34th Asilomar Conference
Signals, Systems, Computers
Oct. 29-Nov. 1, 2000
Asilomar, CA, USA

Illustration of the Duality Between Channel Coding and Rate Distortion with Side Information

Jonathan K. Su
MIT Lincoln Laboratory
Lexington, MA, USA
su@ll.mit.edu/su@LNT.de

Joachim J. Eggers
Telecomm. Laboratory
Univ. Erlangen-Nuremberg
Erlangen, Germany
eggers@LNT.de

Bernd Girod
Information Systems Lab
Stanford University
Stanford, CA, USA
girod@ee.stanford.edu

Abstract

Digital watermarking can be viewed as channel coding with side information at the encoder *(CC-SI); the original data to be watermarked is known to the encoder but not the decoder. Likewise, distributed source coding is* rate distortion with side information at the decoder *(RD-SI); a noisy observation of the source data to be compressed is available to the decoder but not the encoder. For a Gaussian channel or source, CC-SI and RD-SI are shown to be information-theoretic duals. Ideal coding schemes are presented, and the schemes are interpreted geometrically to highlight dual elements and quantities.*

1. Introduction

The duality between *channel coding* (CC) for the Gaussian channel and *rate distortion* (RD) for a Gaussian source has been known for years [5]. Recently, interest has been renewed in two similar scenarios: *channel coding with side information at the encoder* (CC-SI) and *rate distortion with side information at the decoder* (RD-SI). CC-SI relates directly to digital watermarking or data hiding [1, 3, 6], and RD-SI to distributed source coding [9].

The side-information duality was demonstrated in [3] by using examples for discrete memoryless channels and sources [7, 8, 10], but it has not been made explicit for the Gaussian case. That is the goal of this paper. Due to space constraints, derivations are omitted; they appear in [12]. Some of these duality concepts have been previously discovered [2] but not yet been published.

1.1. Channel Coding with Side Information

The Gaussian CC-SI scenario is shown in the top diagram in Fig. 1. In n channel uses, the encoder attempts to communicate a letter m chosen from a finite alphabet. The channel consists of two mutually independent, AWGN sources: the *state* $\mathbf{S} \sim \mathcal{N}(\mathbf{0}, QI)$ and *noise* $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, NI)$, where I is the $n \times n$ identity matrix. The encoder has com-

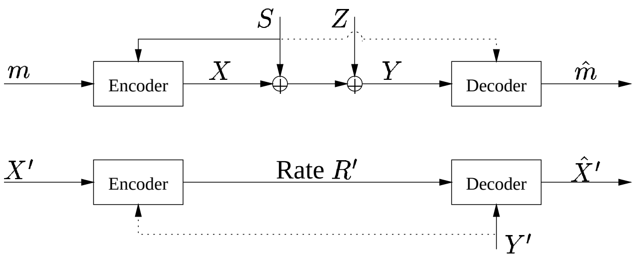

Figure 1. Basic scenarios for CC-SI (top) and RD-SI (bottom)

plete knowledge of the state $\mathbf{S}$ and transmits a length-n signal $\mathbf{X}$ with average power constraint $(1/n)\sum_{k=1}^{n} X^2(k) \leq P$. The decoder receives $\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{Z}$ but does *not* observe $\mathbf{S}$, and it decodes the received message $\hat{m}$. Let C_{enc} denote the capacity of this channel.

If $\mathbf{S}$ is also available at the decoder (dotted line in top diagram in Fig. 1), the decoder can just subtract $\mathbf{S}$ from $\mathbf{Y}$, and the channel capacity is

$$C_{\text{both}} = \tfrac{1}{2}\log_2(1 + P/N). \tag{1}$$

Clearly, $C_{\text{enc}} \leq C_{\text{both}}$. However, Costa [4] proved the remarkable and surprising result that $C_{\text{enc}} = C_{\text{both}}$: *It is possible to communicate at the same rate as when the side information* $\mathbf{S}$ *is known to both the encoder* ***and decoder***.[1]

In blind digital watermarking or data hiding, the state $\mathbf{S}$ represents the original, unwatermarked data, m the hidden information, and $\mathbf{X}$ the signal embedded in $\mathbf{S}$ to convey m. Then $\mathbf{X} + \mathbf{S}$ is the watermarked data, $\mathbf{Z}$ an attack,[2] and $\mathbf{Y}$ the received data, from which $\hat{m}$ is decoded. Costa's result means that, theoretically, the original data, unknown to the decoder, does not impair communication at all.

1.2. Rate Distortion with Side Information

The RD-SI scenario appears in the bottom diagram of Fig. 1. A *source* produces n realizations to form a sequence $\mathbf{X}' \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$. The encoder and decoder communicate

[1] Equality does not necessarily hold for non-Gaussian channels.
[2] A theoretically optimum attack is investigated in [11].

without error at a rate of R' bits per source symbol. The decoder also has an *observation* $\mathbf{Y}' = a(\mathbf{X}' + \mathbf{Z}')$, where $\mathbf{X}'$ and $\mathbf{Z}'$ are independent, $\mathbf{Z}' \sim \mathcal{N}(\mathbf{0}, N'I)$, and $a > 0$ is known. Then the decoder computes an approximation $\hat{\mathbf{X}}'$ of the source sequence $\mathbf{X}'$. We wish to communicate at the lowest possible rate R' such that the average squared-error distortion $(1/n)\sum_{k=1}^{n} \mathrm{E}[(X'(k) - \hat{X}'(k))^2]$ does not exceed D. Denote this lower bound by $R_{\text{dec}}(D)$.

If the encoder also observes $\mathbf{Y}'$ (dotted line in bottom diagram of Fig. 1), then the encoder and decoder can each compute $\boldsymbol{\mu} = \mathrm{E}[\mathbf{X}'|\mathbf{Y}']$. It then suffices to communicate $\mathbf{X}'_{\boldsymbol{\mu}} = \mathbf{X}' - \boldsymbol{\mu}$ with distortion D. Since $\mathbf{X}'_{\boldsymbol{\mu}} \sim \mathcal{N}\left(\mathbf{0}, \frac{N'\sigma^2}{\sigma^2+N'}\right)$, the rate-distortion function $R_{\text{both}}(D)$ is

$$R_{\text{both}}(D) = \begin{cases} \frac{1}{2}\log_2 \frac{N'\sigma^2}{(\sigma^2+N')D}, & 0 \le D \le \frac{N'\sigma^2}{\sigma^2+N'}; \\ 0, & D > \frac{N'\sigma^2}{\sigma^2+N'}. \end{cases} \quad (2)$$

It is clear that $R_{\text{dec}}(D) \ge R_{\text{both}}(D)$. Wyner and Ziv [13, 14] showed that $R_{\text{dec}}(D) = R_{\text{both}}(D)$: *It is possible to communicate at the same rate as when the side information $\mathbf{Y}'$ is known to both the decoder **and encoder**.*[3]

As an example, consider combining images from a space-based telescope and ground-based observatory. Both simultaneously image the same region of space. $\mathbf{X}'$ corresponds to the image at the telescope, which encounters no atmospheric interference, and $\mathbf{Y}'$ to the less-accurate image at the observatory. The telescope transmits information at rate R' to the observatory, which computes the reconstructed image $\hat{\mathbf{X}}'$. Wyner's and Ziv's result means that distributed source coding can, theoretically, operate at a lower rate than conventional lossy source coding (which does not exploit $\mathbf{Y}'$) without sacrificing image quality.

2. CC-SI Interpretation

In CC-SI with discrete memoryless random variables (RVs), the capacity [7, 8] is $C_{\text{enc}} = \max_{p(u,x|s)}\{I(U;Y) - I(U;S)\}$. Maximization is performed over all $p(y,u,x,s)$ of the form $p(y,u,x,s) = p(s)p(u,x|s)p(y|x,s)$, and U is a finite-alphabet auxiliary RV.

Costa [4] applied this result to the Gaussian case and showed that capacity is achieved when $U = U^* = X + \alpha^* S$, where $X \sim \mathcal{N}(0,P)$, $S \sim \mathcal{N}(0,Q)$, and $\alpha^* = P/(P+N)$. Then

$$I(U^*;Y) = \frac{1}{2}\log_2 \frac{P+Q+N}{\frac{P}{P+\alpha^{*2}Q}(1-\alpha^*)^2Q+N}, \quad (3)$$

$$I(U^*;S) = \frac{1}{2}\log_2 \frac{P+\alpha^{*2}Q}{P}, \quad (4)$$

and the rate $C^* = I(U^*;Y) - I(U^*;S) = C_{\text{both}}$ in (1).

Costa's coding scheme is summarized below.

Codebook The codebook $\mathcal{U}$ contains about $2^{n(I(U^*;Y)-\varepsilon)}$ *codevectors* (CVs) $\mathbf{U}$, each drawn $\mathcal{N}(\mathbf{0}, (P+\alpha^{*2}Q)I)$. The CVs are randomly and equiprobably assigned to $2^{n(C^*-2\varepsilon)}$ distinct *bins*, denoted by $\mathcal{U}_m$, where m is the *bin index*. Each bin $\mathcal{U}_m$ contains about $2^{n(I(U^*;S)+\varepsilon)}$ CVs.

Encoding Given m_0 and $\mathbf{S}_0$, search bin $\mathcal{U}_{m_0}$ for the CV $\mathbf{U}_0$ that satisfies

$$\mathbf{U}_0 = \arg\min_{\mathbf{U}\in\mathcal{U}_{m_0}} \|\mathbf{U} - \alpha^*\mathbf{S}_0\|. \quad (5)$$

Compute $\mathbf{X}_0 = \mathbf{U}_0 - \alpha^*\mathbf{S}_0$, and transmit $\mathbf{X}_0$ over the channel.

MAP Decoding Given $\mathbf{Y}_0$, search the entire codebook $\mathcal{U}$ for the CV $\hat{\mathbf{U}}$ that satisfies

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}\in\mathcal{U}} \|\mathbf{Y}_0 - c^*\mathbf{U}\|, \quad (6)$$

where $c^* = (P+\alpha^*Q)/(P+\alpha^{*2}Q)$. Return the decoded message $\hat{m}$, the bin index of the bin $\mathcal{U}_{\hat{m}} \ni \hat{\mathbf{U}}$.

[3]Like CC-SI, equality does not always hold for non-Gaussian sources.

2.1. CC-SI: Single Codevector

We treat random vectors as points in $\mathbb{R}^n$, and for Gaussian random vectors, orthogonality implies independence. The left-hand diagram in Fig. 2 depicts the vector relationships (as if $n=3$) for a single CV $\mathbf{U}$, assumed to belong to bin $\mathcal{U}_m$. According to Costa's construction, $\mathbf{U} = \mathbf{X} + \alpha^*\mathbf{S}$. The small hemisphere depicts a *bin-encoding sphere* of radius $\sqrt{nP}$ centered at $\mathbf{U}$. This represents (5): If $m_0 = m$ and $\alpha^*\mathbf{S}_0$ lies within distance $\sqrt{nP}$ of $\mathbf{U}$, then the encoder chooses $\mathbf{U}_0 = \mathbf{U}$.

The figure also shows that $\mathbf{X}+\mathbf{S} = c^*\mathbf{U}+\mathbf{V}$, with $\mathbf{U} \perp \mathbf{V}$ and $\sigma_V^{*2} = \frac{P}{P+\alpha^{*2}Q}(1-\alpha^*)^2Q$. Hence, by transmitting $\mathbf{X}_0 = \mathbf{U}_0 - \alpha^*\mathbf{S}_0$, the encoder "steers" the state $\mathbf{S}_0$ towards $c^*\mathbf{U}_0$. Also note that encoding is like quantizing $\alpha^*\mathbf{S}_0$ to the nearest CV in $\mathcal{U}_{m_0}$ and transmitting the "quantization error" $\mathbf{X}_0$.

The noise $\mathbf{Z}$ is independent of $\mathbf{X}$, $\mathbf{S}$, $\mathbf{U}$, and $\mathbf{V}$, so $\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{Z} = c^*\mathbf{U} + \mathbf{V} + \mathbf{Z}$. Thus, the received vector $\mathbf{Y}$ lies at a distance of about $\sqrt{n(\sigma_V^{*2}+N)}$ from $c^*\mathbf{U}$. The large hemisphere depicts a *decoding sphere* with this radius and centered at $c^*\mathbf{U}$; it represents (6): Any received vector $\mathbf{Y}_0$ in this sphere is decoded to CV $c^*\hat{\mathbf{U}} = c^*\mathbf{U}$.

Depending on P, Q, and N, the bin-encoding and decoding spheres may intersect. Fig. 2 shows them as non-intersecting for clarity.

2.2. CC-SI: Entire Codebook

The left-hand diagram in Fig. 3 presents an abstract illustration of CC-SI in $\mathbb{R}^n$; the angles cannot be taken literally. The thick concentric circles represent the surfaces of (hyper)spheres with radii $\sqrt{n(P+\alpha^{*2}Q)}$ (inner) and $\sqrt{nc^{*2}(P+\alpha^{*2}Q)}$ (outer). The CVs $\mathbf{U}$ lie near the inner surface and are shown as dots, triangles, and squares; like shapes belong to the same bin $\mathcal{U}_m$. The scaled CVs $c^*\mathbf{U}$ lie near the outer surface. Thus, scaling spreads out the CVs.

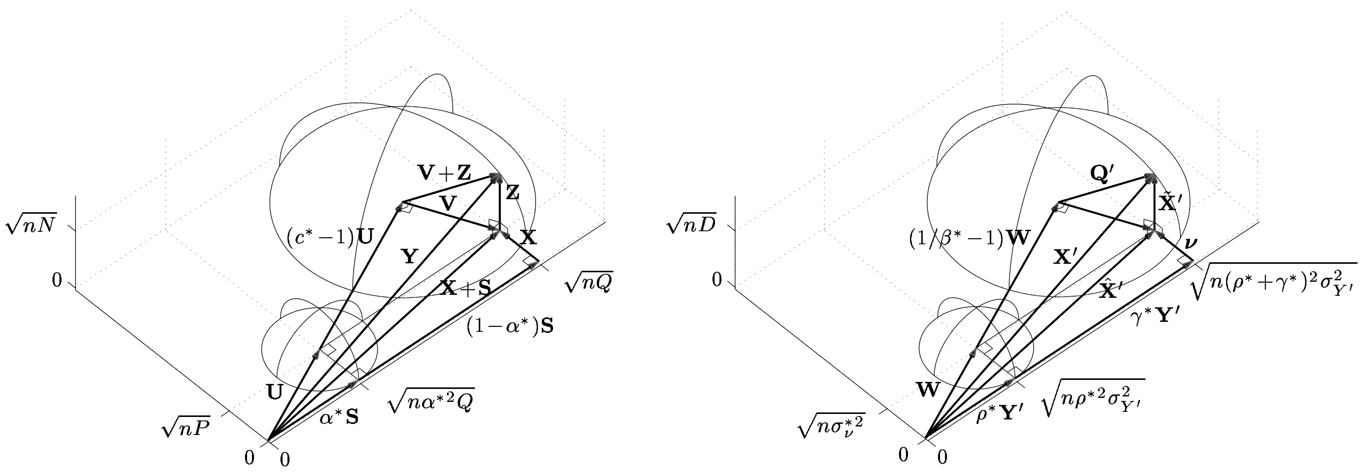

Figure 2. Vector relationships in CC-SI for a single CV U (left) and RD-SI for a single QV W (right)

Encoding has a *sphere-covering* interpretation. The dashed circles on the inner surface depict bin-encoding spheres for a single bin (containing "dots") and are virtually non-intersecting. Also, $\alpha^*\mathbf{S}_0$ lies near the surface of a sphere of radius $\sqrt{n\alpha^{*2}Q}$. Fulfilling the power constraint can be viewed as covering the hull between spheres of radii $\sqrt{n(P+\alpha^{*2}Q)}$ and $\sqrt{n\alpha^{*2}Q}$ with bin-encoding spheres of radius $\sqrt{nP}$. The required number of bin-encoding spheres is lower-bounded by the ratio of the hull volume to the bin-encoding sphere volume. For large n, the ratio is

$$\frac{A_n\left(n(P+\alpha^{*2}Q)\right)^{n/2}}{A_n\left(nP\right)^{n/2}} = 2^{n(I(U^*;S)+\varepsilon)}, \qquad (7)$$

where A_n is a constant that depends on n [5].

The bin-encoding spheres for different bins intersect, as shown by the dotted circles on the inner surface. Since the encoder knows m_0, it never searches the wrong bin.

Decoding has a *sphere-packing* interpretation. The received vector $\mathbf{Y}_0$ lies near the surface of a sphere with radius $\sqrt{n(P+Q+N)}$. The dashed circles on the outer surface in the figure depict decoding spheres, each with radius $\sqrt{n(\sigma_V^{*2}+N)}$, for all scaled CVs. For reliable decoding, the decoding spheres should not intersect, so the number of reliably decodable CVs is upper-bounded by the number of decoding spheres that can be packed into a sphere of radius $\sqrt{n(P+Q+N)}$. For large n, the bound is

$$\frac{A_n\left(n(P+Q+N)\right)^{n/2}}{A_n\left(n\left(\sigma_V^{*2}+N\right)\right)^{n/2}} = 2^{n(I(U^*;Y)-\varepsilon)}. \qquad (8)$$

Although $2^{n(I(U^*;Y)-\varepsilon)}$ CVs can be reliably decoded, all $2^{n(I(U^*;S)+\varepsilon)}$ CVs in a bin $\mathcal{U}_m$ convey the same message m. Hence, the number of different messages that can be communicated is $2^{n(I(U^*;Y)-\varepsilon)} \div 2^{n(I(U^*;S)+\varepsilon)} = 2^{n(C^*-2\varepsilon)}$.

3. RD-SI Interpretation

For RD-SI with discrete memoryless RVs and distortion measure $d(\cdot,\cdot)$, $R_{\text{dec}}(D) = \min_{p(w|x'),f}\{I(X';W) - I(Y';W)\}$ [5, 13, 14]. A double minimization is conducted over all $p(x',y',w)$ of the form $p(x',y',w) = p(x',y')p(w|x')$ and all functions $\hat{x}' = f(w,y')$ such that $\sum_{x',w,y'} p(x',y')p(w|x')d(x',f(w,y')) \leq D$.

For the Gaussian case, we have recently derived $R_{\text{dec}}(D)$ in another manner [12], which shows that $W = W^* \sim \mathcal{N}(0,\sigma_W^{*2})$, where $\sigma_W^{*2} = \left(\sigma^2 - \frac{\sigma^2+N'}{N'}D\right)\left(\frac{N'-D}{N'}\right)$. This choice of W yields

$$I(X';W^*) = \tfrac{1}{2}\log_2\frac{(N'-D)\sigma^2}{N'D}, \qquad (9)$$

$$I(Y';W^*) = \tfrac{1}{2}\log_2\frac{(N'-D)(\sigma^2+N')}{(N')^2}, \qquad (10)$$

so the rate $R^* = I(X';W^*) – I(Y';W^*) = R_{\text{both}}(D)$ in (2).

The RD-SI coding scheme is described below.

Codebook The codebook $\mathcal{W}$ contains about $2^{n(I(X';W^*)+\varepsilon)}$ *quantization vectors* (QVs) $\mathbf{W}$, each drawn $\mathcal{N}(\mathbf{0},\sigma_W^{*2}I)$. The QVs are randomly and equiprobably assigned to $2^{n(R^*+2\varepsilon)}$ distinct *bins*, denoted by $\mathcal{W}_m$, where m is the *bin index*. Each bin $\mathcal{W}_m$ contains about $2^{n(I(Y';W^*)-\varepsilon)}$ QVs.

Encoding Given $\mathbf{X}_0'$, search the entire codebook $\mathcal{W}$ for the QV $\mathbf{W}_0$ that satisfies

$$\mathbf{W}_0 = \arg\min_{\mathbf{W}\in\mathcal{W}} \|\mathbf{X}_0' - (1/\beta^*)\mathbf{W}\|, \qquad (11)$$

where $\beta^* = (N'-D)/N'$. Transmit the bin index m_0 of the bin $\mathcal{W}_{m_0} \ni \mathbf{W}_0$.

MAP Decoding Given m_0 and $\mathbf{Y}_0'$, search bin $\mathcal{W}_{m_0}$ for the QV $\hat{\mathbf{W}}$ that satisfies

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}\in\mathcal{W}_{m_0}} \|\mathbf{W} - \rho^*\mathbf{Y}_0'\|, \qquad (12)$$

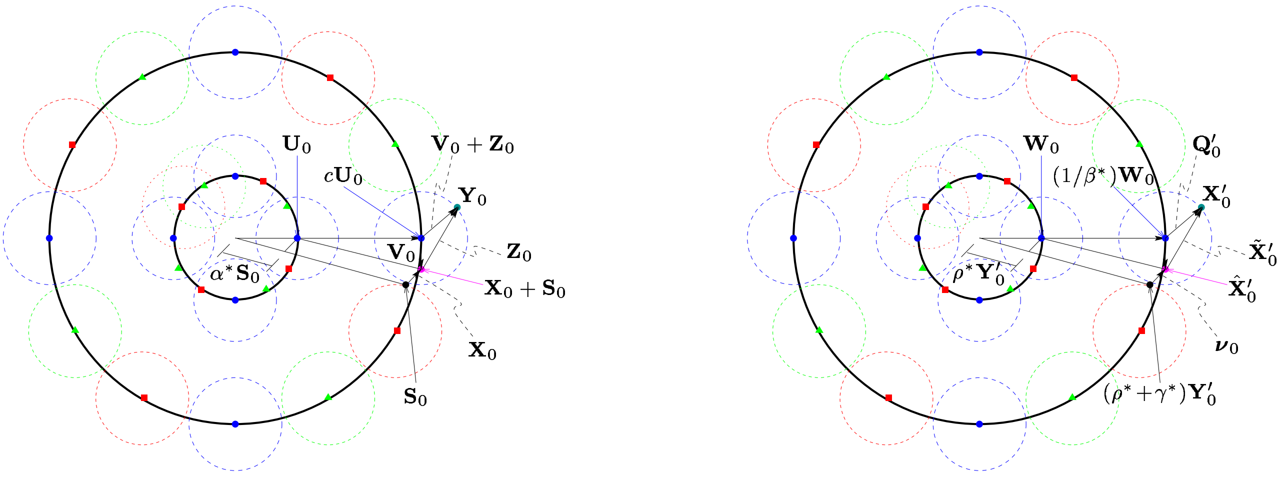

Figure 3. Abstract illustrations of CC-SI (left) and RD-SI (right) coding schemes

with $\rho^* = \frac{1}{a}\left(\frac{\sigma^2}{\sigma^2+N'} - \frac{D}{N'}\right)$. Return the *reconstruction vector* $\hat{\mathbf{X}}'_0 = f(\hat{\mathbf{W}}, \mathbf{Y}'_0) = \hat{\mathbf{W}} + \gamma^*\mathbf{Y}'_0$, where $\gamma^* = D/aN$.

3.1. RD-SI: Single Quantization Vector

The right-hand side of Fig. 2 depicts the vector relationships for a single QV $\mathbf{W}$, assumed to belong to bin $\mathcal{W}_m$. The encoder quantizes $\mathbf{X}'$ to the scaled QV $(1/\beta^*)\mathbf{W}$, and the *quantization-noise vector* $\mathbf{Q}' = \mathbf{X}' - (1/\beta^*)\mathbf{W}$ is independent of $\mathbf{W}$. The large hemisphere depicts a *quantization sphere* centered at $(1/\beta^*)\mathbf{W}$ with radius $\sqrt{n\sigma_{Q'}^{*2}}$, where $\sigma_{Q'}^{*2} = N'D/(N'-D)$. This represents (11): Any source vector $\mathbf{X}'_0$ in the sphere is quantized to $(1/\beta^*)\mathbf{W}$.

The figure also shows that $\mathbf{W} = \boldsymbol{\nu} + \rho^*\mathbf{Y}'$, where $\boldsymbol{\nu} \perp \mathbf{Y}'$. Thus, $\rho^*\mathbf{Y}'$ lies at a distance of about $\sqrt{n\sigma_\nu^{*2}}$ from $\mathbf{W}$, where $\sigma_\nu^{*2} = \frac{N'\sigma^2}{\sigma^2+N'} - D$. The small hemisphere depicts a *bin-decoding sphere* centered at $\mathbf{W}$ and having this radius. It reflects (12): If $m_0 = m$ and $\rho^*\mathbf{Y}'_0$ lies within distance $\sqrt{n\sigma_\nu^{*2}}$ of $\mathbf{W}$, the decoder selects $\hat{\mathbf{W}} = \mathbf{W}$.

The quantization and bin-decoding spheres may intersect, depending on a, σ^2, N', and D. Fig. 2 shows a non-intersecting case for clarity.

With $\hat{\mathbf{W}} = \mathbf{W}$, the reconstruction vector $\hat{\mathbf{X}}' = \mathbf{W} + \gamma^*\mathbf{Y}' = \boldsymbol{\nu} + (\rho^*+\gamma^*)\mathbf{Y}'$; $\hat{\mathbf{X}}'$ is the minimum mean-squared error estimate of $\mathbf{X}'$ given $\mathbf{W}$ and $\mathbf{Y}'$. The *reconstruction-error vector* $\tilde{\mathbf{X}}' = \mathbf{X}' - \hat{\mathbf{X}}'$ is independent of $\mathbf{W}$ and $\mathbf{Y}'$.

3.2. RD-SI: Entire Codebook

The right-hand diagram in Fig. 3 illustrates RD-SI abstractly. The thick concentric circles depict the surfaces of spheres of radii $\sqrt{n\sigma_W^{*2}}$ (inner) and $\sqrt{n\sigma_W^{*2}/\beta^{*2}}$ (outer). The QVs are shown as dots, triangles, and squares near the inner surface; their scaled versions lie near the outer surface. Like shapes belong to the same bin $\mathcal{W}_m$.

Encoding has a *sphere-covering* interpretation. The dashed circles on the outer surface show the quantization spheres, each with radius $\sqrt{n\sigma_{Q'}^{*2}}$, for all scaled QVs $(1/\beta^*)\mathbf{W}$; the spheres are virtually non-intersecting. The source vector $\mathbf{X}'_0$ lies near the surface of a sphere with radius $\sqrt{n\sigma^2}$. This sphere should be covered by the quantization spheres; the required number of quantization spheres is then lower-bounded by

$$\frac{A_n\left(n\sigma^2\right)^{n/2}}{A_n\left(n\sigma_{Q'}^{*2}\right)^{n/2}} = 2^{n(I(X';W^*)+\varepsilon)}. \qquad (13)$$

Decoding has a *sphere-packing* interpretation. The QVs in each bin lie near the inner surface of radius $\sqrt{n\sigma_W^{*2}}$. The dashed circles on the inner surface depict the bin-decoding spheres for a single bin (containing "dots"). The spheres have radii $\sqrt{n\sigma_\nu^{*2}}$ and should not intersect to ensure reliable decoding. The number of reliably decodable QVs is upper-bounded by the number of bin-decoding spheres that can be packed into a sphere of radius $\sqrt{n\sigma_W^{*2}}$, so the bound is

$$\frac{A_n\left(n\sigma_W^{*2}\right)^{n/2}}{A_n\left(n\sigma_\nu^{*2}\right)^{n/2}} = 2^{n(I(Y';W^*)-\varepsilon)}. \qquad (14)$$

The dotted circles on the inner surface show that the bin-decoding spheres for different bins intersect. Because the decoder is given m_0, it never uses the wrong bin.

Finally, $\mathcal{W}$ contains $2^{n(I(X';W^*)+\varepsilon)}$ QVs, but the decoder searches only the $2^{n(I(Y';W^*)-\varepsilon)}$ QVs in bin $\mathcal{W}_{m_0}$. Thus, the required number of bins is $2^{n(I(X';W^*)+\varepsilon)} \div 2^{n(I(Y';W^*)-\varepsilon)} = 2^{n(R^*+2\varepsilon)}$.

4. Duality of CC-SI and RD-SI

The preceding discussion shows that CC-SI and RD-SI have many correspondences. Encoding (decoding) in one

CC-SI	RD-SI
α^*	ρ^*
c^*	$1/\beta^*$
$\mathbf{S}, Q$	$\mathbf{Y}', a_\star^2(\sigma^2 + N')$
$\mathbf{U}, P + \alpha^{*2}Q$	$\mathbf{W}, \sigma_W^{*2}$
$\mathbf{X}, P$	$\boldsymbol{\nu}, \sigma_\nu^{*2}$
$\mathbf{Y}, P + Q + N$	$\mathbf{X}', \sigma^2$
$\mathbf{Z}, N$	$\tilde{\mathbf{X}}', D$
$\mathbf{X} + \mathbf{S}, P + Q$	$\hat{\mathbf{X}}', \sigma^2 - D$
$\mathbf{V} + \mathbf{Z}, \sigma_V^{*2} + N$	$\mathbf{Q}', \sigma_{Q'}^{*2}$

Table 1. Dual elements in CC-SI and RD-SI

scenario is analogous to decoding (encoding) in the other. It is evident why binning is necessary. In CC-SI, the CVs in each bin are close enough together to ensure that, for any m_0 and $\mathbf{S}_0$, the encoder will likely find a CV in bin $\mathcal{U}_{m_0}$ that is close enough to $\alpha^*\mathbf{S}_0$ to satisfy the power constraint. In RD-SI, the encoder does not observe $\mathbf{Y}_0'$ but knows that $\mathbf{Y}_0' = a(\mathbf{X}_0' + \mathbf{Z}_0')$. The QVs in each bin are far enough apart so that, given m_0 and $\mathbf{Y}_0'$, the decoder will likely choose the correct QV from bin $\mathcal{W}_{m_0}$.

In CC-SI, $\mathcal{U}$ can thus be viewed as a channel code with $2^{nI(U^*;Y)}$ CVs that is partitioned into $2^{nI(U^*;S)}$ source codes (bins) $\mathcal{U}_m$. In RD-SI, $\mathcal{W}$ is a source code with $2^{nI(X';W^*)}$ QVs that is partitioned into $2^{nI(Y';W^*)}$ channel codes (bins) $\mathcal{W}_m$. These ideas were recently proposed as guidelines for practical RD-SI and CC-SI schemes [3, 9].

From Figs. 2 and 3, we can identify corresponding CC-SI and RD-SI elements. However, RD-SI involves four parameters (a, σ^2, N', D) but CC-SI only three (P, Q, N); this subtle difference prevents CC-SI and RD-SI from always being exact duals. After CC-SI encoding, $\mathbf{X} + \mathbf{S}$, $\mathbf{X} \perp \mathbf{S}$; after RD-SI decoding, $\hat{\mathbf{X}}' = \boldsymbol{\nu} + (\rho^* + \gamma^*)\mathbf{Y}'$, $\boldsymbol{\nu} \perp \mathbf{Y}'$. For exact duality to hold, it is necessary that $\rho^* + \gamma^* = 1$; this equation is satisfied only for

$$a = a_\star = \sigma^2/(\sigma^2 + N'). \tag{15}$$

Table 1 lists the dual elements when (15) is satisfied.[4] CC-SI with P, Q, and N is the dual of RD-SI via

$$\begin{aligned} a_\star &= \tfrac{Q}{P+Q+N}, \qquad & \sigma^2 &= P + Q + N, \\ N' &= \tfrac{(P+Q+N)(P+N)}{Q}, \qquad & D &= N. \end{aligned} \tag{16}$$

Likewise, RD-SI with σ^2, N', and D (and $a = a_\star$) is the dual of CC-SI via

$$P = \tfrac{N'\sigma^2}{\sigma^2+N'} - D, \quad Q = \tfrac{\sigma^4}{\sigma^2+N'}, \quad N = D. \tag{17}$$

[4]The RD-SI noise $\mathbf{Z}'$ does not correspond directly to a CC-SI entity but forms part of the CC-SI state $\mathbf{S}$.

5. Generalization of Standard Cases

Finally, CC-SI and RD-SI generalize standard CC and RD. When $Q = 0$ or $N' \to \infty$ (no side information), the bins become singleton sets. In CC-SI, $\mathbf{S} \equiv \mathbf{0}$, $c^* = 1$, and $I(U^*; S) = 0$. The encoder transmits $\mathbf{X}_0 = \mathbf{U}_0$, the decoder uses minimum-distance decoding without scaling, and the standard sphere-packing argument applies [5]. In RD-SI, $\beta^* = 1$, $\gamma^* = 0$, and $I(Y'; W^*) = 0$. The encoder quantizes $\mathbf{X}'$ without scaling the QVs $\mathbf{W}$, the decoder returns $\hat{\mathbf{X}}' = \hat{\mathbf{W}} = \mathbf{W}_0$, and standard sphere covering applies.

Acknowledgment: We thank Thomas Cover and Mung Chiang for encouragement and stimulating discussion.

References

[1] B. Chen and G. Wornell. Preprocessed and postprocessed quantization index modulation methods for digital watermarking. *Proc. SPIE Security & Watermarking Multimedia Contents II*, vol. 3971, pp. 48–59, Jan. 2000.

[2] M. Chiang. *A random walk in the information systems: Undergraduate honors thesis*. Supervisors: T. M. Cover, S. Boyd. Stanford Univ., USA, 1999.

[3] J. Chou, S. S. Pradhan, and K. Ramchandran. On the duality between distributed source coding and data hiding. *33rd Asilomar Conf. Signals, Systems, Computers*, 1999.

[4] M. H. M. Costa. Writing on dirty paper. *IEEE Trans. Info. Thy.*, IT-29:439–441, May 1983.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.

[6] J. J. Eggers, J. K. Su, and B. Girod. A blind watermarking scheme based on structured codebooks. *Secure Images and Image Authentication, IEE Colloq.*, pp. 4/1–4/6, Apr. 2000.

[7] S. I. Gel'fand and M. S. Pinsker. Coding for channel with random parameters. *Problems of Control and Info. Thy.*, 9(1):19–31, 1980.

[8] C. Heegard and A. A. El Gamal. On the capacity of computer memory with defects. *IEEE Trans. Info. Thy.*, IT-29(5):731–739, Sep. 1983.

[9] S. S. Pradhan and K. Ramchandran. Distributed source coding using syndromes (DISCUS): Design and construction. *Data Compr. Conf.*, Mar. 1999.

[10] D. Slepian and J. K. Wolf. Noiseless encoding of correlated information sources. *IEEE Trans. Info. Thy.*, IT-19:471–480, Jul. 1973.

[11] J. K. Su, J. J. Eggers, and B. Girod. Analysis of digital watermarks subjected to optimum linear filtering and additive noise. *Signal Processing*, to appear Spring 2001.

[12] J. K. Su, J. J. Eggers, and B. Girod. Channel coding and rate distortion with side information: Geometric interpretation and illustration of duality. Submitted to *IEEE Trans. Info. Thy.*, May 2000.

[13] A. D. Wyner. The rate-distortion function for source coding with side information at the decoder-II: General sources. *Information and Control*, 38:60–80, 1978.

[14] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Info. Thy.*, IT-22(1):1–10, Jan. 1976.