Operational Duality Between Lossy Compression and Channel Coding

Ankit Gupta, *Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—We explore the duality between lossy compression and channel coding in the operational sense: whether a capacity-achieving encoder-decoder sequence achieves the rate-distortion function of the dual problem when the channel decoder [encoder] is the source compressor [decompressor, resp.], and vice versa. We show that, if used as a lossy compressor, the maximum-likelihood channel decoder of a randomly chosen capacity-achieving codebook achieves the rate-distortion function almost surely. However, operational duality does not hold for every capacity achieving encoder-decoder sequence, or rate-distortion achieving compressor-decompressor sequence. We show that there exist optimal channel coding [lossy compression] schemes, which fail when used for the dual lossy compression [channel coding resp.] problem.**

***Index Terms*—Channel coding with cost constraints, discrete memoryless channels, discrete memoryless sources, lossy data compression, rate-distortion theory, source-channel coding duality.**

I. Introduction

A. Functional Duality

THE functional duality between the solutions for the capacity of memoryless channels and the rate-distortion function of memoryless sources was recognized by Claude Shannon since the inception of information theory [1] (and further developed in [2]): the capacity-cost function is obtained by maximizing mutual information over input distributions that satisfy an average cost constraint for a fixed conditional output distribution, while the rate-distortion function is obtained my minimizing mutual information over conditional distributions that satisfy a distortion constraint for a fixed source distribution. Functional duality has been exploited in the design of iterative algorithms for the extremization of mutual information [3], [4] and is retained in the presence of side information at the encoder, at the decoder, or at both [5].

In this paper we deal exclusively with the setup where neither encoder nor decoder have side information. In that model, the following general formalization of functional duality was put forth in [5]: Fix a random transformation $P^*_{Y|X}$, input/output alphabets $\mathcal{A}$ and $\mathcal{B}$, and a cost function $f : \mathcal{A} \to \mathbb{R}$. Consider the following optimization problem:

$$C(W) = \max_{\substack{P_X \\ \mathbb{E}[f(X)]\le W}} I(P_X, P^*_{Y|X}) \tag{1}$$

$$= I(P^*_X, P^*_{Y|X}) \tag{2}$$

$$= I(X^*; Y^*), \tag{3}$$

where the maximum is attained by P^*_X and its corresponding output distribution is denoted by P^*_Y.

On the other hand, fix a distribution P^*_S on $\mathcal{B}$ and a distortion function $d : \mathcal{B} \times \mathcal{A} \to \mathbb{R}$, where $\mathcal{A}$ is the reproduction alphabet. Consider the following optimization problem:

$$R(D) = \min_{\substack{P_{\hat{S}|S} \\ \mathbb{E}[d(S,\hat{S})]\le D}} I(P^*_S, P_{\hat{S}|S}) \tag{4}$$

$$= I(P^*_S, P^*_{\hat{S}|S}) \tag{5}$$

$$= I(S^*; \hat{S}^*) \tag{6}$$

where the minimum is attained by $P^*_{\hat{S}|S}$, and $P^*_{\hat{S}}$ is the marginal distribution of $P^*_S P^*_{\hat{S}|S}$.

The optimization problems (1) and (4) are said to be dual if

$$P^*_X(a)P^*_{Y|X}(b|a) = P^*_S(b)P^*_{\hat{S}|S}(a|b), \quad \forall(a,b) \in \mathcal{A}\times\mathcal{B}. \tag{7}$$

In which case

$$C(W) = R(D). \tag{8}$$

It was shown in [5] that duality holds if $\{P^*_{Y|X}, f, W, P^*_S, d, D\}$ is such that

$$P^*_S(b) = P^*_Y(b) \quad \forall b \in \mathcal{B} \tag{9}$$

$$d(b,a) = \alpha \log \frac{1}{P^*_{Y|X}(b|a)} + \beta(b) \tag{10}$$

$$D = \mathbb{E}[d(Y^*, X^*)] = \alpha H(Y^*|X^*) + \mathbb{E}[\beta(Y^*)] \tag{11}$$

are satisfied *or* whenever

$$P^*_{Y|X}(b|a) = P^*_{S|\hat{S}}(b|a) \quad \forall(a,b) \in \mathcal{A}\times\mathcal{B} \tag{12}$$

$$f(a) = \alpha D(P^*_{S|\hat{S}=a} || P^*_S) + \gamma \tag{13}$$

$$W = \mathbb{E}[f(\hat{S}^*)] \tag{14}$$

are satisfied, where $\alpha > 0, \gamma$ are arbitrary constants and $\beta : \mathcal{B} \to \mathbb{R}$ is an arbitrary function.

Manuscript received January 21, 2009; revised June 27, 2010; accepted December 09, 2010. Date of current version May 25, 2011. This work was partially supported by NSF Grants CCF-0635154 and CCF-0728445. This paper was presented in part at the 2009 Information Theory and Applications Workshop, San Diego, CA.

A. Gupta is with Samsung Telecommunications America, Richardson, TX 75082 USA (e-mail: gupta.ankit@samsung.com).

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: verdu@princeton.edu).

Communicated by E. Yang, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2011.2136910

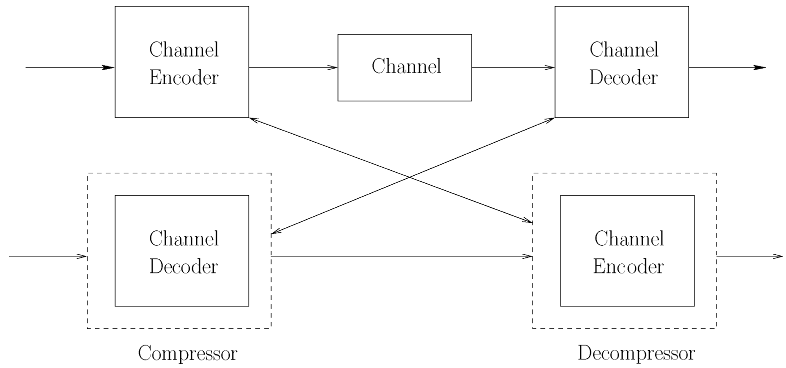

Fig. 1. Operational duality: channel decoder as lossy compressor.

B. Operational Duality

Operational duality refers to the property that optimal encoding/decoding schemes for one problem lead to optimal encoding/decoding schemes for the corresponding dual problem. As established in various degrees of generality in [6], [7], operational duality holds between linear lossless compression and linear encoding for channels with additive discrete noise. In that setup, the source realization is linearly transformed by the channel code parity-check matrix and the decompressor is the syndrome decoder of the channel code. This duality was exploited in [7] and [8] to obtain lossless compressors from capacity-achieving sparse-graph codes. These compressors have performance comparable to or better than existing data compressors such as Lempel-Ziv or Context-Tree Weighting for a wide variety of sources. Returning to lossy compression, it was shown in [9] that linear encoding cannot achieve the rate-distortion function. Therefore, a different approach must be explored for operational duality between lossy compression and channel coding.

It is natural to explore the operational duality depicted in Fig. 1 where the channel decoder becomes the lossy compressor and the channel encoder becomes the lossy decompressor, and vice versa, where the channel encoder/decoder are chosen as a given lossy decompressor/compressor. In addition to the insight offered by functional duality, the operational duality depicted in Fig. 1 might be fruitful because the channel decoder partitions the set of channel output n-sequences into subsets of sequences that are similar in the sense that they are the likely responses of the channel to a given codeword. Analogously a lossy compressor partitions the source n-sequences into subsets of sequences that are similar in the sense that they can be represented by an n-sequence within some distortion. The outputs of both the channel decoder and the lossy compressor simply identify the subset corresponding to the channel output and source realization, respectively.

One of the incentives of establishing operational duality is to capitalize on technological advances in one area in order to provide practical efficient schemes in the other (see Section I-C). Another incentive is the possibility of borrowing proof techniques from one problem to the other. An operational duality argument is used in [10] to prove the achievability side of the rate-distortion theorem. The twist in this argument, which follows the paradigm in Fig. 1 is that good compressors are obtained via large error-probability channel codes. Specifically, let P_S^* be a memoryless source with a distortion function $d(\cdot,\cdot)$ and corresponding optimal $P^*_{S|\hat{S}}$ obtained through (4). Now consider coding for the memoryless channel with the transition probability given in (12). Let $T^n_{P_{\hat{S}^*}}$ denote n-sequences typical according to $P^*_{\hat{S}}$, and let $B(\hat{s}^n)$ denote the set of sequences in $\mathcal{B}^n$ that are typical according to $\prod_{i=1}^n P^*_{S_i|\hat{S}_i=\hat{s}_i}$. Now choose codewords $\{c_1,\ldots,c_M\} \in T^n_{P_{\hat{S}^*}}$ and decoding sets $\{\mathcal{D}^{-1}(c_1),\ldots,\mathcal{D}^{-1}(c_M)\}$ for channel coding such that

$$\mathcal{D}^{-1}(c_i) = B(c_i) - \cup_{j=1}^{i-1}\mathcal{D}^{-1}(c_j) \tag{15}$$

and

$$P^*_{S^n|\hat{S}^n}[\mathcal{D}^{-1}(c_i)|c_i] > 1-\epsilon \tag{16}$$

until all sequences in $T^n_{P_{\hat{S}^*}}$ which satisfy both (15) and (16) are exhausted. Let $\hat{\epsilon}$ be the probability that the distortion between the source and its reproduction exceeds D, when compressed by this maximal[1] channel code in the manner shown in Fig. 1. It is shown in [10] that for any $\tau > 0$

$$\frac{1}{n}\log M \le R(D) + \tau \tag{17}$$

and

$$\hat{\epsilon} \le 1 - \epsilon + 2\tau \tag{18}$$

for all sufficiently large n. Thus, by choosing a large error probability for channel coding ($\epsilon \approx 1$), we can achieve the rate-distortion limits asymptotically. We show this result in greater generality in Theorem 3.

In the special case of the binary symmetric channel (BSC) and the dual problem of compressing the fair coin source with bit error distortion, it was shown in [11] that any sequence of codes for the BSC(p) that have asymptotic rate $1-h(\delta)$, with $p < \delta < \frac{1}{2}$ and probability of error vanishing exponentially in the blocklength, achieve asymptotic distortion less than $\delta + \sqrt{2(h(\delta)-h(p))}$ for compressing the fair coin source under bit error rate criterion, when the maximum-likelihood decoder is used as the lossy compressor.

[1]Where by maximal channel code we mean that no other codeword satisfying both (15) and (16) can be added to the codebook, and thus, its size cannot be increased any further.

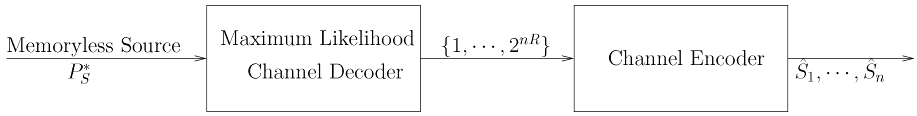

Fig. 2. Lossy compression setup.

C. Sparse-Graph Based Schemes

In view of the fact that sparse-graph codes have been highly successful for approaching channel capacity with practical complexity, a number of schemes have been proposed recently which use sparse-graph channel codes in the fashion shown in Fig. 1 to obtain a lossy compressor. Sparse-graph code based compressors are constructed in [12] and [13] which are optimal for compressing the binary symmetric source with Hamming distortion. In [14] and [15] asymptotically optimal sparse graph codes are constructed for the problem of compressing the m-ary source with Hamming distortion, and the discrete memoryless source with separable distortion respectively. The empirical performance achieved by these schemes is also very close to the optimal rate-distortion tradeoff as seen in [15], [16], and [17].

D. Organization

In Section II we show that a single randomly constructed codebook with maximum likelihood channel decoding is almost surely optimal for both problems. In Section III we show that there exist channel codes which are optimal for channel coding but operate far from the rate-distortion function, when used for lossy compression of the dual source. In Section IV we establish that if a lossy-compression scheme achieves the optimal rate-distortion performance, its dual channel coding scheme (from Fig. 1) may not achieve vanishing probability of error for channel coding. Thus, while operational duality does not hold in general, a random construction in conjunction with the maximum likelihood channel decoder is asymptotically optimal for both problems almost surely. In Section V we revisit the maximal channel coding scheme used in [10] (Section I-B) in greater generality by considering a pair of functionally dual problems, and by removing the restriction of typicality on the maximal codebook selection.

II. Random Codebook Optimality

A channel codebook with blocklength n and rate R is a $2^{nR} \times n$ matrix $\mathcal{C}$ whose coefficients are drawn from $\hat{\mathcal{A}}$. The collection of codebooks with blocklength n and rate R and average cost W, whose maximum likelihood decoder achieves error probability not larger than ϵ is denoted by $\mathcal{F}(n, R, W, \epsilon)$. A channel codebook $\mathcal{C}$, with blocklength n and rate R can be used for lossy compression as depicted in Fig. 2. In that case an n-tuple source realization is compressed into nR bits by passing it through the maximum likelihood decoder of the channel codebook $\mathcal{C}$. The collection of codebooks with blocklength n, rate R such that when used for lossy compression achieve distortion less than or equal to D with probability exceeding $1-\epsilon$ is denoted by $\mathcal{G}(n, R, D, \epsilon)$. The following result shows that a randomly constructed codebook with rate below the capacity of the channel, operates with vanishing error probability on the channel and achieves distortion arbitrarily close to D (for compressing the dual source) as the rate approaches capacity (and $n \to \infty$), with high probability.

Theorem 1: Fix a memoryless channel $P^*_{Y|X}$, and a codeword cost function and cost constraint $f : \mathcal{A} \to R$, and W respectively. Choose a source P^*_S, a bounded distortion function $d : \mathcal{B} \times \mathcal{A} \to \mathbb{R}$, $d(\cdot,\cdot) \leq D_{max}$ and D such that (9)–(11) are satisfied. Let $\mathcal{C}$ denote a $2^{n(R(D)-\gamma)} \times n$ random matrix with coefficients chosen independently from a capacity-achieving distribution P^*_X. Then

$$\lim_{n\to\infty} \mathbb{P}[\mathcal{C} \in \mathcal{F}(n, C(W)-\gamma, W+\gamma, \epsilon) \cap \mathcal{G}(n, R(D)-\gamma, D+\tau(\gamma), \epsilon)] = 1 \quad (19)$$

for all $\epsilon > 0$ and $\gamma > 0$, for some function $\tau(\gamma)$, which vanishes as $\gamma \to 0$.

Proof: Before providing the formal proof, we present an intuitive justification for this result. Thanks to functional duality, the distribution according to which the random channel code is drawn is also the optimal distribution for the random source code. Further, from (10) the maximum likelihood decoder also acts as the minimum distance lossy compressor. Note that for the channel code to work, we need a rate slightly lower than capacity, which means that for the source coding problem, we are working at a rate slightly lower than the rate-distortion function. Thus, in this theorem we show that the slight decrease in the source code rate will only incur a slight penalty in distortion. It must also be emphasized that the penalty is caused from two factors: a) operating at a rate below the rate-distortion function and b) from using a distribution for the random codebook that is not the prescribed one at the given rate. The proof of Theorem 1 is in essence a continuity argument and seeks to show that this penalty vanishes with the gap from capacity.

Our goal is to show that for any $\gamma > 0$, there exists $\tau(\gamma)$, such that $\tau(\gamma) \to 0$ as $\gamma \to 0$, and

$$\lim_{n\to\infty} \mathbb{P}[\mathcal{C} \in \mathcal{F}(n, C(W)-\gamma, W+\gamma, \epsilon)] = 1 \quad (20)$$

and

$$\lim_{n\to\infty} \mathbb{P}[\mathcal{C} \in \mathcal{G}(n, R(D)-\gamma, D+\tau(\gamma), \epsilon)] = 1. \quad (21)$$

The codebook $\mathcal{C}$ contains $2^{n(C(W)-\gamma)} \times n$ symbols, each of which is sampled independently from P^*_X. Since according to (1), $\mathbb{E}[f(X^*)] \leq W$, the law of large numbers implies that $\mathcal{C}$ satisfies the cost constraint averaged over the message w.p.1 (as $n \to \infty$). Further the random-coding proof of the achievability part of the channel coding theorem shows that $\mathcal{C}$ achieves probability of error less than ϵ (for any $\epsilon > 0$), when used for channel coding w.p.1 as $n \to \infty$. Thus

$$\lim_{n\to\infty} \mathbb{P}[\mathcal{C} \in \mathcal{F}(n, C(W)-\gamma, W+\gamma, \epsilon)] = 1. \quad (22)$$

We now compute the distortion achieved with $\mathcal{C}$ when used for lossy compression. First of all note that, for the dual lossy compression problem $P_S^* = P_Y^*$ and $P_{\hat{S}}^* = P_X^*$. Define

$$\lambda_D(d) = \inf_{P_{Z,\hat{Z}}} D(P_{Z,\hat{Z}} \| P_S^* \times P_{\hat{S}}^*) \tag{23}$$

$$= \inf_{P_{Z,\hat{Z}}} [I(Z;\hat{Z}) + D(P_{\hat{Z}} \| P_{\hat{S}}^*)] \tag{24}$$

where $P_{\hat{S}}^*$ is the optimum reproduction distribution when the allowed distortion is D and where the minimum is chosen over all joint distributions $P_{Z,\hat{Z}}$, such that

$$P_Z = P_S^* \tag{25}$$

and

$$E_{P_{Z,\hat{Z}}}[d(Z,\hat{Z})] \le d. \tag{26}$$

The function $\lambda_D(d)$ was first introduced in [19]. Informally, $\lambda_D(d)$ is the minimum rate required for a random codebook with entries chosen independently from a given $P_{\hat{S}}^*$ so as to compress the source P_S^* with average asymptotic distortion less than d [18]. Note that $\lambda_D(D_{max}) = 0$ (by substituting $P_{Z,\hat{Z}} = P_S^* \times P_{\hat{S}}^*$). Further, dropping the non-negative term $D(P_{\hat{Z}} \| P_{\hat{S}}^*)$ from (24), we get

$$\lambda_D(d) \ge \inf_{\substack{P_{Z,\hat{Z}} \\ \mathbb{E}[d(Z,\hat{Z})] \le d \\ P_Z = P_S^*}} I(Z;\hat{Z}) = R(d) \tag{27}$$

with equality at $d = D$, because in that case the optimizing distribution in (24) is $P_{Z\hat{Z}} = P_{S,\hat{S}}^*$. Therefore

$$\lambda_D(D) = R(D). \tag{28}$$

The following result shows that $\lambda_D(d)$ is strictly decreasing in the neighborhood of $d = D$.

Lemma 1: Let $R(D) > 0$, for $0 < \gamma < R(D)/2$ there exists $\rho(\gamma)$ such that

$$\lim_{\gamma \to 0} \rho(\gamma) = 0. \tag{29}$$

and

$$\lambda_D(D + \rho(\gamma)) < R(D) - \gamma. \tag{30}$$

Proof: Note that, by its definition, (24), $\lambda_D(d)$ is convex and nonincreasing. Due to convexity, it must be continuous in its region of definition. Moreover, if it is constant in some neighborhood of distortion d_1 it must be constant for all $d > d_1$. Since $\lambda_D(d)$ cannot be negative and $\lambda_D(D_{max}) = 0$, we conclude that it must be strictly decreasing for any d such that $\lambda_D(d) > 0$. ■

Define the per-letter distortion as

$$d^n(a^n, b^n) = \frac{1}{n} \sum_{i=1}^{n} d(a_i, b_i) \tag{31}$$

and

$$\mathcal{E}(n,\delta) = \left\{ s^n \in \mathcal{B}^n : \frac{1}{n} \log \frac{1}{P_{\hat{S}^n}^*[d^n(s^n, \hat{S}^n) \le d]} < \lambda_D(d) + \delta \right\}. \tag{32}$$

It is shown in [18] that for all $\delta > 0$

$$\lim_{n \to \infty} P_{S^n}^*(\mathcal{E}(n,\delta)) = 1. \tag{33}$$

For convenience we denote, with a slight abuse of notation, the minimum distortion between a source sequence s^n and a codeword in $\mathcal{C}$ by

$$d^n(s^n, \mathcal{C}) = \min_{c^n \in \mathcal{C}} d^n(s^n, c^n). \tag{34}$$

Since we have $2^{nC(W) - n\gamma} = 2^{nR(D) - n\gamma}$ codewords

$$\mathbb{P}[d^n(s^n, \mathcal{C}) \ge D + \rho(2\gamma)]$$
$$< (1 - 2^{-n(\lambda_D(D + \rho(2\gamma)) + \delta) + o(n)})^{2^{n(C(W) - \gamma)}} \tag{35}$$
$$\le e^{-2^{n(C(W) - \gamma - \lambda_D(D + \rho(2\gamma)) - \delta)}} \tag{36}$$

for $s^n \in \mathcal{E}(n,\delta)$, where the randomness in (35) is only due to $\mathcal{C}$ and (36) is obtained by using the inequality $(1-x)^y \le e^{-xy}$. The probability of exceeding distortion $D + \rho(2\gamma)$ on any element of $\mathcal{E}(n,\delta)$ is given by

$$\mathbb{P}\left[\bigcup_{s^n \in \mathcal{E}(n,\delta)} \{d^n(s^n, \mathcal{C}) \ge D + \rho(2\gamma)\} \right]$$
$$\le |\mathcal{E}(n,\delta)| e^{(-2^{n(C(W) - \gamma - \lambda_D(D + \rho(2\gamma)) - \delta)})} \tag{37}$$
$$\le 2^{n \log |\mathcal{B}|} e^{(-2^{n(C(W) - \gamma - \lambda_D(D + \rho(2\gamma)) - \delta)})} \tag{38}$$

where (38) follows from the fact that $\mathcal{E}(n,\delta) \subset \mathcal{B}^n$. Therefore, if $\delta \le \gamma$, then

$$\lim_{n \to \infty} \mathbb{P}\left[\bigcup_{s^n \in \mathcal{E}(n,\delta)} \{d^n(s^n, \mathcal{C}) \ge D + \rho(2\gamma)\} \right] = 0. \tag{39}$$

Define $\mathcal{H}(n, \delta, \hat{\epsilon}, R)$ as the set of all codebooks $\mathcal{C}$ of rate R such that

$$\max_{s^n \in \mathcal{E}(n,\delta)} d^n(s^n, \mathcal{C}) \le D + \hat{\epsilon}. \tag{40}$$

From (39) we have

$$\lim_{n \to \infty} \mathbb{P}[\mathcal{C} \in \mathcal{H}(n, \gamma, \rho(2\gamma), R(D) - \gamma)] = 1. \tag{41}$$

From (33), for n large enough and arbitrary $\epsilon > 0$

$$P_{S^n}^*(\mathcal{E}^c(n,\gamma)) \le \epsilon \tag{42}$$

thus for sufficiently large n

$$\mathcal{H}(n,\gamma,\rho(2\gamma),R(D)-\gamma)\subset\mathcal{G}(n,R(D)-\gamma,D+\rho(2\gamma),\epsilon). \quad (43)$$

From (41)

$$\lim_{n\to\infty}\mathbb{P}[\mathcal{C}\in\mathcal{G}(n,R(D)-\gamma,D+\rho(2\gamma),\epsilon)]=1. \quad (44)$$

From (22) and (44)

$$\lim_{n\to\infty}\mathbb{P}[\mathcal{C}\in\mathcal{G}(n,R(D)-\gamma,D+\rho(2\gamma),\epsilon) \cap\mathcal{F}(n,C(W)-\gamma,W+\gamma,\epsilon)]=1. \quad (45)$$

Any codebook in the intersection in (45) achieves a probability of error less than ϵ for channel coding, and achieves a distortion $D+\rho(2\gamma)$ with probability greater than $1-\epsilon$ in lossy compression. Thus, we can simply choose $\tau(\gamma)=\rho(2\gamma)$ to conclude that a random codebook operating at rate $C(W)-\gamma$, where γ is sufficiently small is good for both source and channel coding, almost surely. ■

III. Optimal Channel Codes That are Not Optimal for Lossy Compression

In Section II, we established that with overwhelming probability, a maximum likelihood channel decoder is a lossy compressor that achieves the rate-distortion function of the dual lossy compression problem for a capacity-achieving random channel codebook ensemble. This prompts a natural question: Is a sequence of channel encoders/decoders good for lossy compression as long it operates close to capacity and achieves vanishing probability of error? In this section we show that the answer is negative. Indeed we show that for any regular channel codebook[2] that achieves capacity with a maximum likelihood decoder, there exists an alternative capacity-achieving decoder that leads to distortion far from the ideal when used as a lossy compressor.

Theorem 2: Fix a memoryless channel $P^*_{Y|X}$, a codeword cost function $f:\mathcal{A}\to\mathbb{R}$, and cost constraint W respectively. Choose a source P^*_S, a distortion function $d(\cdot,\cdot):\mathcal{B}\times\mathcal{A}\to\mathbb{R}$ and D such that (9)–(11) are satisfied, and $d(\cdot,\cdot)$ and $\beta(\cdot)$ are bounded. Choose any sequence of regular codebooks with asymptotic rate strictly smaller than $C(W)$, whose maximum likelihood decoders achieve

$$\lim_{n\to\infty}\epsilon_n=0. \quad (46)$$

Then, there exists a sequence of channel decoders with vanishing error probability that achieve distortion D_n when used as lossy compressors of the dual source, such that

$$\liminf_{n\to\infty}D_n\geq D+\alpha C(W) \quad (47)$$

with α specified in (10).[3]

[2]A regular codebook [20] only contains symbols $a\in\mathcal{A}$ that achieve $D(P^*_{Y|X=a}||P^*_Y)=C(W)$ for the capacity-achieving output distribution P^*_Y.

[3]Note that the lower bound on the gap from the optimal distortion: $\alpha C(W)$, is independent of the gap from capacity.

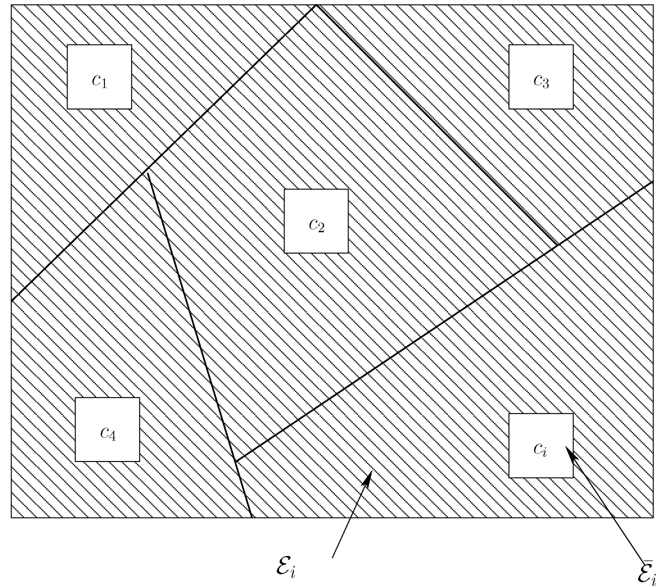

Fig. 3. Graphical representation of the sets defined in (50) and (54).

Proof: Denote the number of codewords in the n^{th} codebook by M_n, and let

$$\gamma=C(W)-\limsup_{n\to\infty}\frac{\log M_n}{n}>0. \quad (48)$$

The set of channel output sequences is partitioned by

$$\bigcup_{i=1}^{M_n}\mathcal{D}^{-1}(c_i)=\mathcal{B}^n. \quad (49)$$

where $\mathcal{D}^{-1}(c_i)\subset\mathcal{B}^n$ is the subset of channel outputs for which the maximum-likelihood decoder $\mathcal{D}$ declares c_i. We will show that there exist subsets $\mathcal{E}_i\subset\mathcal{D}^{-1}(c_i)$ (see Fig. 3), such that the probability that the channel output falls on $\mathcal{E}_i$ given that the codeword c_i was transmitted vanishes asymptotically, and the probability that the source $P^*_{S^n}$ takes values on $\bigcup_{i=1}^{M_n}\mathcal{E}_i$ goes to 1 asymptotically. Therefore, if we modify the channel decoder so that it outputs a fixed codeword when the channel output is in $\bigcup_{i=1}^{M_n}\mathcal{E}_i$, it will still be asymptotically optimal for channel coding, but will perform far from optimal as a lossy compressor, because it outputs the same reproduction sequence for almost all source sequences. The following lemma gives a construction of the desired subsets $\{\mathcal{E}_i\}_{i=1}^{M_n}$:

Lemma 2: Define for arbitrary $\delta>0$

$$\mathcal{E}_i=\left\{s^n\in\mathcal{D}^{-1}(c_i):\frac{1}{n}\sum_{\ell=1}^{n}\imath(s_\ell,c_i[\ell])<C(W)-\delta\right\} \quad (50)$$

with the information density denoted by

$$\imath(b,a)=\log\frac{P^*_{Y|X}(b|a)}{P^*_Y(b)}. \quad (51)$$

If the codebook is regular, then

$$\lim_{n\to\infty}P^*_{Y^n|X^n}(\mathcal{E}_i|c_i)=0 \quad (52)$$

for every codeword c_i, where the convergence in (52) is uniform.

Proof: Thanks to the assumption that the code is regular and the alphabets are finite, $\frac{1}{n}\sum_{\ell=1}^{n} \imath(s_\ell, c_i[\ell])$ is equal to the arithmetic average of the random variables

$$Y_n[a] = \frac{1}{n} \sum_{\ell \in \mathcal{S}(a)} \imath(s_\ell, a) \quad (53)$$

where $j \in \mathcal{S}(a)$ if and only if $c_i[j] = a$. For those, $a \in \mathcal{A}$ such that $|\mathcal{S}(a)| \to \infty$, the law of large numbers ensures that $Y_n \to C(W)$ in probability. For the other symbols, their contribution to $\frac{1}{n}\sum_{\ell=1}^{n} \imath(s_\ell, c_i[\ell])$ vanishes. Therefore, regardless of the empirical distribution of c_i, (52) follows. ■

Henceforth, we choose $\delta = \gamma/2$ in definition (50). Let

$$\bar{\mathcal{E}}_i = \mathcal{D}^{-1}(c_i) \setminus \mathcal{E}_i. \quad (54)$$

We have (cf. Fig. 3)

$$(\bigcup_{i=1}^{M_n} \mathcal{E}_i) \cap (\bigcup_{i=1}^{M_n} \bar{\mathcal{E}}_i) = \phi \quad (55)$$

and

$$(\bigcup_{i=1}^{M_n} \mathcal{E}_i) \cup (\bigcup_{i=1}^{M_n} \bar{\mathcal{E}}_i) = \bigcup_{i=1}^{M_n} \mathcal{D}^{-1}(c_i) = \mathcal{B}^n. \quad (56)$$

We now show that the source realization lies in the union of the sets $\{\mathcal{E}_i\}_{i=1}^{M_n}$ with overwhelming probability.

Lemma 3:

$$\lim_{n\to\infty} P^*_{S^n}(\bigcup_{i=1}^{M_n} \mathcal{E}_i) = 1. \quad (57)$$

Proof: Denoting the output of the maximum-likelihood channel decoder as $\mathcal{D}(\cdot)$, from (54) we have

$$P^*_{S^n}(s^n) \le 2^{-n(C(W)-\gamma/2)} P^*_{S^n|\hat{S}^n}(s^n|\mathcal{D}(s^n)) \quad (58)$$

for any sequence $s^n \in \bar{\mathcal{E}}_i$, therefore

$$\begin{aligned} &P^*_{S^n}(\bigcup_{i=1}^{M_n} \bar{\mathcal{E}}_i) \\ &\le \sum_{s^n \in \bigcup_{i=1}^{M_n} \bar{\mathcal{E}}_i} P^*_{S^n|\hat{S}^n}(s^n|\mathcal{D}(s^n)) 2^{-n(C(W)-\gamma/2)} && (59) \\ &\le 2^{-n(C(W)-\gamma/2)} \sum_{i=1}^{M_n} \sum_{s^n \in \mathcal{E}_i} P^*_{S^n|\hat{S}^n}(s^n|c_i) && (60) \\ &\le M_n 2^{-n(C(W)-\gamma/2)} && (61) \\ &\le 2^{-n\gamma/2} && (62) \end{aligned}$$

where (62) follows from (48). Using (62) along with (56), we obtain the desired result upon taking limits. ■

We now modify the decoder $\mathcal{D}$ such that the new decoder $\bar{\mathcal{D}}$ has an asymptotically vanishing error probability

$$\bar{\mathcal{D}}^{-1}(c_1) = \mathcal{D}^{-1}(c_1) \cup (\bigcup_{i=1}^{M_n} \mathcal{E}_i) \quad (63)$$

and for $j \neq 1$

$$\bar{\mathcal{D}}^{-1}(c_j) = \mathcal{D}^{-1}(c_j) \setminus \mathcal{E}_j. \quad (64)$$

With this modification for $j = 1$

$$\begin{aligned} P^*_{Y^n|X^n}(\bar{\mathcal{D}}^{-1}(c_1)|c_1) &\ge P^*_{Y^n|X^n}(\mathcal{D}^{-1}(c_1)|c_1) && (65) \\ &\ge 1 - \epsilon_n && (66) \end{aligned}$$

and for $j \neq 1$

$$\begin{aligned} P^*_{Y^n|X^n}(\bar{\mathcal{D}}^{-1}(c_j)|c_j) &\ge P^*_{Y^n|X^n}(\mathcal{D}^{-1}(c_j) \setminus \mathcal{E}_j|c_j) && (67) \\ &\ge 1 - \epsilon_n - P^*_{Y^n|X^n}(\mathcal{E}_j|c_j). && (68) \end{aligned}$$

Accordingly, the modified scheme has an error rate

$$\begin{aligned} \epsilon'_n &\le \epsilon_n + \max_{i=1,\ldots,M_n} P^*_{Y^n|X^n}(\mathcal{E}_i|c_i) && (69) \\ &= o(1). && (70) \end{aligned}$$

Therefore, the codebook/decoder sequence $\{\mathcal{C}_n, \bar{\mathcal{D}}\}$ is asymptotically optimal for channel coding. We now show that the distortion obtained when this channel code is used to compress the dual source, in the manner shown in Fig. 1, satisfies (47). Define

$$\imath^n(y^n, x^n) = \frac{1}{n} \sum_{\ell=1}^{n} \imath(y_\ell, x_\ell) \quad (71)$$

where $\imath(\cdot,\cdot)$ is given in (51). The average distortion is equal to

$$\begin{aligned} D_n &= \frac{\alpha}{n} \sum_{s^n \in \mathcal{A}^n} P^*_{S^n}(s^n) \log \frac{1}{P^*_{S^n|\hat{S}^n}(s^n|\bar{\mathcal{D}}(s^n))} + \mathbb{E}[\beta(S^*)] && (72) \\ &= \frac{\alpha}{n} \sum_{s^n \in \mathcal{A}^n} P^*_{S^n}(s^n) \log \frac{P^*_{S^n}(s^n)}{P^*_{S^n|\hat{S}^n}(s^n|\bar{\mathcal{D}}(s^n))} + \alpha H(S^*) \\ &\quad + \mathbb{E}[\beta(S^*)] && (73) \\ &= \alpha H(S^*|\hat{S}^*) + \alpha C(W) + \mathbb{E}[\beta(S^*)] \\ &\quad - \alpha \sum_{s^n \in \mathcal{A}^n} P^*_{S^n}(s^n) \imath^n(s^n, \bar{\mathcal{D}}(s^n)) && (74) \\ &= \alpha C(W) + D - \alpha \sum_{s^n \in \cup_{i=1}^{M_n} \mathcal{E}_i} P^*_{S^n}(s^n) \imath^n(s^n, \bar{\mathcal{D}}(s^n)) \\ &\quad - \alpha \sum_{s^n \in \cup_{i=1}^{M_n} \bar{\mathcal{E}}_i} P^*_{S^n}(s^n) \imath^n(s^n, \bar{\mathcal{D}}(s^n)) && (75) \end{aligned}$$

where (75) follows from (11) and (56). Define

$$r_n = \alpha \sum_{s^n \in \cup_{i=1}^{M_n} \bar{\mathcal{E}}_i} P^*_{S^n}(s^n) \imath^n(s^n, \bar{\mathcal{D}}(s^n)) \quad (76)$$

and

$$l_n = \alpha \sum_{s^n \in \cup_{i=1}^{M_n} \bar{\mathcal{E}}_i} P^*_{S^n}(s^n) \imath^n(s^n, c_1). \quad (77)$$

From Lemma 4 in the Appendix and (62), we have

$$\lim_{n\to\infty} r_n = \lim_{n\to\infty} l_n = 0. \quad (78)$$

We now evaluate (75) in the limit

$$\begin{aligned}\limsup_{n\to\infty} D_n &= \alpha C(W) + D - \lim_{n\to\infty} r_n \\ &\quad -\alpha \limsup_{n\to\infty} \sum_{s^n\in\cup_{i=1}^{M_n}\mathcal{E}_i} P^*_{S^n}(s^n)\imath^n(s^n,c_1) \quad (79)\\ &= \alpha C(W) + D + \lim_{n\to\infty} l_n \\ &\quad -\alpha \limsup_{n\to\infty} \sum_{s^n\in\mathcal{A}^n} P^*_{S^n}(s^n)\imath^n(s^n,c_1) \quad (80)\\ &= \alpha C(W) + D \\ &\quad + \limsup_{n\to\infty} \frac{\alpha}{n} D(P^*_{S^n}\|P^*_{S^n|\hat{S}^n=c_1}) \quad (81)\\ &\geq \alpha C(W) + D. \quad (82)\end{aligned}$$

Hence, we have shown a channel coding scheme that has vanishing probability of error for channel coding and rate arbitrarily close to capacity, but performs far from the rate-distortion function when used as a lossy compressor. ■

IV. Optimal Lossy Compressors for Channel Coding

We now consider the operational duality setup that is complementary to the one resolved in Section III. We analyze whether a rate-distortion achieving lossy compression scheme also achieves vanishing probability of error for channel coding, when the lossy compressor is used as a channel decoder and the lossy decompressor is used as the channel encoder. Note that the rate-distortion achieving schemes that we consider must approach $R(D)$ from below. Otherwise, the channel coding theorem ensures that probability of error for channel coding is bounded away from zero.

A simple counterexample shows that optimal rate-distortion performance is not a guarantee of achieving channel-capacity for the functionally dual problem. Consider the problem of compressing the fair coin source with Hamming distortion δ. The dual channel-coding problem is transmission over the binary symmetric channel with crossover probability δ. Given a sequence of codebooks and lossy compressors, that have asymptotic rate $R(\delta)-\epsilon$ and asymptotic average distortion $\delta+\gamma$, we can obtain a new sequence of compressors and decompressors by doubling the number of codewords simply adding a neighbor at unit Hamming distance to each codeword in the codebook, and using the maximum-likelihood channel decoder as a lossy compressor. For this new sequence the asymptotic rate-distortion performance is unchanged, while the channel error probability P_e is lower bounded by δ: the probability that the transmitted codeword will be received as its unit Hamming distance neighbor. Thus, irrespective the magnitude of ϵ and γ the probability of error for channel coding is bounded away from zero, establishing that rate-distortion achieving schemes may fail for the functionally-dual channel-coding problem.

V. Lossy Data Compressor From Maximal Channel Code

In the spirit of [10], we give a duality result based on a maximal channel code construction. In contrast to [10], we consider a pair of functionally dual problems from which our codebook selection follows naturally, without imposing any typicality constraints.

A. Code Construction

Consider the rate-distortion problem

$$R(D) = \min_{\substack{P_{\hat{S}|S}\\ \mathbb{E}[d(S,\hat{S})]\leq D}} I(P^*_S, P_{\hat{S}|S}) = I(P^*_S, P^*_{\hat{S}|S}) \quad (83)$$

and the dual channel coding problem $(R(D) = C(W))$

$$C(W) = \max_{\substack{P_{\hat{S}}\\ \mathbb{E}[f(\hat{S})]\leq W}} I(P_{\hat{S}}, P^*_{S|\hat{S}}) = I(P^*_{\hat{S}}, P^*_{S|\hat{S}}). \quad (84)$$

Fix arbitrary $\delta, \delta' > 0$ and $0 < \epsilon < 1$. Define the balls

$$B(x^n) = \{s^n \in \mathcal{B}^n : d^n(s^n, x^n) \leq D + \delta\}. \quad (85)$$

Consider a codebook $\mathcal{C} = \{c_i \in \mathcal{A}^n : i = 1, \ldots, M\}$ such that each codeword satisfies

$$f^n(c_i) = \frac{1}{n}\sum_{j=1}^{n} f(c_i[j]) \leq W + \delta' \quad (86)$$

and

$$P^*_{S^n|\hat{S}^n}(\mathcal{D}^{-1}(c_i)|c_i) > 1 - \epsilon \quad (87)$$

where

$$\mathcal{D}^{-1}(c_i) = B(c_i) - \cup_{j=1}^{i-1} B(c_j). \quad (88)$$

Let the channel decoder[4] be defined as

$$\mathcal{D}(b^n) = \begin{cases} c_i, & b^n \in \mathcal{D}^{-1}(c_i) \\ a_0^n, & \text{otherwise} \end{cases} \quad (89)$$

for an arbitrary fixed $a_0^n \in \mathcal{A}^n$.

B. Analysis

Theorem 3: Fix $\delta > 0$, $\delta' > 0$, and $\epsilon > 0$. For the code construction in Section V-A

$$\limsup_{n\to\infty} \frac{1}{n}\log M \leq C(W + \delta'). \quad (90)$$

If the codebook is maximal, i.e., we cannot find $a^n \in \mathcal{A}^n \setminus \mathcal{C}$ that satisfies (86)–(87), then it attains the excess distortion probability

$$\limsup_{n\to\infty} \mathbb{P}[d^n(Y^n, \mathcal{D}(Y^n)) > D + \delta] \leq 1 - \epsilon, \quad (91)$$

where Y^n is distributed according to $P^*_{S^n}$.

Proof: Let $\{\epsilon_k\}$ be a sequence such that

$$\lim_{k\to\infty} \epsilon_k = 0 \quad (92)$$

and

$$\lim_{k\to\infty} \sqrt{k}\epsilon_k = \infty. \quad (93)$$

[4]Note that this decoder outputs a codeword rather than a message.

Let $T^k_{P_{\hat{S}}}$ denote elements of $\mathcal{A}^k$, which are ϵ_k-typical according to $P_{\hat{S}}$. Any codeword c_i satisfying (86) is in $T^n_{P'_{\hat{S}}}$ with

$$P'_{\hat{S}}(a) = \frac{1}{n}\sum_{i=1}^{n} 1\{c_i = a\} \tag{94}$$

which satisfies

$$\mathbb{E}[f(\hat{S}')] \leq W + \delta' \tag{95}$$

and therefore

$$I(P'_{\hat{S}}, P^*_{S|\hat{S}}) \leq C(W + \delta'). \tag{96}$$

From [10, Corollary 2.1.4], for arbitrary $\varsigma > 0$ there exists $n(\varsigma)$, such that for $n > n(\varsigma)$

$$|\mathcal{C} \cap T^n_{P'_{\hat{S}}}| \leq \exp(n(I(P'_{\hat{S}}, P^*_{S|\hat{S}}) + \varsigma)) \tag{97}$$

$$\leq \exp(n(C(W + \delta') + \varsigma)) \tag{98}$$

since the number of types is upper bounded by $(n+1)^{|\mathcal{A}|}$

$$M \leq (n+1)^{|\mathcal{A}|} \exp(n(C(W + \delta') + \varsigma)) \tag{99}$$

therefore, the strong converse statement in (90) follows.

We now proceed to show (91). To that end, note from (85) that every element of

$$B = \bigcup_{i=1}^{M} B(c_i) \tag{100}$$

is reconstructed with distortion less than or equal to $D + \delta$. For Y^n distributed according to $P^*_{S^n}$, we have

$$\mathbb{P}[d^n(Y^n, \mathcal{D}(Y^n)) > D + \delta] \leq P^*_{S^n}[B^c]. \tag{101}$$

If $\hat{s}^n \in \mathcal{A}^n$ and

$$f(\hat{s}^n) \leq W + \delta' \tag{102}$$

then

$$P^*_{S^n|\hat{S}^n}[B^c \cap B(\hat{s}^n)|\hat{s}^n] < 1 - \epsilon \tag{103}$$

because the code is maximal. Since $\mathbb{E}[f(\hat{S}^*)] \leq W$, there exists $n_0(\delta')$ such that for $n > n_0(\delta')$

$$f(a^n) \leq W + \delta', \quad \forall a^n \in T^n_{P^*_{\hat{S}}}. \tag{104}$$

Now, let $\hat{s}^n \in T^n_{P^*_{\hat{S}}}$ and let S^n be generated through the random transformation $P^*_{S^n|\hat{S}^n=\hat{s}^n}$. Fix an arbitrary $\tau > 0$. The law of large numbers guarantees the existence of $n_1(\tau)$ such that for $n > n_1(\tau)$

$$P^*_{\hat{S}^n}[T^n_{P^*_{\hat{S}}}] > 1 - \tau \tag{105}$$

and

$$\mathbb{P}[(S^n, \hat{s}^n) \in T^n_{P^*_{S\hat{S}}}] > 1 - \tau. \tag{106}$$

Furthermore, since $\mathbb{E}[d(S^*, \hat{S}^*)] \leq D$, there exists $n_2(\delta)$, such that for $n > n_2(\delta)$

$$d^n(b^n, a^n) \leq D + \delta, \quad \forall (b^n, a^n) \in T^n_{P^*_{S\hat{S}}}. \tag{107}$$

Combining (106) and (107), for $n > \max\{n_1(\tau), n_2(\delta)\}$

$$\mathbb{P}[d^n(S^n, \hat{s}^n) \leq D + \delta] > 1 - \tau \tag{108}$$

which, according to (54) implies that

$$P^*_{S^n|\hat{S}^n}[B(\hat{s}^n)|\hat{s}^n] > 1 - \tau. \tag{109}$$

Therefore, for $n > \max\{n_0(\delta'), n_1(\tau), n_2(\delta)\}$, and $\hat{s}^n \in T^n_{P^*_{\hat{S}}}$

$$P^*_{S^n|\hat{S}^n}[B|\hat{s}^n] = 1 - P^*_{S^n|\hat{S}^n}[B^c|\hat{s}^n] \tag{110}$$

$$= 1 - P^*_{S^n|\hat{S}^n}[B^c \cap B(\hat{s}^n)|\hat{s}^n] - P^*_{S^n|\hat{S}^n}[B^c \cap B^c(\hat{s}^n)|\hat{s}^n] \tag{111}$$

$$> \epsilon - P^*_{S^n|\hat{S}^n}[B^c \cap B^c(\hat{s}^n)|\hat{s}^n] \tag{112}$$

$$> \epsilon - P^*_{S^n|\hat{S}^n}[B^c(\hat{s}^n)|\hat{s}^n] \tag{113}$$

$$> \epsilon - \tau \tag{114}$$

where we get (112) by using (103), and (114) follows from (109). We may use (114) to conclude that for $n > \max(n_0(\delta'), n_1(\tau), n_2(\delta))$

$$P^*_{S^n}[B] = \sum_{\hat{s}^n \in \mathcal{A}^n} P^*_{S^n|\hat{S}^n}[B|\hat{s}^n] P^*_{\hat{S}^n}[\hat{s}^n] \tag{115}$$

$$\geq \sum_{\hat{s}^n \in T^n_{P^*_{\hat{S}}}} P^*_{S^n|\hat{S}^n}[B|\hat{s}^n] P^*_{\hat{S}^n}[\hat{s}^n] \tag{116}$$

$$\geq (\epsilon - \tau)(1 - \tau) \tag{117}$$

$$\geq \epsilon - 2\tau. \tag{118}$$

Finally, (91) follows from (118) and (101). ■

Appendix

Lemma 4: Consider the rate-distortion problem given in (4) with the distortion function $d : \mathcal{B} \times \mathcal{A} \to \mathbb{R}$ given in (10). Let $0 \leq d(\cdot,\cdot) \leq D_{max}$ and $|\beta(\cdot)| \leq \beta_{max}$, then for an arbitrary $\hat{s}^n \in \mathcal{A}^n$

$$|\imath^n(s^n, \hat{s}^n)| \leq K \tag{119}$$

whenever $P^*_{S^n}(s^n) > 0$, for some constant K, where $\imath^n(\cdot,\cdot)$ is specified in (71).

Proof: In view of (10), for any $(a, b) \in \mathcal{A} \times \mathcal{B}$ such that $P^*_S(b) > 0$, we have

$$\min_{\substack{b' \in \mathcal{B} \\ P^*_S(b') \neq 0}} \log P^*_S(b') - \frac{\beta_{max}}{\alpha} \leq -\imath(b, a) \leq \frac{D_{max} + \beta_{max}}{\alpha}. \tag{120}$$

Since by assumption $P_S^*(s_j) > 0,\ j \in \{1, 2, \ldots, n\}$, (71) and (120) enable us to conclude that (119) holds with

$$K = \max \left\{ \left| \min_{\substack{b' \in \mathcal{B} \\ P_S^*(b') \neq 0}} \log P_S^*(b') - \frac{\beta_{max}}{\alpha} \right|, \left| \frac{D_{max} + \beta_{max}}{\alpha} \right| \right\}. \tag{121}$$

■

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul./Oct. 1948.

[2] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *Proc. IRE Nat. Conv. Rec.*, Mar. 1959, pp. 142–163.

[3] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 460–473, Jul. 1972.

[4] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 14–20, Jan. 1972.

[5] S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1181–1203, May 2003.

[6] T. C. Ancheta, "Syndrome source coding and its universal generalizations," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 3, pp. 423–428, Mar. 1976.

[7] G. Caire, S. Shamai, and S. Verdú, "Lossless data compression with error correcting codes," *Adv. Netw. Inf. Theory*, vol. 66, pp. 263–284, 2004, DIMACS Ser. Discrete Mathematics and Theoretical Computer Science, AMS.

[8] G. Caire, S. Shamai, A. Shokrollahi, and S. Verdú, "Fountain codes for lossless data compression," *Algebraic Coding Theory Inf. Theory*, vol. 68, pp. 1–20, 2006, DIMACS Ser. Discrete Mathematics and Theoretical Computer Science.

[9] J. L. Massey, "Joint source and channel coding," *Commun. Syst. Random Process. Theory*, vol. 11, pp. 279–293, 1978.

[10] I. Csiszar and J. Korner, *Information Theory, Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[11] V. Chandar, "Iterative Algorithms for Lossy Source Coding," M.S. thesis, Dept. Elect. Eng. Comput. Sci., Mass. Inst. Technol., Cambridge, MA, 2006.

[12] Y. Matsunaga and H. Yamamoto, "A coding theorem for lossy data compression by LDPC codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2225–2229, Sep. 2003.

[13] E. Martinian and M. J. Wainwright, "Low-density codes achieve the rate-distortion bound," in *Proc. Data Compression Conf.*, Snowbird, UT, 2006, pp. 153–162.

[14] S. Miyake, "Lossy data compression over Z_q by LDPC code," in *Proc. 2006 IEEE Int. Symp. Information Theory*, Seattle, WA, Jul. 2006, pp. 813–816.

[15] A. Gupta and S. Verdú, "Nonlinear sparse-graph codes for lossy compression," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 1961–1975, May 2009.

[16] E. Maneva and M. J. Wainwright, "Lossy source encoding via message-passing and decimation over generalized codewords of LDGM codes," in *Proc. 2005 IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sep. 2005, pp. 1493–1497.

[17] S. Ciliberti, K. Mezard, and R. Zecchina, "Message passing algorithms for non-linear nodes and data compression," *ComPlexUs*, vol. 3, pp. 58–65, Aug. 2006.

[18] A. Dembo and I. Kontoyiannis, "Source coding, large deviations and approximate pattern matching," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1590–1615, Jun. 2002.

[19] E. H. Yang and J. C. Kieffer, "On the performance of data compression algorithms based upon string matching," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 47–65, Jan. 1998.

[20] S. Shamai (Shitz) and S. Verdú, "The empirical distribution of good codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 836–846, May 1997.

Ankit Gupta (S'05–M'09) received the B.Tech. degree in 2003 from the Indian Institute of Technology, Delhi, India, and the M.A. and Ph.D. degrees in 2006 and 2009, respectively, from Princeton University, Princeton, NJ, all in electrical engineering.

He is a Senior Research Engineer with Samsung Telecommunications America.

Sergio Verdú (S'80–M'84–SM'88–F'93) received the Telecommunications Engineering degree from the Universitat Politècnica de Barcelona, Barcelona, Spain, in 1980 and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1984.

Since 1984 he has been a member of the faculty of Princeton University, Princeton, NJ, where he is the Eugene Higgins Professor of Electrical Engineering.

Dr. Verdú is the recipient of the 2007 Claude E. Shannon Award and the 2008 IEEE Richard W. Hamming Medal. He is a member of the National Academy of Engineering and was awarded a Doctorate Honoris Causa from the Universitat Politècnica de Catalunya in 2005. He is a recipient of several paper awards from the IEEE: the 1992 Donald Fink Paper Award, the 1998 Information Theory Outstanding Paper Award, an Information Theory Golden Jubilee Paper Award, the 2002 Leonard Abraham Prize Award, the 2006 Joint Communications/Information Theory Paper Award, and the 2009 Stephen O. Rice Prize from the IEEE Communications Society. He has also received paper awards from the Japanese Telecommunications Advancement Foundation and from Eurasip. In 1998, Cambridge University Press published his book *Multiuser Detection,* for which he received the 2000 Frederick E. Terman Award from the American Society for Engineering Education. He served as President of the IEEE Information Theory Society in 1997. He is currently the Editor-in-Chief of the *Foundations and Trends in Communications and Information Theory*.