in Section I, however, given the performance of separated schemes this is the more interesting region; see Fig. 1.

Finally, notice that Spielman introduced the error-reducing codes in order to show that those can be used recursively to produce an error-correcting code which is both efficiently decodable and has a positive relative distance. Although (upto universality in δ), the error-reducing codes are precisely the JSCC codes, it does not follow, however, that any generic JSCC code can be bootstrapped into an error-correcting code. Indeed, the expander-based code (14) possesses an additional very special property: it reduces distortion to exactly zero provided that the parity bits are error free (and the message bits are not too noisy).

B. L-multiple packings and list-decodable codes

Another instructive connection is between the JSCC and L-multiple packings. A set $\mathcal{C}\subset\mathbb{F}_2^n$ is called an L-multiple packing of radius r if

$$\forall y\in\mathbb{F}_2^n :\quad |\mathcal{B}_n(y,r)\cap\mathcal{C}|\le L\,,$$

equivalently,

$$\forall x_1,\ldots,x_{L+1}\in\mathcal{C} :\quad \mathrm{erad}(\{x_1,\ldots,x_{L+1}\}) > r\,, \tag{16}$$

[2] (12) may be interepreted as follows: The goal of the encoder f is to cover the space $\mathbb{F}_2^n$ with sets $U_x = f(B(x,kD))$ in such a way that surrounding every point y is a large δn-ball fully contained in one of U_x's. In other words, the best encoder f will maximize the Lebesgue's number of the covering $\{U_x, x\in\mathbb{F}_2^k\}$.

equivalently, balls of radius r centered at points of $\mathcal{C}$ cover the space with multiplicity at most L; equivalently, $\mathcal{C}$ is an r-error-correcting code with a decoder of list size L. We define

$$A_\ell(n, r, L) = \max\{|\mathcal{C}| : \mathcal{C} \text{ is an } L\text{-packing of radius } r\}\,.$$

The asymptotics of $A_\ell(n, r, L)$ was studied in a number works including [1], [3], [6]. In particular, a simple counting and random coding argument show that when L is growing to infinity with n, e.g. exponentially, we have

$$A_\ell(n, \delta n, \exp\{\lambda n\}) = \exp\{n(1 - h(\delta) + \lambda) + o(n)\}\,. \quad (17)$$

The asymptotic for a fixed L is more delicate. In particular, an elegant upper bound was shown in [3] that interpolates between the Elias-Bassalygo and Hamming bounds as L grows from 1 to ∞.

The connection to the JSCC comes from the following simple observation:

Proposition 2: The image $\operatorname{Im} f$ of any $f : \mathbb{F}_2^k \to \mathbb{F}_2^n$ which is (D, δ)-decodable is an L-multiple packing of radius δn with

$$L = |\mathcal{B}_k(0, kD)| = \sum_{0 \le j \le kD} \binom{k}{j}\,.$$

Proof: Indeed, by (9) every preimage $f^{-1}\mathcal{B}_n(y, \delta n)$ must be contained inside some $\mathcal{B}_k(x_0, kD)$. ■

In view of (17), we see however that asymptotically the converse bound of Proposition 2 reduces to the information theoretic (2). Thus, although it is easy to construct a large constellation $\mathcal{C}$, achieving (17) and such that any δn-ball contains almost exactly $2^{n\lambda}$ points, it is much harder (in fact impossible for most values of ρ and δ) to then label the points of $\mathcal{C}$ with elements of $\mathbb{F}_2^k$ so as to guarantee that each such $2^{n\lambda}$-list has a small radius in $\mathbb{F}_2^k$. In other words, only very special L-multiple packings are good JSCC.

C. Distance-preserving embeddings

In this section we show a connection of JSCC for BSSC to distance-preserving embedding in Hamming space. Distance-preserving embedding of and metric space into other is a well-studied problem, most notably for the ℓ_2-spaces where we have the celebrated Johnson-Lindenstrauss lemma [7]. In Hamming space, there is no such lemma in general as per our knowledge. However some weaker results are true. Below we describe one such result and its consequence.

Suppose, we have an encoding $f : \mathbb{F}_2^k \to \mathbb{F}_2^n$ that is (D, δ)-decodable. A sufficient condition for this property is that for any two points in $\mathbb{F}_2^k$ whose distance is greater than or equal to Dk, must be mapped to two points in $\mathbb{F}_2^n$ at least distance $2\delta n$ apart. To be precise, we have the following lemma.

Lemma 3: Suppose, $f : \mathbb{F}_2^k \to \mathbb{F}_2^n$ and for any $x, y \in \mathbb{F}_2^k$, $|x - y| \ge Dk$ implies $|f(x) - f(y)| \ge 2\delta n$. Then the optimal decoder given in (10) achieves a distortion D. Moreover, the suboptimal decoder $g : \mathbb{F}_2^n \to \mathbb{F}_2^k$ given by the map $g(y) = \arg\min_{x \in \mathbb{F}_2^k} |f(x) - y|$ (resolving ties arbitrarily) achieves distortion D at noise level δ.

Proof: From the condition of the theorem, $\forall y \in \mathbb{F}_2^n$, $\operatorname{erad}(f^{-1}\mathcal{B}_n(y, \delta n)) \le Dk$. Hence the optimal decoder must achieve the distortion D.

We check the case of the suboptimal minimum-distance decoder now. Let $x \in \mathbb{F}_2^k$ is to be encoded and $y = f(x) + u \in \mathbb{F}_2^n$ is received, where $|u| \le \delta n$. Let the decoder output $z \in \mathbb{F}_2^k$. This implies, $|y - f(z)| \le |y - f(x)| = |u| \le \delta n$. And hence, $|f(z) - f(x)| \le |f(x) - y| + |y - f(z)| \le 2\delta n$. But this implies $|z - x| \le Dk$. ■

At this point, we quote a lemma from [14].

Lemma 4: Let $0 \le \epsilon \le \frac{1}{2}$ and $1 \le l \le k$. Let each entry of an $n \times k$ binary matrix R be independent Bernoulli(ϵ^2/l). Then for some $C > 0$ and any $u, v \in \mathbb{F}_2^k$:

$$|u - v| > \frac{l}{2\epsilon} \implies \mathbb{P}\left[|Ru - Rv| > \frac{(1-\epsilon)\epsilon n}{2}\right] \ge 1 - e^{-C\epsilon^3 n}\,.$$

We have the following theorem.

Theorem 5: Let $0 \le \epsilon \le \frac{1}{2}$ and $1 \le l \le k$. Suppose, $C\epsilon^3\rho > \ln 2$, where C is the constant mentioned in Lemma 4 and $\rho = n/k$. Then there exists a matrix R such that for all $u \in \mathbb{F}_2^k$, if $|u| > l/(2\epsilon)$ then $|Ru| > (1-\epsilon)\epsilon n/2$.

Proof: The proof is immediate from Lemma 4, where we take one of the vectors from the pair to be the zero vector. If we had a random matrix as in that lemma, then the probability that the distance condition is not satisfied for at least one $u \in \mathbb{F}_2^k$ is at most $2^k e^{-C\epsilon^3 n} < 1$, from the union bound. This shows existence of a desired matrix. ■

For the case of $f : \mathbb{F}_2^k \to \mathbb{F}_2^n$ being linear, i.e., a $k \times n$ matrix G, the condition of Lemma 3 simplifies. Indeed, G is (D, δ) decodable, if $|u| > Dk$ implies $|Gu| > 2\delta k$ for all $u \in \mathbb{F}_2^k$. This brings us to the following result.

Theorem 6: There exists a linear JSCC given by a matrix G that achieves a distortion D over a BSSC(δ), for $\delta = \frac{\lambda(2D-\lambda)}{16D^2}$, for any $\rho = n/k > \frac{8\ln 2 . D^3}{C\lambda^3}$, where C is the constant given by Lemma 4 and $0 < \lambda < 1$ can be chosen arbitrarily.

Proof: Let us choose, $l = \lambda k$ and $\epsilon = \lambda/(2D)$ in Theorem 5. It is evident that there exists a matrix such that for all $u \in \mathbb{F}_2^k$, if $|u| > Dk$ then $|Ru| > \frac{\lambda(2D-\lambda)}{16D^2}$. ■

The result of Lemma 4 certainly does not permit us to come up with a strong JSCC code. However if this lemma can be replaced with a similar statement with stronger guarantee, then one will be able to construct stronger JSCCs. Nonetheless, this section outlines the connection between JSCCs and the rich literature of distance-preserving embeddings.

III. 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 [8]. In this section we show that such a repetition code can be decoded optimally with very small complexity and "on-the-fly", that is without having to wait for the entire channel output y^n.

Fix an arbitrary encoder given by the mapping $f_1 : \mathbb{F}_2^u \to \mathbb{F}_2^v$, to be called "small code". Based on f_1 we construct longer codes by L-repetition to obtain an $f_L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ with $k = Lu, n = Lv$, and

$$f_L(x_1, \ldots, x_L) = (f_1(x_1), \ldots, f_1(x_L))\,.$$

This yields a sequence of codes with $\rho = {}^{n}\!/\!_{k} = {}^{v}\!/\!_{u}$. It is convenient to think of inputs and outputs in terms the u-ary and v-ary super-letters. To that end we introduce the input $\mathcal{X} = \mathbb{F}_2^u$ and the output $\mathcal{Y} = \mathbb{F}_2^v$ alphabets.

Note that the expressions for the optimal decoder (10) and (12) are too complicated to draw any immediate conclusions. In particular they do not appear to operate on the basis of super-letters. It turns out, however, that there exists a much simpler asymptotically optimal decoder, whose structure we describe next.

Given a transition probability kernel $P_{\hat{S}|Y} : \mathcal{Y} \to \mathcal{X}$ we construct the decoder $g : \mathcal{Y}^L \to \mathcal{X}^L$ as follows. First the estimate $\hat{s}^L \in \mathcal{X}^L$ is initialized to all blanks. Then given a letter $b \in \mathcal{Y}$, the decoder scans y^L for occurrences of b and fills the associated entries of $\hat{s}^L$ with symbols from $\mathcal{X}$ in proportion specified by $P_{\hat{S}|Y}$ (or the best possible rational approximation). The operation is repeated for each $b \in \mathcal{Y}$. Note that this procedure can be realized by a finite-state (Markov) machine: for each letter y_j the decoder outputs the letter $\hat{s}_j$ from $\mathcal{X}$ and updates its state so as to repeatedly cycle through all of $\mathcal{X}$ in proportion specified by $P_{\hat{S}|Y}$ (a good rational approximation of the kernel may need to be precomputed first). A decoder constructed as above will be called a *Markov decoder*. Note that block-by-block decoders discussed in [8] are a special case corresponding to matrices $P_{\hat{S}|Y}$ of 0 and 1.

Theorem 7: Fix a small code $f_1 : \mathbb{F}_2^u \to \mathbb{F}_2^v$ and consider an L-repetition construction. The limit

$$D_\infty(f_1, \delta) = \lim_{L\to\infty} D_{opt}(f_L, \delta)\,,$$

cf. (11), exists and is a non-negative concave function of δ. As any such function it has a dual representation:

$$D_\infty(f_1, \delta) = \inf_{\lambda \ge 0} \lambda\delta - D^*_\infty(f, \lambda)\,,$$

where the concave conjugate $D^*_\infty(f, \lambda)$ is given by any of the following equivalent expressions:

$$\begin{aligned} D^*_\infty(f_1, \lambda) &\triangleq \inf_{\delta \ge 0} \lambda\delta - D_\infty(f_1, \delta) \\ &= \max_{P_{\hat{S}|Y}} \min_{P_{SY}} \mathbb{E}\left[\lambda \frac{|f_1(S) - Y|}{v} - \frac{|S - \hat{S}|}{u}\right] \\ &= \min_{P_{SY}} \max_{P_{\hat{S}|Y}} \mathbb{E}\left[\lambda \frac{|f_1(S) - Y|}{v} - \frac{|S - \hat{S}|}{u}\right]\,, \quad (18) \end{aligned}$$

where the probability space is a Markov chain $S \to Y \to \hat{S}$ with $S, \hat{S} \in \mathcal{X} = \mathbb{F}_2^u$ and $Y \in \mathcal{Y} = \mathbb{F}_2^v$. Moreover, solutions to outer maximizations $\max_{P_{\hat{S}|Y}}$ and $\min_{P_{SY}}$ yield the asymptotically optimal Markov decoder and the worst-case source-adversary realization, respectively.

Remark: We may further simplify (18) to get:

$$D_\infty(f_1, \delta) = \frac{1}{u} \max_{P_{S,Y} : \mathbb{E}[|f_1(S) - Y|] \le \delta v} \sum_y P_Y(y)\, \overline{\mathrm{rad}}(P_{S|Y=y})\,,$$

where

$$\overline{\mathrm{rad}}(P_S) = \min_{\hat{s}} \sum_s P_S(s) |s - \hat{s}|$$

is the moment of inertia of distribution P_S on $\mathbb{F}_2^u$. This moment of inertia (in the special case of uniform P_S on a subset of $\mathbb{F}_2^u$) plays an important role in the study of L-multiple packings [1]–[3].

IV. Compound Adversary

In this section we break from the BSSC model in two ways.

First, we go beyond the binary case. As in [8] we define an adversary (P, W) as one that has to output a source sequence that is strongly-typical w.r.t. P (in the sense of [5]), and a channel output that is strongly-typical given the input. A scheme is said to be (k, n, D)-adversarial for (P, W) similarly to Definition 1. Many of the results presented in the Introduction carry over to this model.

Second, we consider a compound adversary. That is, the adversary can choose to be typical w.r.t. any pair (P, W) in some class $\mathcal{A}$. We say that a JSSC scheme is (k, n, D)-adversarial for $\mathcal{A}$ if it is (k, n, D)-adversarial for all pairs $(P, W) \in \mathcal{A}$. For asymptotic analysis we can define:

$$\begin{aligned} \overline{D}^*_{ad}(\mathcal{A}, \rho) = \limsup_{k\to\infty} \inf\{D : \exists (k, \lfloor \rho k \rfloor, D) \\ \text{adversarial JSCC for } \mathcal{A}\}\,, \quad (19) \\ \underline{D}^*_{ad}(\mathcal{A}, \rho) = \liminf_{k\to\infty} \inf\{D : \exists (k, \lfloor \rho k \rfloor, D) \\ \text{adversarial JSCC for } \mathcal{A}\}\,. \quad (20) \end{aligned}$$

In order to bound these quantities, we first make a digression from JSCC, and treat unequal error protection (UEP) in the combinatorial setting. We then use this for JSCC, not unlike the way in which the achievability JSCC excess-distortion exponent is shown by Csiszár [4].

A. Combinatorial Unequal Error Protection Coding

In the UEP setting, the adversary may choose the channel to be strongly typical with respect to any element W of a set $\mathcal{A}_C$. The encoder is assumed to know W, but the decoder does not. The rate that is sent may depend upon W, and the goal is to have "good" tradeoff between the rates $R(W)$ given different channels.

If W was known to the decoder, then the best known achievable rate is given by the GV bound. In order to generalize the bound (5) beyond the binary case, we use the following definitions. For any input $x \in \mathcal{X}^n$ transmitted through the channel W, let $U_W(x)$ be the set of all possible outputs of the channel W and the *confusability set* of the channel W is defined by $V_W(x) = \{y \in \mathcal{X}^n : U_W(x) \cap U_W(y) \neq \emptyset\}$. Then there exists a codebook for this channel of size $\frac{|\mathcal{X}|^n}{\max_{x\in\mathcal{X}^n} |V_W(x)|}$ [9].[3] The asymptotic rate of this code is

$$\begin{aligned} R_{GV}(W) &\triangleq \lim_{n\to\infty} \frac{1}{n} \log_2 \frac{|\mathcal{X}|^n}{\max_{x\in\mathcal{X}^n} |V_W(x)|} \\ &= \log_2 |\mathcal{X}| - \frac{1}{n} \lim_{n\to\infty} \log_2 \max_{x\in\mathcal{X}^n} |V_W(x)|. \quad (21) \end{aligned}$$

Note that in the case of a binary input-output adversarial channel that can introduce at most δn errors, this rate reduces indeed to (5), as the confusability set is a ball of radius 2δ.

[3] Indeed, the denominator can be improved to $1 + \frac{1}{|\mathcal{X}|^n} \sum_x |V_W(x)|$ for general adversarial channels [13], however we will not use that in the following.

Back in the UEP setting assume that the set $\mathcal{A}_C$ is *degraded*, i.e., it is totally ordered w.r.t. the stochastic channel degradation relation. Then, the following states that $R_{GV}(W)$ is achievable even if W is not known to the decoder.

Theorem 8: If $\mathcal{A}_C$ is a degraded class of channels of size that is polynomial in n, then for any channel $W \in \mathcal{A}_C$ an asymptotic rate of $R_{GV}(W)$ can be achieved.

Proof: Suppose, $W_1, \dots, W_{\ell_n}$ are the channels in $\mathcal{A}_C$ ordered from worst to the best, $|\mathcal{A}_C| = \ell_n$. Let the maximum size of the confusability set of the channel W_i be given by $V_i = \max_{x \in \mathcal{X}^n} |V_{W_i}(x)|$. Let us construct a code by the following greedy (Gilbert) algorithm. Start with any $x_1^{(1)} \in \mathcal{X}^n$ and include it in the codebook. Choose the next codeword $x_2^{(1)}$from $\mathcal{X}^n \setminus V_{W_1}(x_1^{(1)})$. Then, choose $x_j^{(1)}$ from $\mathcal{X}^n \setminus \cup_{k<j} V_{W_1}(x_k^{(1)})$ for $j = 2, \dots, M_1$ where $M_1 = \left\lfloor \frac{|\mathcal{X}|^n}{\ell_n V_1} \right\rfloor$.

In general, set $M_i = \left\lfloor \frac{|\mathcal{X}|^n}{\ell_n V_i} \right\rfloor$, for $i = 1, \dots, \ell_n$. Choose codeword $x_j^{(i)}$ from

$$\mathcal{X}^n \setminus \Big(\cup_{l<i} \cup_{1 \le k \le M_l} V_{W_l}(x_k^{(l)}) \Big) \setminus \Big(\cup_{k<j} V_{W_i}(x_k^{(i)}) \Big).$$

Clearly, the code $\mathcal{C}$ can be partitioned into $\mathcal{C} = \cup_i \mathcal{C}_i$, where $\mathcal{C}_i = \{x_1^{(i)}, \dots, x_{M_i}^{(i)}\}$.

Suppose, the encoder knows now that the channel is W_i. It then chooses its codebook to be $\cup_{j \le i} \mathcal{C}_j$. As $V_{W_i}(x)$ does not contain any codeword other than x for all $x \in \cup_{j \le i} \mathcal{C}_j$ the following decoding is always successful. Suppose $x \in \mathcal{C}_j, j \le i$ is transmitted. Having received $y \in U_{W_i}(x) \subseteq U_{W_j}(x)$, the decoder, for each of $l = 1, 2, \dots,$ tries to find a codeword $\hat{x}$ in $\cup_{k \le l} \mathcal{C}_k$ such that $y \in U_{W_k}(\hat{x})$. It stops whenever it finds one with the smallest l. This would not be possible for any $l < j$ and the correct transmitted vector will then be found. The size of the code is $\sum_{j=1}^{i} M_j = \sum_{j=1}^{i} \left\lfloor \frac{|\mathcal{X}|^n}{\ell_n V_j} \right\rfloor \ge \frac{|\mathcal{X}|^n}{\ell_n V_i}$. Hence the rate of the code is asymptotically $R_{GV}(W_i)$ as ℓ_n is only polynomially growing with n. ■

Note that the assumption on $|\mathcal{A}_C|$ is not restrictive, since for finite alphabet, the total number of possible channels (conditional types) is only polynomial in n.

B. Compound JSCC

We restrict our attention to the case where the class $\mathcal{A}$ is *degraded*. A source Q is (stochastically) degraded w.r.t. a source P if there exists a channel with input distribution P and output distribution P. Degradedness of channels is defined in the same usual way. This is extended to classes as follows.

Definition 2: A class of sources $\mathcal{A}_S$ is degraded if it is totally ordered w.r.t. the stochastic source degradation relation. A class of channels $\mathcal{A}_C$ is degraded if it is totally ordered w.r.t. the stochastic channel degradation relation. A class of source-channel pairs $\mathcal{A}$ is degraded, if it is a subset of the product of degraded source and channel classes $\mathcal{A}_S \times \mathcal{A}_C$.

In such a compound setting, a separation-based scheme suffers from an additional drawback compared to the strongly-typical case. Since the source-channel interface rate is fixed, it must be suitable for the worst-case source and worse-case channel. It is not difficult to see that a separation-based scheme can achieve a distortion D if and only if for any allowed P there exists an ECC of rate $R(P, D)$ for the worst W s.t. $(P, W) \in \mathcal{A}_C$. A general JSCC scheme may do better than this. Intuitively speaking, when the source is such that $R(P, D)$ is low, the rate of the source-channel interface may be adapted in order to accommodate for lower-capacity channels. In order to show this, we use UEP coding.

Theorem 9: If

$$\inf_{(P,W) \in \mathcal{A}} kR(P, D) - nC(W) > 0$$

then no JSCC scheme can achieve D at blocklengths (k, n). For a degraded class, asymptotically, if

$$\inf_{(P,W) \in \mathcal{A}} R(P, D) - \rho R_{GV}(W) < 0$$

then D is achievable.

Proof: The first (converse) part is trivial given the strong-typicality version of (2). For the second (direct) part, we provide the following construction. The encoder observes the source type P, then uses an optimal combinatorial RDF-achieving ("type-covering") codebook of rate $R(P, D)$. The outputs of these possible codebooks are all mapped to an UEP channel codebook, where for any source type P ot is assumed that the channel type is

$$W(P) = \arg \min_{W:(P,W) \in \mathcal{A}} R_{GV}(W).$$

Note that since the number of possible source types is polynomial in n, then so is the number of possible channel assumptions, as required in Theorem 8. By the degradedness property, correct encoding will hold for any allowed adversary in $\mathcal{A}$. ■

Remark: the analysis in this section can also be applied to the BSSC setting.

References

[1] R. Ahlswede and V. Blinovsky. Multiple packing in sum-type metric spaces. *Elect. Notes Disc. Math.*, 21:129–131, 2005.

[2] V. Blinovsky. Plotkin bound generalization to the case of multiple packings. *Prob. Peredachi Inform.*, 45(1):1–4, 2009.

[3] V. M. Blinovsky. Bounds for codes in the case of list decoding of finite volume. *Prob. Peredachi Inform.*, 22(1):7–19, 1986.

[4] I. Csiszár. Joint source-channel error exponent. *Prob. Peredachi Inform.*, 9(5):315–328, 1980.

[5] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic, New York, 1981.

[6] P. Elias. Error-correcting codes for list decoding. *IEEE Trans. Inf. Theory*, 37(1):5–12, 1991.

[7] W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemp. Math.*, 26:189–206, 2006.

[8] Y. Kochman, A. Mazumdar, and Y. Polyanskiy. The adversarial joint source-channel problem. In *Proc. 2012 IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, July 2012.

[9] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, 1997.

[10] 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.

[11] D. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Trans. Inf. Theory*, 42(6):1723–1731, 1996.

[12] M. Sudan. *Algorithmic Introduction to Coding Theory: Lecture Notes*. MIT: 6.897, 2001.

[13] L. Tolhuizen. The generalized Gilbert-Varshamov bound is implied by Turan's theorem. *IEEE Trans. Inf. Theory*, 43(5):1605–1606, 1997.

[14] S. S. Vempala. *The Random Projection Method*. American Mathematical Society, 2004.