CRYPTOGRAPHIC MEASURES IN INFORMATION HIDING

Phillip A. Regalia

Department of Electrical Engineering and Computer Science
Catholic University of America
Washington, DC 20064

ABSTRACT

Recent information hiding schemes are scrutinized in terms of their cryptographic performance. We establish conditions under which the key equivocation function is optimal for the studied schemes, and show that, under a reasonable key generation model, the perfect secrecy property is nearly satisfied, limited by a mutual information measure that decreases exponentially with the block length. The novelty of the work is to extend classical cryptographic analysis results to schemes involving cover signals, a component absent from standard cryptography. The schemes show unexpectedly good cryptographic security, although we observe that information embedding with robustness has steganographic weaknesses.

Index Terms— Information hiding; wet paper coding; key equivocation; message equivocation; perfect secrecy.

1. INTRODUCTION

Contemporary information hiding has recast watermarking on solid information theoretic grounds to assess embedding capacity and robustness to attack [1]–[3]. An equally important consideration from a cryptographic or steganographic viewpoint, however, is how well the message is indeed hidden in the first place. Although one can always, in principle, encrypt a message prior to embedding it, this presents additional overhead, and ignores the issue of whether information hiding algorithms have any intrinsic security features.

With this view, we reexamine information embedding [1]–[3] and wet paper coding [4], [5] (representing two offshoots of dirty paper coding [6]), by exploiting the crossroads of coding and cryptography (e.g., [7]–[11]). Although not designed from cryptographic considerations, the schemes under study are shown to exhibit unexpectedly good cryptographic security for message and key equivocation functions, due in essence to the additional randomness injected by the cover signal. This may obviate further message obfuscation stages in some applications.

Section 2 reviews the hiding schemes under study, and our main results are collected in Section 3.

This work is supported by the National Science Foundation under grant CCF 0634757.

2. PROBLEM SET-UP

We begin with the basic set-up of Figure 1. A cover signal (image, audio, video, etc.) generates a binary carrier sequence by way of a parity function which is assumed publicly known [12], [4]. The carrier sequence is modified to embed a given (plain text) message, producing a watermarked (or cipher) signal. The original cover signal is then modified into the stego signal whose parity function output agrees with the cipher signal [4], [12], as indicated by the inverse parity function block. The sender and receiver agree on a private key.

Random variables are denoted with upper case italic letters, with lower case bold letters denoting a particular realization. Thus S denotes the carrier sequence from the parity function (with $\mathbf{s}$ a particular realization), M denotes the message to embed, K denotes the key, and C denotes the cipher signal produced by the hiding function. We treat all signals as vectors of bits (each 0 or 1), using componentwise modulo-2 addition over the Galois field $\mathcal{F}_2$. In particular:

- The carrier sequence contains n bits, derived from n parity checks that comprise the parity function, so that $S \in \mathcal{F}_2^n$. We assume all 2^n realizations of S are equally probable, as furnished by a "good" parity function.
- The (plain text) message M collects q bits ($M \in \mathcal{F}_2^q$) with $q < n$, and all 2^q configurations of M are assumed equally probable (though, e.g., compression [7]).
- The key $K \in \mathcal{F}_2^{q\times n}$ is a $q \times n$ parity check matrix. For a particular realization $\mathbf{k}$ of this matrix, and a particular message $\mathbf{m}$, the chosen cipher signal $\mathbf{c} \in \mathcal{F}_2^n$ satisfies
$$\mathbf{m} = \mathbf{kc}$$
among other constraints to be detailed below. Thus if the receiver knows the key $\mathbf{k}$, the hidden message $\mathbf{m}$ can be recovered from the cipher text $\mathbf{c}$.

Entropy is denoted by $H(\cdot)$ and mutual information by $I(\cdot\,;\cdot)$:
$$H(S) = -\sum_{\mathbf{s}\in\mathcal{F}_2^n} \Pr(\mathbf{s}) \log_2 \Pr(\mathbf{s})$$
$$I(M;C) = \sum_{\mathbf{m}\in\mathcal{F}_2^q} \sum_{\mathbf{c}\in\mathcal{F}_2^n} \Pr(\mathbf{m},\mathbf{c}) \log_2 \frac{\Pr(\mathbf{m},\mathbf{c})}{\Pr(\mathbf{m})\Pr(\mathbf{c})}$$

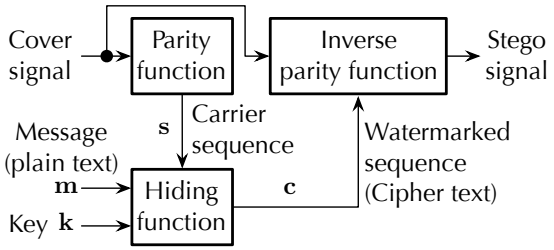

Fig. 1. Information hiding set-up with parity-function data.

For a given key $\mathbf{k}$, assumed of full rank q, the set of binary vectors $\mathbf{b} \in \mathcal{F}_2^n$ that lie in its null space defines a code of rate $r = (n-q)/n$, denoted $G_{\mathbf{k}}(\mathbf{0})$:

$$G_{\mathbf{k}}(\mathbf{0}) = \{\mathbf{b} \in \mathcal{F}_2^n : \mathbf{0} = \mathbf{kb}\}.$$

The set of binary vectors which produce instead a given "syndrome" $\mathbf{m}$ defines a coset [13] $G_{\mathbf{k}}(\mathbf{m})$ for that syndrome:

$$G_{\mathbf{k}}(\mathbf{m}) = \{\mathbf{b} \in \mathcal{F}_2^n : \mathbf{m} = \mathbf{kb}\}.$$

The member of $G_{\mathbf{k}}(\mathbf{m})$ of lowest Hamming weight (smallest number of 1s) is the *coset leader* for the syndrome $\mathbf{m}$.[1]

Modern information embedding [1]–[3] (here specialized to the binary case) produces a cipher signal $\mathbf{c}$ minimizing the Hamming distance $d(\mathbf{s}, \mathbf{c})$ subject to the constraint $\mathbf{m} = \mathbf{kc}$. The amounts to "quantizing" the carrier sequence $\mathbf{s}$ to the coset $G_{\mathbf{k}}(\mathbf{m})$. If $\mathbf{e}$ denotes the coset leader for the syndrome $\mathbf{m} - \mathbf{ks}$, then $\mathbf{c} = \mathbf{s} + \mathbf{e}$ is the closest member of $G_{\mathbf{k}}(\mathbf{m})$ to $\mathbf{s}$. For any $\mathbf{m}$, define the average distortion (per bit) as

$$D = \frac{1}{n} \sum_{\mathbf{s} \in \mathcal{F}_2^n} \Pr(\mathbf{s}) \min_{\mathbf{c} \in G_{\mathbf{k}}(\mathbf{m})} d(\mathbf{s}, \mathbf{c})$$

Since each $\mathbf{s}$ is quantized to a code of rate $r = (n-q)/n$, the average distortion is lower bounded through the rate-distortion function [14, Thm. 13.3.1] as

$$H_2(D) \geq H'(S) - r = 1 - r$$

in which $H_2(D) = -D \log_2 D - (1-D) \log_2 (1-D)$ is the binary entropy function, and $H'(S)$ is the per-bit entropy rate of the carrier sequence S (with $H'(S) = H(S)/n$ for large enough n); here $H'(S) = 1$ since S is uniformly distributed. The lower bound on D [where $H_2(D) = 1 - r$] is achieved if the (error correction) code $G_{\mathbf{k}}(\mathbf{0})$ achieves channel capacity over a binary symmetric channel with error probability D (e.g., [14], [15]). Design methods for such codes [16], [17] may thus be used to generate "low distortion" keys.

Wet paper coding [4], [5] likewise produces a cipher sequence $\mathbf{c}$ fulfilling $\mathbf{m} = \mathbf{kc}$, but with different constraints. An index set—call it $\mathbf{t}$—collects q integers from $\{1, 2, \ldots, n\}$

[1] In case of nonuniqueness, a particular lowest-weight member is arbitrarily selected as the unique coset leader.

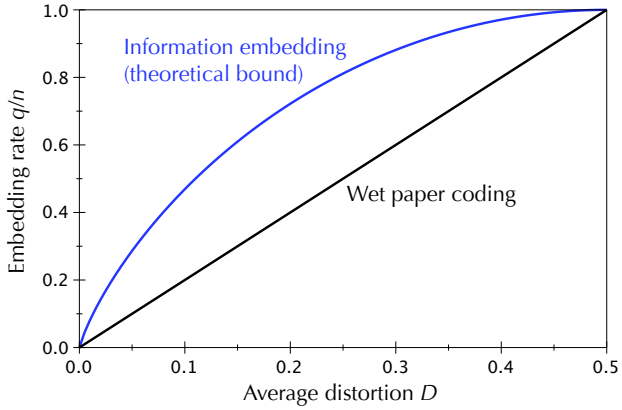

Fig. 2. Embedding rate versus distortion for the two schemes.

and permits $\mathbf{c}$ to differ from the carrier sequence $\mathbf{s}$ only in the positions comprising $\mathbf{t}$. This index set gives an "information set" [9] provided a $q \times q$ submatrix of $\mathbf{k}$—built by retaining columns whose indices are in $\mathbf{t}$—has full rank, giving then a unique solution for $\mathbf{c}$. The index set is randomly selected, and need not be known to the receiver; the randomness is to better evade detection by an eavesdropper [12], [4]. We thus introduce T (a "tool" which complements the key) as a random variable comprising the q indices used in the hiding stage, with $\mathbf{t}$ denoting a particular outcome. The average distortion from this method is $D = 0.5q/n$ [4], with q/n the embedding rate. This is larger than the average distortion attainable using the information embedding construct reviewed above; cf. Fig. 2. An intermediary between the two curves of Fig. 2 is obtained by allowing $\mathbf{c}$ and $\mathbf{s}$ to differ in l positions, with $q \leq l \leq n$, giving 2^{l-q} possibilities for $\mathbf{c}$ and thus generally lower distortion [5] than the formulation of [4]. For clarity, the extremes $l = q$ and $l = n$ are treated in what follows.

2.1. Embedding with robustness

Information embedding can also allow the message M to be recovered even if the cipher text C suffers further distortion [1]–[3]. The basic construct is to partition the key $\mathbf{k}$ row-wise and choose the closest $\mathbf{c}$ to $\mathbf{s}$ that satisfies

$$\begin{matrix} q \\ q' \end{matrix} \left\{ \begin{bmatrix} \mathbf{m} \\ \mathbf{0} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{k}_1 \\ \mathbf{k}_2 \end{bmatrix}}_{\mathbf{k}} \mathbf{c} \right.$$

in which the null space of $\mathbf{k}_2$ gives a "good" error correction code, and that of $\mathbf{k}$ a "good" quantization code [2], [13]. If a sufficient number of cipher texts $\mathbf{c}$ are observed for the same key $\mathbf{k}$, the linear subspace they span builds the orthogonal complement to $\mathbf{k}_2$. This reveals information on the key $\mathbf{k}$ and, more seriously from a steganographic viewpoint, alerts an observer that $\mathbf{c}$ may contain a hidden message. For this reason, we shall not integrate message robustness in the analysis

to follow. This amounts to removing $\mathbf{k}_2$, and the construct reverts to its simpler form $\mathbf{m} = \mathbf{kc}$.

3. CRYPTOGRAPHIC SECURITY MEASURES

We begin with the key equivocation function [10] $I(K;C) = H(K) - H(K|C)$ which measures how much information may be revealed about the key K from observations of the cipher text C, and then study the message equivocation function $I(M;C)$ underlying the perfect secrecy [10], [11] condition.

Lemma 1 *The key equivocation function is given by*

$$\begin{aligned} I(K;C) &= H(C) - H(M) - I(S;K,C) \\ &\qquad \textit{(information embedding)} \\ &= H(C) - H(M) - I(T;K,C) - I(S;K,C,T) \\ &\qquad \textit{(wet paper encoding)} \end{aligned}$$

These differ from a standard result [10, Thm. 2.10] by the inclusion of mutual information terms involving the carrier sequence S and/or tool set T, which are absent in classical cryptography. For the proof, expand the joint entropy as

$$\begin{aligned} H(C,K,M,T,S) &= H(K,M,T,S) + \underbrace{H(C|K,M,T,S)}_{=0} \\ &= H(K) + H(M) + H(T) + H(S) \end{aligned}$$

in which $H(C|K,M,T,S) = 0$ since the key, message, carrier and tool together determine the cipher text; the key, message, carrier signal and tool are likewise assumed mutually independent. By a separate expansion, we also have

$$\begin{aligned} H(C,K,M,T,S) &= H(C) + H(K|C) + \underbrace{H(M|K,C)}_{=0} \\ &\quad + H(T|C,K,M) + H(S|C,K,M,T) \end{aligned}$$

in which $H(M|K,C) = 0$, $H(T|C,K,M) = H(T|C,K)$ and $H(S|C,K,M,T) = H(S|C,K,T)$ since the key and cipher text determine the message. Equating the expansions and isolating $H(K) - H(K|C) = I(K;C)$ gives the statement for wet paper encoding. The information embedding expression then follows by removing the tool variable T. ◇

The following theorem gives conditions which ensure that the key is not revealed by the cipher text, and that the cipher text has maximum entropy:

Theorem 1 *If $H(S) = n$ (all carrier sequences equally probable) and $H(M) = q$ (all messages equally probable), then for either scheme,*

$$I(K;C) = 0 \quad \textit{and} \quad H(C) = n.$$

For the proof, consider first the information embedding scheme. Insert $I(S;K,C) = H(S) - H(S|K,C)$ into the expression of lemma 1, and isolate $H(S|K,C)$ as

$$\begin{aligned} H(S|K,C) &= I(K;C) + H(M) + H(S) - H(C) \\ &= I(K;C) + q + n - H(C) \end{aligned} \tag{1}$$

We claim now that $H(S|K,C) \le q$. To verify, consider any realization $(\mathbf{k},\mathbf{c})$, and set $\mathbf{m} = \mathbf{kc}$. To construct a candidate carrier sequence $\mathbf{s}$, we note that $\mathbf{s}$ and $\mathbf{c}$ must differ by a coset leader, of which there are 2^q. (Pick any $\mathbf{d} \in \mathcal{F}_2^q$ and let $\mathbf{e}$ be the coset leader for $\mathbf{m} - \mathbf{d}$, to get $\mathbf{s} = \mathbf{c} + \mathbf{e}$). This limits $\mathbf{s}$ to one of 2^q configurations. As such, the entropy of S given any fixed couple $(\mathbf{k},\mathbf{c})$ is bounded as $H(S|\mathbf{k},\mathbf{c}) \le q$. By averaging over the joint probability $\Pr(\mathbf{k},\mathbf{c})$,

$$H(S|K,C) = \sum_{\mathbf{k},\mathbf{c}} \Pr(\mathbf{k},\mathbf{c})\, H(S|\mathbf{k},\mathbf{c}) \le q$$

as well. This gives, via (1), $I(K;C) + q + n - H(C) \le q$ or

$$I(K;C) \le H(C) - n.$$

But as the cipher signal has n bits, necessarily $H(C) \le n$, giving $I(K;C) \le 0$. As mutual information is nonnegative, we get $I(K;C) = 0$, $H(C) = n$, and $H(S|K,C) = q$.

For the wet paper scheme, lemma 1 reads

$$H(S|K,C,T) = I(K;C) + I(T;K,C) + q + n - H(C)$$

using still $H(M) = q$ and $H(S) = n$. We again claim $H(S|K,C,T) \le q$, since knowledge of the index set $\mathbf{t}$ and cipher signal $\mathbf{c}$ identifies the q positions in which $\mathbf{s}$ and $\mathbf{c}$ potentially differ. This likewise reduces $\mathbf{s}$ to one of 2^q possibilities, to give $H(S|C,K,T) \le q$ as above, and thus

$$I(K;C) + I(T;K,C) \le H(C) - n$$

with still $H(C) \le n$. This implies $I(K;C) = 0$, $H(C) = n$, $I(T;K,C) = 0$, and $H(S|K,C,T) = q$. ◇

We consider next the message equivocation $I(M;C)$ that measures how much information the cipher signal C reveals about the message M. By rearranging two expansions of the joint entropy $H(M,C,K,T,S)$, we can show

$$\begin{aligned} I(M;C) &= H(M) - H(M|C) \\ &= H(C) - H(K) - H(T) - H(S) \\ &\quad + H(K|M,C) + H(T|M,K,C) \\ &\quad + H(S|M,K,C,T) \\ &= q - I(K;M,C) \end{aligned}$$

using $H(C) = H(S) = n$, $I(T;M,C,K) = I(T;C,K) = 0$ and $H(S|M,K,C,T) = H(S|K,C,T) = q$ from theorem 1. The same expression applies to the information embedding scheme upon removing the tool variable T. Now, the non-negativity of $I(M;C)$ implies $I(K;M,C) \le q$. But knowledge of the message M and cipher signal C reveals q parity constraints on the key; if a full q bits of information are imparted so that $I(K;M,C) = q$, then the *perfect secrecy* [10], [11] condition $I(M;C) = 0$ will be satisfied.

In this direction, suppose we model the elements K_{ij} of the key as i.i.d. Bernoulli random variables, with

$$\Pr(K_{ij} = 1) = 1 - \Pr(K_{ij} = 0) = p.$$

(The value p is small, and diminishes as $1/n$, for a good low density parity check matrix). We then claim:

Theorem 2 *For sufficiently large nq,*

$$I(M;C) \approx 2^{-n} nq H_2(p)$$

where $H_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$.

As $I(M;C) = q - I(K;M,C)$, we examine $I(K;M,C) = H(K) - H(K|M,C)$. Since the nq elements of the key K are i.i.d. and Bernoulli, $H(K) = nqH_2(p)$. The conditional entropy $H(K|M,C)$ may then be expanded as

$$H(K|M,C) = \sum_{\mathbf{m},\mathbf{c}} H(K|\mathbf{m},\mathbf{c})\ \Pr(\mathbf{m},\mathbf{c})$$

Consider first the event $\mathbf{c} = \mathbf{0}$. This implies $\mathbf{m} = \mathbf{0}$; trivially all keys satisfy $\mathbf{0} = \mathbf{k0}$. Thus $H(K|\mathbf{0},\mathbf{0}) = nqH_2(p)$. Now, $\Pr(\mathbf{m}=\mathbf{0},\mathbf{c}=\mathbf{0}) = 2^{-n}$ since $H(C) = n$ implies $\Pr(\mathbf{c}) = 2^{-n}$ for each $\mathbf{c}$, and

$$\Pr(\mathbf{c}=\mathbf{0}) = \Pr(\mathbf{m}=\mathbf{0},\mathbf{c}=\mathbf{0}) + \sum_{\mathbf{m}\neq\mathbf{0}} \underbrace{\Pr(\mathbf{m},\mathbf{c}=\mathbf{0})}_{=0}$$

Thus $H(K|\mathbf{0},\mathbf{0})\ \Pr(\mathbf{0},\mathbf{0}) = 2^{-n}nqH_2(p)$.

Introduce now the typical set of keys

$$A_{nq}^{\epsilon} = \left\{\mathbf{k} : nq(H_2(p)-\epsilon) \le -\log_2 \Pr(\mathbf{k}) \le nq(H_2(p)+\epsilon)\right\}.$$

For any fixed ϵ, the probability mass of the typical set A_{nq}^{ϵ} approaches 1 arbitrarily closely as nq grows, and has cardinality $|A_{nq}^{\epsilon}| \approx 2^{nqH_2(p)}$ [14]. Now, for any cipher text $\mathbf{c} \neq \mathbf{0}$ and any message $\mathbf{m}$, denote by

$$A_{nq}^{\epsilon}(\mathbf{m},\mathbf{c}) = \left\{\mathbf{k} \in A_{nq}^{\epsilon} : \mathbf{m} = \mathbf{kc}\right\}$$

the subset of typical keys consistent with the given $(\mathbf{m},\mathbf{c})$. As the equation $\mathbf{m} = \mathbf{kc}$ introduces q parity constraints, a fraction $1/2^q$ of the typically keys will satisfy them, so that $|A_{nq}^{\epsilon}(\mathbf{m},\mathbf{c})| \approx 2^{nqH_2(p)}/2^q$. For sufficiently small ϵ, the probability mass function in the typical set is nearly uniform [$\Pr(\mathbf{k}) \approx 2^{-nqH_2(p)}$ for $\mathbf{k} \in A_{nq}^{\epsilon}$], so that

$$\begin{aligned} H(K|\mathbf{m},\mathbf{c}) &= -\sum_{\mathbf{k}\in A_{nq}^{\epsilon}(\mathbf{m},\mathbf{c})} \frac{\Pr(\mathbf{k})}{|A_{nq}^{\epsilon}(\mathbf{m},\mathbf{c})|}\ \log_2 \frac{\Pr(\mathbf{k})}{|A_{nq}^{\epsilon}(\mathbf{m},\mathbf{c})|} \\ &\approx nqH_2(p) - q, \qquad \text{for each } \mathbf{m} \text{ and } \mathbf{c} \neq \mathbf{0}. \end{aligned}$$

Thus $\sum_{\mathbf{m},\mathbf{c}\neq\mathbf{0}} H(K|\mathbf{m},\mathbf{c})\Pr(\mathbf{m},\mathbf{c}) \approx nqH_2(p) - q$. Combining the pieces,

$$\begin{aligned} I(K;M,C) &= H(K) - H(K|M,C) \\ &\approx nqH_2(p) - 2^{-n}nqH_2(p) - nqH_2(p) + q \end{aligned}$$

giving $I(M;C) = q - I(K;M,C) \approx 2^{-n}nqH_2(p)$. ◇

Alternatively, if $\mathbf{k}$ is a "good" parity check matrix, then so is $\mathbf{P}_q\mathbf{k}\mathbf{P}_n$, where $\mathbf{P}_q$ and $\mathbf{P}_n$ are permutation matrices. This gives up to $q!n!$ keys to choose from, each equally probable. The same conclusion in theorem 2 can be reached, without resorting to typical sets. Thus perfect secrecy is approached exponentially fast in the block length n.

4. REFERENCES

[1] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Information Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.

[2] R. J. Barron, B. Chen, and G. W. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Information Theory*, vol. 49, no. 5, pp. 1159–1180, May 2003.

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

[4] J. Fridrich, M. Goljan, P. Lisoněk, and D. Soukal, "Writing on wet paper," *IEEE Trans. Signal Processing*, vol. 10, no. 53, pp. 3923–3935, Oct. 2005.

[5] J. Fridrich, M. Goljan, and D. Soukal, "Wet paper codes with improved coding efficiency," *IEEE Trans. Information Forensics and Security*, vol. 1, no. 1, pp. 102–110, Mar. 2006.

[6] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[7] M. E. Hellman, "An extension of the Shannon theory approach to cryptography," *IEEE Trans. Information Theory*, vol. 23, no. 3, pp. 289–294, May 1977.

[8] A. Canteaut and F. Chabaud, "A new algorithm for finding minimum-weight words in a linear code: Application to McEliece's cryptosystem and to narrow-sense BCH codes of length 511," *IEEE Trans. Information Theory*, vol. 44, no. 1, pp. 367–378, Jan. 1998.

[9] T. Johansson and F. Jönsson, "On the complexity of some cryptographic problems based on the general decoding problem," *IEEE Trans. Information Theory*, vol. 48, no. 10, pp. 2669–2678, Oct. 2002.

[10] D. R. Stinson, *Cryptography: Theory and Practice*, Chapman and Hall/CRC, Boca Raton, FL, 3rd edition, 2006.

[11] W. Trappe and L. C. Washington, *Introduction to Cryptography with Coding Theory*, Prentice-Hall, Upper Saddle River, NJ, 2nd edition, 2006.

[12] R. J. Anderson and F. A. P. Petitcolas, "On the limits of steganography," *IEEE J. Selected Areas in Communications*, vol. 16, no. 4, pp. 474–481, Apr. 1998.

[13] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Information Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, Hoboken, NJ, 2nd edition, 2007.

[15] E. Martinian and M. J. Wainwright, "Low density codes achieve the rate-distortion bound," in *Proc. Data Compression Conference*, Snow Bird, UT, Mar. 2006.

[16] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Information Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[17] A. W. Eckford, F. R. Kschischang, and S. Pasupathy, "On designing good LDPC codes for Markov channels," *IEEE Trans. Information Theory*, vol. 53, no. 1, pp. 5–21, Jan. 2007.