A Random Coding Error Exponent for Joint Quantization and Watermarking of Gaussian Sources under Memoryless Gaussian Attacks*

Yadong Wang, Fady Alajaji and Tamás Linder
Department of Mathematics and Statistics
Queen's University, Kingston, ON, K7L 3N6, Canada
Email: {yadong, fady, linder}@mast.queensu.ca

***Abstract*— We study joint quantization and watermarking of a memoryless Gaussian source under memoryless additive Gaussian attacks in a private scenario. The achievable region involving the quantization and the watermarking rate pairs has been established by Karakos and Papamarcou (2003). In this paper, we refine the analysis of the watermarking decoding error probability for given achievable rate pairs by deriving a computable random coding lower bound to the error exponent. The random coding exponent is positive within almost the entire achievable region of Karakos and Papamarcou.**

I. Introduction

In a joint compression and embedding information-hiding model, the watermarker encodes a watermark and a covertext jointly to output a (compressed) stegotext. Denoting the *quantization rate* by R_Q and the *watermarking rate* by R_W, the main goal is to determine the achievable rate pairs (R_Q, R_W) under transparency and robustness constraints on the system (detailed definitions are given in Section II).

Karakos and Papamarcou [1] study the tradeoff between R_Q and R_W for Gaussian host data and additive memoryless Gaussian attacks in a private scenario (where the host data is available at the decoder). The main result of [1] is a coding theorem which establishes the achievable region for rate pairs (R_Q, R_W). Maor and Merhav [2], [3] study a similar tradeoff problem for discrete memoryless sources in a public scenario (where the host data is not available at the decoder). The work in [2] focuses on the attack-free problem, while [3] extends the model in [2] to include stationary memoryless discrete attacks. In both works, coding theorems are established in which a single-letter expression involving the maximum achievable watermarking rate, the compression rate and the distortion threshold are obtained. Yang and Sun [4] study a similar private joint compression/watermarking problem with abstract alphabets. Other related works include [5] and [6].

In this work, we focus on the problem introduced and investigated in [1], i.e., the joint quantization and watermarking of memoryless Gaussian sources under additive white Gaussian noise (AWGN) attacks in a private scenario. We refine the analysis of the probability of error in decoding the watermarks for any achievable rate pairs (R_Q, R_W). Using a random coding technique that incorporates Gallager's method [7], we obtain a computable random coding lower bound to the error exponent of watermark decoding. In a sense, our problem can be described as a joint source-channel coding problem with side information at both the encoder and the decoder, and we study this problem from an error exponent viewpoint.

It is worth pointing out that Merhav [8] and Somekh-Baruch and Merhav [9] studied the error exponent performance for systems with finite alphabets in a private scenario. In [8], a single-letter expression of the Gallager random coding lower bound to the error exponent is obtained, while in [9], an asymptotic expression for the exact error exponent is derived. Note that the results of [8], [9] do not apply to the Gaussian Karakos-Papamarcou setup studied here, as they depend on the finiteness of the covertext and attack channel alphabets. Furthermore in [8], [9], different distortion constraints are imposed at the encoder.

Throughout, random variables, their realizations and alphabets are denoted by capital letters, lower case letters and calligraphic letters, respectively, e.g., X, x, and $\mathcal{X}$. Random vectors and their alphabets are denoted by capital letters and calligraphic letters superscripted by their lengths, respectively, e.g., X^n and $\mathcal{X}^n$, and the realizations are denoted by boldface lower case letters, e.g., $\mathbf{x} \triangleq (x_1, x_2, ..., x_n)$. $\mathbb{E}(X)$ denotes the expectation of X. All logarithms are in the natural base.

II. Problem Description

A general model for joint compression and watermarking in a private scenario is given in Fig. 1. Let $\{U_i\}_{i=1}^{\infty}$ be an independent and identically distributed (i.i.d.) sequence of zero mean Gaussian random variables with variance σ_u^2. Let $\mathcal{U} = \mathcal{Y} = \mathbb{R}$, and $d : \mathcal{U} \times \mathcal{Y} \to [0, \infty)$ be a single-letter distortion measure. For $\mathbf{u} \in \mathcal{U}^n$ and $\mathbf{y} \in \mathcal{Y}^n$, define $d(\mathbf{u}, \mathbf{y}) = \sum_{i=1}^n d(u_i, y_i)$. In this paper, we consider the squared distortion measure, i.e., $d(\mathbf{u}, \mathbf{y}) = \|\mathbf{u} - \mathbf{y}\|^2 = \sum_{i=1}^n (u_i - y_i)^2$. Let $A_{Z|Y}$ be an AWGN channel with input alphabet $\mathcal{Y}$, output alphabet $\mathcal{Z}$ ($\mathcal{Z} = \mathcal{Y} = \mathbb{R}$) so that $Z = Y + N$, where N is Gaussian with mean zero and variance D_A and is independent of Y.

Definition 1: An (R_Q, R_W, n) joint quantization and watermarking code consists of an encoder-decoder pair $(\varphi^{(n)}, \psi^{(n)})$:

$$\varphi^{(n)} : \mathcal{W} \times \mathcal{U}^n \to \mathcal{Y}^n, \qquad \psi^{(n)} : \mathcal{Z}^n \times \mathcal{U}^n \to \mathcal{W}, \tag{1}$$

where $\mathcal{W} = \{1, 2, \ldots, M_W\}$ is the watermark set and $M_W \triangleq \lceil e^{nR_W} \rceil$. Given $w \in \mathcal{W}$ and $\mathbf{u} \in \mathcal{U}^n$, the stegotext $\mathbf{y}$ takes values from a set $\mathbf{c}$ of $M_Q \triangleq \lceil e^{nR_Q} \rceil$ codevectors, i.e., $\mathbf{c} \triangleq \{\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(M_Q)\}$.

Definition 2: Given an (R_Q, R_W, n) code, the conditional probability of error in decoding a watermark index w is given

[*] This work was supported in part by NSERC of Canada.

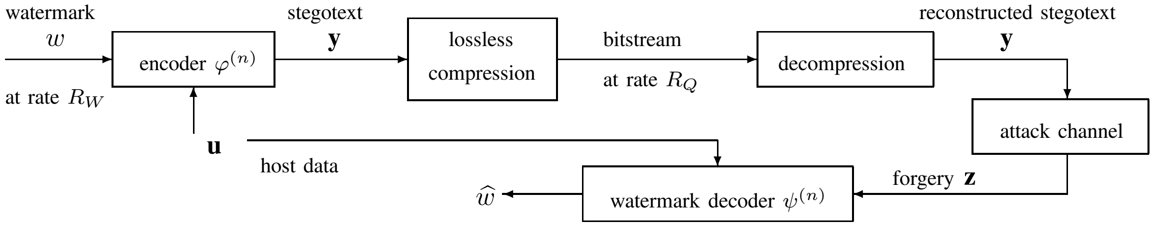

Fig. 1. A model for joint compression and watermarking in a private scenario.

by $P_{e,w}^{(n)} = \Pr\{\widehat{w} \neq w | w \text{ is embedded}\}$, where $\widehat{w}$ is the decoded watermark message. Furthermore, if we assume that all watermark indices are equiprobable, the average probability of decoding error is given by $P_e^{(n)} = \frac{1}{M_W}\sum_{w=1}^{M_W} P_{e,w}^{(n)}$.

Definition 3: Given an (R_Q, R_W, n) joint quantization and watermarking code, the average distortion between the host data and the stegotext is given by $D^{(n)} \triangleq \mathbb{E}\left[\frac{1}{n}d\Big(U^n, \varphi^{(n)}(W, U^n)\Big)\right]$.

Definition 4: Given $D_Q > 0$, the transparency and robustness conditions for a sequence of (R_Q, R_W, n) joint quantization and watermarking codes require that for any $\epsilon > 0$ and $\delta > 0$, $D^{(n)} \leq D_Q + \delta$ and $P_e^{(n)} \leq \epsilon$ for n sufficiently large.

Definition 5: A quadruple $(R_Q, R_W; D_Q, D_A)$ is said to be achievable if for any $\epsilon, \delta > 0$, there exists a sequence of (R_Q, R_W, n) joint quantization and watermarking codes such that $P_e^{(n)} \leq \epsilon$ and $D^{(n)} \leq D_Q + \delta$ for n sufficiently large. Given (D_Q, D_A), denote by $\mathcal{R}_{D_Q,D_A}$ the achievable region of all rate pairs (R_Q, R_W) such that $(R_Q, R_W; D_Q, D_A)$ is achievable.

The achievable rate region has been derived for memoryless Gaussian sources and memoryless Gaussian attacks. The main result is summarized in the following theorem.

Theorem 1: [1] The achievable rate region is given by

$$\begin{aligned}\mathcal{R}_{D_Q,D_A} = \Big\{(R_Q, R_W) : R_Q \geq \left[\frac{1}{2}\log\Big(\frac{\sigma_u^2}{D_Q}\Big)\right]^+, \\ R_W \leq \max_{\tau\in[\max\{1,\sigma_u^2/D_Q\},e^{2R_Q}]} \\ \min\Big[R_Q - \frac{1}{2}\log(\tau), \frac{1}{2}\log\Big(1+\frac{P_W(\tau)}{D_A}\Big)\Big]\Big\},\end{aligned}$$

where

$$P_W(\tau) \triangleq \frac{\tau(\sigma_u^2 + D_Q) - 2\sigma_u^2 + 2\sqrt{\sigma_u^2(\tau D_Q - \sigma_u^2)(\tau - 1)}}{\tau^2}.$$

III. Main Results

Given an i.i.d. Gaussian covertext $\{U_i\}_{i=1}^{\infty}$, a distortion threshold D_Q, and a Gaussian attack variance D_A, consider a rate pair $(R_Q, R_W) \in \mathcal{R}_{D_Q,D_A}$ (in this paper, we assume that $D_Q < \sigma_u^2$, which is a reasonable assumption in most practical applications). Our main result is the following theorem. A sketch of the proof is provided in Section V.

Theorem 2: Given $\delta > 0$, $s \geq 0$, $\rho \in [0, 1]$, $\beta^2 \in (D_Q, \sigma_u^2)$, and $\gamma \in (\sigma_u^2/D_Q, e^{2(R_Q - R_W)})$, there exists a sequence of (R_Q, R_W, n) joint quantization and watermarking codes such that

$$D^{(n)} \leq D_Q + \delta, \tag{2}$$

$$P_e^{(n)} \leq 4\exp\Big\{-n\big[\Lambda(\gamma, \beta, \rho, s) - o(1)\big]\Big\}, \tag{3}$$

for n sufficiently large, where $\Lambda(\gamma, \beta, \rho, s) \triangleq \min\big\{\Lambda_1(\gamma, \beta, \rho, s), \Lambda_2(\gamma, \beta)\big\}$, where

$$\begin{aligned}&\Lambda_1(\gamma, \beta, \rho, s) \\ &\quad= \frac{1}{2}\log\left(\frac{1 + 2s\beta^2(\gamma-1)D_Q + 2s(1+\rho)\sigma_u^2\theta}{\gamma}\right) \\ &\quad+ \frac{\rho}{2}\log\Big(\frac{1 + 2s\beta^2(\gamma-1)D_Q + \frac{(\gamma-1)D_Q}{(1+\rho)D_A}}{\gamma}\Big) - \rho R_W,\end{aligned} \tag{4}$$

with $\theta = \beta^2 - D_Q - 2s\beta^2(\gamma-1)D_Q^2$, and

$$\begin{aligned}&\Lambda_2(\gamma, \beta) = \\ &\min\left\{\frac{1}{2}\Big(\frac{\beta^2}{\sigma_u^2} - 1 - \log\frac{\beta^2}{\sigma_u^2}\Big), \frac{1}{2}\Big(\frac{\gamma D_Q}{\sigma_u^2} - 1 - \log\frac{\gamma D_Q}{\sigma_u^2}\Big)\right\},\end{aligned} \tag{5}$$

and $o(1) \to 0$ as $n \to \infty$.

The error exponent bound in (3) can be tightened by optimizing it with respect to γ, β, ρ and s. This yields the following random coding error exponent,

$$\begin{aligned}&E_R(R_Q, R_W; D_Q, D_A) \\ &\triangleq \sup_{\gamma\in(\frac{\sigma_u^2}{D_Q}, e^{2(R_Q-R_W)}),\, \beta^2\in(D_Q,\sigma_u^2),\, \rho\in[0,1],\, s\geq 0} \Lambda(\gamma, \beta, \rho, s).\end{aligned}$$

Remarks.

- From (5), it is clear that $\Lambda_2(\gamma, \beta)$ is always positive by the choice of $\gamma > \sigma_u^2/D_Q$ and $\beta^2 < \sigma_u^2$. The condition $\gamma < e^{2(R_Q - R_W)}$ is equivalent to $R_W < R_Q - \frac{1}{2}\log\gamma$.
- The term $\Lambda_1(\gamma, \beta, \rho, s)$ is similar to the random coding lower bound derived in [7, pp. 337–343] for AWGN channels. However, here we deal with a distortion constraint at the channel input instead of a power constraint. The term $\Lambda_2(\gamma, \beta)$ is somewhat similar to the reliability function for Gaussian sources with respect to the rate-distortion pair $(R_Q - R_W, D_Q)$ [10].
- $\Lambda(\gamma, \beta, \rho, s)$ is maximized over $s \geq 0$ by[1]

$$s^* = \frac{1 - 2abc + \sqrt{(1-2abc)^2 + \frac{4ac(\rho a + b)(2+\rho)}{1+\rho}}}{4ac(2+\rho)}, \tag{6}$$

where

$$a \triangleq \frac{1}{\beta^2(\gamma-1)D_Q + (1+\rho)\sigma_u^2(\beta^2 - D_Q)},$$

$$b \triangleq \frac{1}{\beta^2(\gamma-1)D_Q} + \frac{1}{(1+\rho)\beta^2 D_A},\ c \triangleq \sigma_u^2\beta^2(\gamma-1)D_Q^2.$$

[1] An analytical derivation of the other three optimizing parameters is not readily available. However the optimization can be carried numerically (e.g., see Section IV).

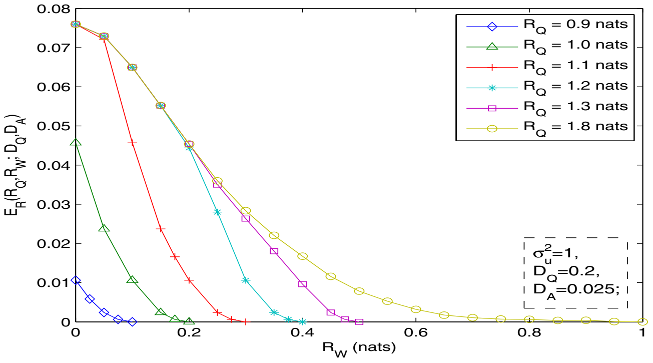

Fig. 2. $E_R(R_Q, R_W; D_Q, D_A)$ v.s. R_W for various values of R_Q.

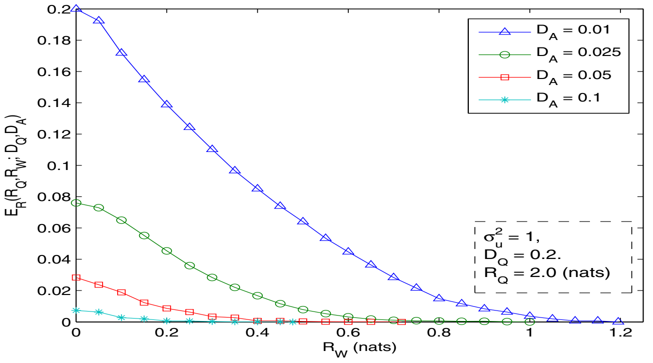

Fig. 3. $E_R(R_Q, R_W; D_Q, D_A)$ v.s. R_W for various values of D_A.

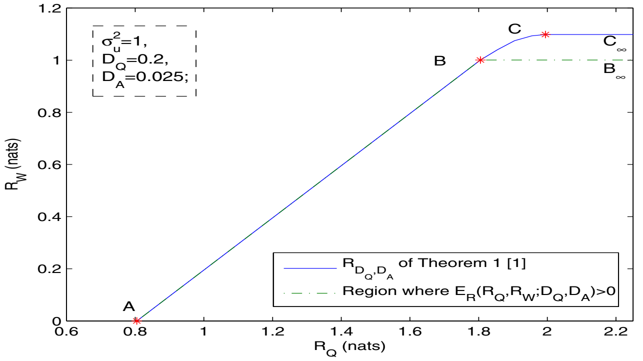

Fig. 4. $\mathcal{R}_{D_Q,D_A}$ of Theorem 1 and the region where the exponent $E_R(R_Q, R_W; D_Q, D_A) > 0$.

IV. Examples

We next present some numerical examples to illustrate the results of the previous section. Figs. 2 and 3 show the random coding error exponent v.s. the watermarking rate R_W for various quantization rates R_Q and channel noise levels D_A. Fig. 4 shows a typical region of rate pairs where the random coding error exponent is positive, in addition to the overall achievable region $\mathcal{R}_{D_Q,D_A}$ of Theorem 1 [1]. We note that $E_R(R_Q, R_W; D_Q, D_A) > 0$ nearly everywhere in $\mathcal{R}_{D_Q,D_A}$.

Here, point A is given by $R_Q = \frac{1}{2}\log(\frac{\sigma_u^2}{D_Q}), R_W = 0$; B is given by $R_Q = \frac{1}{2}\log(\frac{\sigma_u^2}{D_Q} + \frac{\sigma_u^2 - D_Q}{D_A}), R_W = \frac{1}{2}\log(1 + \frac{D_Q - D_Q^2/\sigma_u^2}{D_A})$; and C is given by $R_Q = \frac{1}{2}\log(1 + \frac{\sigma_u^2}{D_Q} + \frac{\sigma_u^2 + D_Q}{D_A}), R_W = \frac{1}{2}\log(1 + \frac{D_Q}{D_A})$ [1]. The figure shows that we can achieve all rates under the segments AB and BB_∞. In fact, for segment AB, i.e., $R_Q < \frac{1}{2}\log(\frac{\sigma_u^2}{D_Q} + \frac{\sigma_u^2 - D_Q}{D_A})$, given any $(R_Q, R_W) \in \mathcal{R}_{R_Q,R_W}$, we have that $\sup_{\rho,s,\beta^2} \Lambda_1(\gamma,\beta,\rho,s) > 0$ for any given γ. Since $\Lambda_2(\gamma,\beta) > 0$ implies that $R_W < R_Q - \frac{1}{2}\log\gamma$, we can approach segment AB by letting $\gamma \to \sigma_u^2/D_Q$. For segment BB_∞, if $R_W \geq \frac{1}{2}\log(1 + \frac{D_Q - D_Q^2/\sigma_u^2}{D_A})$, we will have $\Lambda_1(\gamma,\beta,\rho,s) \leq 0$ for any γ, β, ρ and s, which means $\Lambda(\gamma,\beta,\rho,s) \leq 0$. On the other hand, for $R_W < \frac{1}{2}\log(1 + \frac{D_Q - D_Q^2/\sigma_u^2}{D_A})$, we have $\sup_{\gamma,\beta^2,\rho,s} \Lambda_1(\gamma,\beta,\rho,s) > 0$. Since $\Lambda_2(\gamma,\beta)$ is always positive, we get a positive random coding error exponent. In this case, when we choose s as in (6), and letting $\gamma \to \sigma_u^2/D_Q$, $\beta \to \sigma_u^2$, and $\rho \to 0$, we can approach segment BB_∞ with $R_W \to \frac{1}{2}\log(1 + \frac{D_Q - D_Q^2/\sigma_u^2}{D_A})$.

V. Sketch of the Proof of Theorem 2

A. Code Construction

Given an i.i.d. Gaussian covertext $\{U_i\}_{i=1}^\infty$ with mean zero and variance σ_u^2, a distortion threshold D_Q, and a Gaussian attack variance D_A, assume that $(R_Q, R_W) \in \mathcal{R}_{D_Q,D_A}$. Let $M \triangleq \lceil e^{n(R_Q - R_W)} \rceil$, choose $\gamma \in (\frac{\sigma_u^2}{D_Q}, e^{2(R_Q - R_W)})$ and $\beta^2 \in (D_Q, \sigma_u^2)$. Now consider a code $\mathbf{c}$ described as follows.

Random Code Generation. The code $\mathbf{c}$ contains $M_W = \lceil e^{nR_W} \rceil$ "subcodes", where $\mathbf{c} \triangleq \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{M_W}\}$ is assigned a product density function $q(\mathbf{c}) = \prod_{i=1}^{M_W} q(\mathbf{c}_i)$. Each $\mathbf{c}_i$ contains M codewords, i.e., $\mathbf{c}_i = \{\mathbf{y}(i,1), \ldots, \mathbf{y}(i,M)\}$, where each codeword $\mathbf{y}(i,j)$ is i.i.d. drawn according to $q(\mathbf{y}) = \prod_{k=1}^n q(y_k)$. Here $q(y_k)$ is the Gaussian density with mean zero and variance $\sigma_y^2 = (\gamma - 1)D_Q$. Thus for each $\mathbf{c}_i$, we have $q(\mathbf{c}_i) = \prod_{j=1}^M q(\mathbf{y}(i,j))$. Given the watermark index w, the subcode $\mathbf{c}_w$ will be used for quantizing the covertext $\mathbf{u} \in \mathcal{U}^n$.

Encoding. Given a watermarking index w and a covertext $\mathbf{u}$, the encoder chooses the first codeword $\mathbf{y}(w,t)$ in $\mathbf{c}_w$ such that $\|\mathbf{u} - \mathbf{y}(w,t)\|^2 \leq nD_Q$, i.e.,

$$\begin{aligned} &\|\mathbf{u} - \mathbf{y}(w,i)\|^2 > nD_Q, \ i = 1, \ldots, t-1, \\ &\|\mathbf{u} - \mathbf{y}(w,t)\|^2 \leq nD_Q, \ t \leq M. \end{aligned} \tag{7}$$

Denote the chosen codevector by $\mathbf{y}(\mathbf{c}_w, \mathbf{u})$. If no such $\mathbf{y}(w,t)$ exists, an error is declared and $\mathbf{y}(\mathbf{c}_w, \mathbf{u}) = \mathbf{0}$ will be sent.

Decoding. The decoder has full knowledge of $\mathbf{u}$, and thus can generate all possible watermarked versions $\{\mathbf{y}(\mathbf{c}_i, \mathbf{u})\}_{i=1}^{M_W}$. Upon receiving the "forgery" $\mathbf{z} = \mathbf{y}(\mathbf{c}_w, \mathbf{u}) + \mathbf{v}$, the decoder compares it with all $\{\mathbf{y}(\mathbf{c}_i, \mathbf{u})\}_{i \in \mathcal{I}}$, where $\mathcal{I} \triangleq \{i \in \mathcal{W} : \|\mathbf{u} - \mathbf{y}(\mathbf{c}_i, \mathbf{u})\|^2 \leq nD_Q\}$, and chooses an output $\hat{w}$ using the maximum-likelihood decoding criterion: $\hat{w} = \arg\max_{i \in \mathcal{I}} f\big(\mathbf{z}|\mathbf{y}(\mathbf{c}_i, \mathbf{u})\big)$, where $f(\mathbf{z}|\mathbf{y}) = \prod_{j=1}^n f(z_j|y_j) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi D_A}} \exp\big(-\frac{(z_j - y_j)^2}{2D_A}\big)$.

B. Analysis for the Average Distortion

Define the event $\mathcal{E}_1(\mathbf{u}) \triangleq \{\mathbf{c}_w : \text{embedding } w \text{ into } \mathbf{u} \text{ with } \mathbf{c}_w \text{ is failed}\}$. Then, the distortion averaged over the random choice of $\mathbf{c}$ can be written as

$$\overline{D}^{(n)} \leq D_Q + \frac{1}{M}\sum_{w=1}^{M}\frac{1}{n}\int_{\mathcal{U}^n} p(\mathbf{u})\|\mathbf{u}\|^2 \int_{\mathcal{E}_1(\mathbf{u})} q(\mathbf{c}_w)\, d\mathbf{c}_w\, d\mathbf{u} \quad (8)$$

where $d\mathbf{c}_w = d\mathbf{y}(w,1)\ldots d\mathbf{y}(w,M)$. Define $\Phi_{D_Q}(\mathbf{y};\mathbf{u}) = 1$ for $d(\mathbf{y},\mathbf{u}) \leq nD_Q$; $\Phi_{D_Q}(\mathbf{y};\mathbf{u}) = 0$ for $d(\mathbf{y},\mathbf{u}) > nD_Q$, and let $P_{ex}(\mathbf{u}, Y^n) \triangleq \int_{\mathcal{Y}^n} q(\mathbf{y})\big[1 - \Phi_{D_Q}(\mathbf{y};\mathbf{u})\big]\, d\mathbf{y}$. Then we have $\int_{\mathcal{E}_1(\mathbf{u})} q(\mathbf{c}_w)\, d\mathbf{c}_w = \big[P_{ex}(\mathbf{u}, Y^n)\big]^M$.

We need the following lemma.

Lemma 1: [10] Let $\{X_i\}$ be an i.i.d. Gaussian source with distribution $X \sim \mathcal{N}(0,\sigma^2)$. For any $\Delta > 0$,
(a) if $a^2 = \sigma^2 + \Delta$, we have

$$\lim_{n\to\infty}\frac{1}{n}\log\Pr\Big(\frac{1}{n}\sum_{i=1}^{n}|X_i - a|^2 \leq \Delta\Big) = -\frac{1}{2}\log\frac{a^2}{\Delta};$$

(b) if $0 < \beta < \sigma$, then

$$\lim_{n\to\infty}\frac{1}{n}\log\Pr\Big(\frac{1}{n}\|X^n\|^2 < \beta^2\Big) = -\frac{1}{2}\Big(\frac{\beta^2}{\sigma^2} - 1 - \log\frac{\beta^2}{\sigma^2}\Big);$$

(c) if $\alpha > \sigma$, then

$$\lim_{n\to\infty}\frac{1}{n}\log\Pr\Big(\frac{1}{n}\|X^n\|^2 > \alpha^2\Big) = -\frac{1}{2}\Big(\frac{\alpha^2}{\sigma^2} - 1 - \log\frac{\alpha^2}{\sigma^2}\Big).$$

Let $\alpha^2 \triangleq \sigma_y^2 + D_Q = \gamma D_Q$ (recall that $\gamma > \sigma_u^2/D_Q$ implies $\alpha^2 > \sigma_u^2$), $\beta_1^2 \in (D_Q, \sigma_u^2)$, $\delta_0 \triangleq \sigma_u^2 - \beta_1^2$, and define $\mathcal{B}_n(\alpha,\beta_1) = \{\mathbf{u} \in \mathcal{U}^n : n\beta_1^2 \leq \|\mathbf{u}\|^2 \leq n\alpha^2\}$. Given any $0 < \epsilon_1^{(n)} < \epsilon_0^{(n)}$ such that $R_Q - R_W \geq \frac{1}{2}\log\gamma + \epsilon_0^{(n)}$ (here $\gamma < e^{2(R_Q - R_W)}$ guarantees the existence of such $\epsilon_0^{(n)}$) and applying Lemma 1, we obtain that for $n \geq N_1$ and $\mathbf{u} \in \mathcal{B}_n(\alpha,\beta_1)$ that $[P_{ex}(\mathbf{u},Y^n)]^M \leq \exp\big\{-\exp\big(n(\epsilon_0^{(n)} - \epsilon_1^{(n)})\big)\big\}$. For $\mathbf{u} \in (\mathcal{B}_n(\alpha,\beta_1))^c$, by Lemma 1, there exists $\epsilon_2^{(n)} > 0$ and $n \geq N_2$ such that

$$\begin{aligned}
&\frac{1}{n}\int_{(\mathcal{B}_n(\alpha,\beta_1))^c} p(\mathbf{u})\|\mathbf{u}\|^2\big[P_{ex}(\mathbf{u},Y^n)\big]^M\, d\mathbf{u} \\
&\leq \quad \delta_0 + \sigma_u^2\left(\exp\Big\{-n\Big[\frac{1}{2}\Big(\frac{\beta_1^2}{\sigma_u^2} - 1 - \log\frac{\beta_1^2}{\sigma_u^2}\Big) - \epsilon_2^{(n)}\Big]\Big\}\right. \\
&\qquad \left. + \exp\Big\{-n\Big[\frac{1}{2}\Big(\frac{\alpha^2}{\sigma_u^2} - 1 - \log\frac{\alpha^2}{\sigma_u^2}\Big) - \epsilon_2^{(n)}\Big]\Big\}\right). \quad (9)
\end{aligned}$$

Now choosing $n \geq \max\{N_1, N_2\}$ we obtain $\overline{D}^{(n)} \leq D_Q + \bar{\delta}^{(n)}$, where $\bar{\delta}^{(n)}$ can be made arbitrarily small by choosing $\sigma_u^2 - \beta_1^2$ sufficiently small and n sufficiently large.

C. Analysis for the Average Probability of Error

Given a random codebook $\mathbf{c} = \{\mathbf{c}_1, \ldots, \mathbf{c}_{M_W}\}$, denote by $\Pr\big(\text{error}|w, \mathbf{y}(\mathbf{c}_w,\mathbf{u}), \mathbf{z}\big)$ the probability of decoding error conditioned, first, on w and $\mathbf{u}$ entering the encoder, second, on the selection of a codeword $\mathbf{y}(w,j) \in \mathbf{c}_w$, denoted as $\mathbf{y}(\mathbf{c}_w,\mathbf{u})$, and on the channel output $\mathbf{z}$. Then the probability of decoding error given that watermark index w was embedded, averaged over the random choice of $\mathbf{c}$, satisfies

$$\begin{aligned}
\overline{P}_{e,w}^{(n)} \leq &\int_{\mathcal{B}_n(\alpha,\beta)} p(\mathbf{u})\int_{(\mathcal{E}_1(\mathbf{u}))^c} q(\mathbf{c}_w)\int_{\mathcal{Z}^n} f\big(\mathbf{z}|\mathbf{y}(\mathbf{c}_w,\mathbf{u})\big)\times \\
&\Pr\big(\text{error}|w, \mathbf{y}(\mathbf{c}_w,\mathbf{u}), \mathbf{z}\big)\, d\mathbf{z}\, d\mathbf{c}_w\, d\mathbf{u} + \int_{(\mathcal{B}_n(\alpha,\beta))^c} p(\mathbf{u})\, d\mathbf{u} \\
&+ \int_{\mathcal{B}_n(\alpha,\beta)} p(\mathbf{u})\int_{\mathcal{E}_1(\mathbf{u})} q(\mathbf{c}_w)\, d\mathbf{c}_w\, d\mathbf{u}. \quad (10)
\end{aligned}$$

Using Gallager's technique for deriving the random coding lower bound for the channel error exponent [7], and applying Lemma 1 and the inequality $\Phi_{D_Q}(\mathbf{y},\mathbf{u}) \leq \exp\big\{s\big[nD_Q - d(\mathbf{y},\mathbf{u})\big]\frac{\|\mathbf{u}\|^2}{n}\big\}$ for $s \geq 0$, we can upper bound the above integrals. Define $\Lambda_1(\gamma,\beta,\rho,s)$ and $\Lambda_2(\gamma,\beta)$ by (4) and (5), respectively. For $\forall\ n \geq \max\{N_1, N_2\}$, we obtain $\overline{P}_{e,w}^{(n)} \leq 4\exp\{-n(\min\big[\Lambda_1(\gamma,\beta,\rho,s), \Lambda_2(\gamma,\beta)\big])\} \triangleq \bar{\epsilon}^{(n)}$. Since the above bound is independent of the watermark index w, we then obtain a random coding upper bound for $\overline{P}_e^{(n)}$.

D. The Existence of A Sequence of (R_Q, R_W, n) Codes

Let $\mathcal{A}$ be the set of all the codes with $P_e^{(n)}(\mathbf{c}) \leq (\bar{\epsilon}^{(n)})^{1-\frac{1}{\sqrt{n}}}$, i.e., $\mathcal{A} \triangleq \{\mathbf{c} : P_e^{(n)}(\mathbf{c}) \leq (\bar{\epsilon}^{(n)})^{1-\frac{1}{\sqrt{n}}}\}$. Clearly, since $P_e^{(n)}(\mathbf{c})$ is a random variable (a function of the random code $\mathbf{c}$), it follows from Markov's inequality that $\Pr(\mathcal{A}) \geq 1 - (\bar{\epsilon}^{(n)})^{\frac{1}{\sqrt{n}}}$ for n sufficiently large. Thus, we have

$$\int_{\mathcal{A}}\frac{q(\mathbf{c})D^{(n)}(\mathbf{c})}{\Pr(\mathcal{A})}\, d\mathbf{c} \leq \frac{1}{\Pr(\mathcal{A})}\int q(\mathbf{c})D^{(n)}(\mathbf{c})\, d\mathbf{c} \leq \frac{D_Q + \bar{\delta}^{(n)}}{1 - (\bar{\epsilon}^{(n)})^{\frac{1}{\sqrt{n}}}}.$$

Since $(\bar{\epsilon}^{(n)})^{\frac{1}{\sqrt{n}}} \leq 4^{\frac{1}{\sqrt{n}}}\exp\{-\sqrt{n}\Lambda(\gamma,\beta,\rho,s)\}$, which goes to 0 as $n \to \infty$, there exists at least one sequence of codes $\{\widetilde{\mathbf{c}}\}$ satisfying $P_e^{(n)}(\widetilde{\mathbf{c}}) < (\bar{\epsilon}^{(n)})^{1-\frac{1}{\sqrt{n}}}$ and $D^{(n)}(\widetilde{\mathbf{c}}) \leq \frac{D_Q + \bar{\delta}^{(n)}}{1-(\bar{\epsilon}^{(n)})^{\frac{1}{\sqrt{n}}}} \leq D_Q + \delta$ simultaneously for n sufficiently large.

References

[1] D. Karakos and A. Papamarcou, "A relationship between quantization and watermarking rates in the presence of additive Gaussian attacks," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1970-1982, Aug. 2003.

[2] A. Maor and N. Merhav, "On joint information embedding and lossy compression," *IEEE Trans. Inform. Theory*, vol. 51, Aug. 2005.

[3] A. Maor and N. Merhav, "On joint information embedding and lossy compression in the presence of a memoryless attack," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3166-3175, Sep. 2005.

[4] E.-H. Yang and W. Sun, "On watermarking and compression rates of joint compression and private watermarking systems with abstract alphabets," *Proc. of the 2005 Canadian Workshop on Information Theory*, Montreal, Quebec, Canada, June 5-8, pp. 296-299, 2005.

[5] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, May 2001.

[6] N. Merhav and E. Ordentlich, "On causal and semicausal codes for joint information embedding and source coding," *IEEE Trans. Inform. Theory*, vol. 52, pp. 213-226, Jan. 2006.

[7] R.G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[8] N. Merhav, "On random coding error exponent of watermarking systems,", *IEEE Trans. Inform. Theory*, vol. 46, no. 2, Mar. 2000.

[9] A. Somekh-Baruch and N. Merhav, "On the error exponent and capacity games of private watermarking systems," *IEEE Trans. Inform. Theory*, vol. 49, pp. 537-563, Mar. 2003.

[10] S. Ihara and M. Kubo, "Error exponent for coding of memoryless Gaussian sources with a fidelity criterion," *IEICE Trans. Fundamentals*, vol. E83-A, no. 10, Oct. 2000.