Error Exponents for Channel Coding with Side Information

Pierre Moulin and Ying Wang
University of Illinois, Beckman Institute, CSL and Dept of Electrical and Computer Engineering,
Urbana, IL 61801, USA
{moulin,wang11}@cs.uiuc.edu

Abstract

Capacity formulas and random-coding and sphere-packing exponents are derived for a generalized family of Gel'fand-Pinsker coding problems. Information is to be reliably transmitted through a noisy channel with random state sequence. Partial information about the state sequence is available to the encoder and decoder. Two families of channels are considered: 1) compound discrete memoryless channels (C-DMC), and 2) channels with arbitrary memory, subject to an additive cost constraint, or more generally to a constraint on the conditional type of the channel output given the input. Both problems are closely connected. For the C-DMC case, our random-coding and sphere-packing exponents coincide at high rates, thereby determining the reliability function of the channel family. The random-coding exponent is achieved using a 3-D binning scheme and a maximum penalized mutual information decoder. In the case of arbitrary channels with memory, a larger random-coding error exponent than in the C-DMC case is obtained. Applications of this study include watermarking, data hiding, communication in presence of partially known interferers, and problems such as broadcast channels, all of which involve the fundamental idea of binning.

I Introduction

In 1980, Gel'fand and Pinsker studied the problem of coding for a discrete memoryless channel (DMC) with random state parameters, which are observed by the encoder but not by the decoder [1]. They derived the capacity of this channel and showed it is achievable by a random binning scheme. Applications of their work include computer memories with defects [2] and writing on dirty paper [3]. In the late 1990's, it was found that the problem of hiding information in cover signals is closely related to the Gel'fand-Pinsker problem: the cover signal plays the role of the state sequence in the Gel'fand-Pinsker problem. Capacity expressions were derived under expected distortion constraints for the transmitter and a memoryless attacker [4].

In problems such as data hiding, the assumption of a fixed channel is untenable when the channel is under partial control of an adversary. This motivated the game-theoretic approach of [4], where the worst channel in a class of memoryless channels was derived, and capacity is the solution to a maxmin mutual-information game. This game-theoretic approach was recently extended by Somekh-Baruch and Merhav [5, 6], who considered a broad class of channels with memory, subject to almost-sure distortion constraints. In the special case of private data hiding, in which the cover signal is known to both the encoder and the decoder, they also derived random-coding and sphere-packing exponents [5]. No binning scheme is necessary in this scenario. Another related problem is the classical memoryless arbitrary varying channel (AVC) [7] in which no side information is available to the encoder or decoder. The AVC model is often used to analyze jamming problems. Error exponents for this model were derived in [8, 9].

The coding problems considered in this paper are motivated by data hiding applications, in which the decoder has partial or no knowledge of the cover signal. In all cases capacity is achievable by binning schemes. Finding the best error exponents for such schemes is challenging, see recent results in [10, 11], obtained independently of ours. The exponents we have derived cannot be achieved by standard binning schemes and standard maximum mutual information (MMI) decoders. Our random-coding error exponent is achieved using a stack of variable-size codeword arrays indexed by the type of the state sequence. This code may be viewed as a 3-D binning scheme. The appropriate decoder is a maximum penalized mutual information (MPMI) decoder, where the penalty is a function of the state sequence type.

II Statement of the Problem

Figure 1: Communication with side information at the encoder and decoder.

Our generic problem of communication with side information at the encoder and decoder is modeled in Fig. 1

1This research was supported by NSF under ITR grants CCR 00-81268 and CCR 00-22024.

For instance, the decoder may have access to a noisy, compressed version of the original cover signal.
Table 1: Relation between S\textsuperscript{e}, S\textsuperscript{a}, S\textsuperscript{d} and S for various coding problems with side information.

<table>
<thead>
<tr>
<th>Problem</th>
<th>S\textsuperscript{e}</th>
<th>S\textsuperscript{a}</th>
<th>S\textsuperscript{d}</th>
<th>Binary?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gel'fand-Pinsker</td>
<td>S</td>
<td>S</td>
<td>∅</td>
<td>yes</td>
</tr>
<tr>
<td>Cover-Chiang</td>
<td>S\textsuperscript{e}</td>
<td>(S\textsuperscript{e}, S\textsuperscript{d})</td>
<td>S\textsuperscript{d}</td>
<td>yes</td>
</tr>
<tr>
<td>Public Watermarking</td>
<td>S</td>
<td>∅</td>
<td>∅</td>
<td>yes</td>
</tr>
<tr>
<td>Semiblind Watermarking</td>
<td>S</td>
<td>∅</td>
<td>f(S)</td>
<td>yes</td>
</tr>
<tr>
<td>Private Watermarking</td>
<td>S</td>
<td>∅</td>
<td>S</td>
<td>no</td>
</tr>
<tr>
<td>Jamming</td>
<td>S</td>
<td>0</td>
<td>S</td>
<td>no</td>
</tr>
</tbody>
</table>

Figure 2: Representation of binning scheme as a stack of variable-size arrays indexed by state sequence type.

and includes the problems listed in Table II as special cases. The state sequence \( S = (S_1, \ldots, S_N) \) consists of independent and identically distributed (i.i.d.) samples drawn from a probability mass function (p.m.f.) \( p_S(s) \), \( s \in S \). Degraded versions \( S^e, S^a, \) and \( S^d \) of the state sequence \( S \) are available to the encoder, adversary and decoder, respectively. The triple \( (S^e, S^a, S^d) \) is the output of a DMC with conditional p.m.f. \( p(s^e, s^a, s^d | s) \); without significant loss of generality, we assume that \( (s^e, s^a, s^d) \) is an invertible function of \( s \).

A message \( M \) is to be transmitted to a decoder; \( M \) is uniformly distributed over the message set \( \mathcal{M} \). The transmitter produces a sequence \( X \) through a function \( f_N(S^e, M) \), in an attempt to reliably transmit the message \( M \) to the decoder. The adversary passes \( X \) through some attack channel \( p_{Y|X|S^e}(y|x,s^e) \) to produce corrupted data \( Y \). The decoder does not know \( p_{Y|X|S^e} \)-selected by the adversary and does not have access to the original covertext. The decoder produces an estimate \( \hat{M} = g_N(Y, S^d) \in \mathcal{M} \) of the transmitted message. The alphabets for \( X \) and \( Y \) are denoted by \( \mathcal{X} \) and \( \mathcal{Y} \), respectively. We allow the encoder/decoder pair \( (f_N, g_N) \) to be randomized, i.e., the choice of \( (f_N, g_N) \) is a function of a random variable known to the encoder and decoder but not to the attacker. We can think of this random variable as a secret key. Formally, the randomized code will be denoted by \( (F_N, G_N) \).

**A Notation**

We use uppercase letters for random variables, lowercase letters for individual values, and boldface fonts for sequences. Entropy of a random variable \( X \) is denoted by \( H(X) \), and mutual information between two random variables \( X \) and \( Y \) is denoted by \( I(X;Y) \); this is also denoted by \( I_{XY}(p_{XY}) \) when viewed as a function of the underlying \( p_{XY} \). Kullback-Leibler divergence between two p.m.f.'s \( p \) and \( q \) is denoted by \( D(p||q) \).

Following the notation in Csiszár and Körner [7], let \( p_x \) denote the type of a sequence \( x \in \mathcal{X}^N \) (\( p_x \) is a p.m.f. over \( \mathcal{X} \)) and \( T_x \) the type class associated with \( p_x \), i.e., the set of all sequences of type \( p_x \). Likewise, we define the joint type \( p_{xy} \) of a pair of sequences \( (x, y) \in \mathcal{X}^N \times \mathcal{Y}^N \) (a p.m.f. over \( \mathcal{X} \times \mathcal{Y} \)) and \( T_{xy} \) the type class associated with \( p_{xy} \), i.e., the set of all sequences of type \( p_{xy} \). Finally, we define the conditional type \( p_{y|x} \) of a pair of sequences \( (x, y) \) as \( p_{y|x} = \frac{p_{xy}}{p_x(x)} \) for all \( x \in \mathcal{X} \) such that \( p_x(x) > 0 \). The conditional type class \( T_{y|x} \) is the set of all sequences \( y \) such that \( (x, y) \in T_{xy} \). We denote by \( H(x) \) the entropy of the p.m.f. \( p_x \) and by \( I(x; y) \) the mutual information for the joint p.m.f. \( p_{xy} \). We let \( P(\mathcal{X}) \) and \( P[\mathcal{X}] \) represent the set of all p.m.f.'s and all empirical p.m.f.'s, respectively, on the alphabet \( \mathcal{X} \). Likewise, \( P(\mathcal{X} \times \mathcal{Y}) \) and \( P[\mathcal{X} \times \mathcal{Y}] \) denote the set of all conditional p.m.f.'s and all empirical conditional p.m.f.'s on the alphabet \( \mathcal{Y} \). We define \( |t|^+ = \max(|0, t|) \).

**B Constrained Side-Information Codes**

A cost function \( \Gamma : \mathcal{S}^e \times \mathcal{X} \to \mathbb{R}^+ \) is defined to quantify the cost \( \Gamma(s^e, x) \) of transmitting symbol \( x \) when the channel state at the encoder is \( s^e \). This definition is extended to \( N \)-vectors using \( \Gamma^N(s^e, x) = \frac{1}{N} \sum_{i=1}^{N} \Gamma(s_i^e, x_i) \).

In information embedding applications, \( \Gamma \) is a distortion function measuring the distortion between host signal and marked signal.

We now define a class of codes that satisfy maximum cost constraints (Def. II.1) and a class of codes that satisfy average cost constraints (Def. II.2). The latter class is of course larger than the former.

**Definition II.1** A length-\( N \), rate-\( R \), randomized code with side information and maximum cost \( D_1 \) is a triple \( (\mathcal{M}, F_N, G_N) \), where:

- \( \mathcal{M} \) is the message set of cardinality \( |\mathcal{M}| = 2^{NR} \);
- \( (F_N, G_N) \) has joint distribution \( p(f_N, g_N) \);
- \( f_N : (\mathcal{S}^e)^N \times \mathcal{X} \to \mathcal{X}^N \) is the encoder mapping the state sequence \( s^e \) and message \( m \) to the transmitted sequence \( x = f_N(s^e, m) \).

The mapping is subject to the cost constraint

\[
\Gamma^N(s^e, f_N(s^e, m)) \leq D_1 \quad \text{almost surely } (p_{s^e}p_{f_N}p_M);
\]

- \( g_N : (\mathcal{Y} \times \mathcal{S}^d)^N \to \mathcal{M} \) is the decoder mapping the received sequence \( y \) and channel state sequence \( s^d \) to a decoded message \( \hat{m} = g_N(y, s^d) \).
Definition II.2 A length-$N$, rate-$R$, randomized code with side information and expected cost $D_1$ is a triple $(\mathcal{M}, F_N, G_N)$ which satisfies the same conditions as in Def. II.1, except that (1) is replaced with the looser constraint
\begin{equation}
\sum_{s^e} p(s^e) \sum_{f_N} \sum_{m \in \mathcal{M}} \frac{1}{|\mathcal{M}|} F_N(s^e, f_N(s^e, m)) \leq D_1.
\end{equation}

Remark II.1 Def. II.2 is analogous to the definition of a length-$N$ information hiding code in [4]. The common source randomness between encoder and decoder appears via the distribution $p(f_N, g_N)$ whereas in [4] it appears via a cryptographic key sequence $k$ with finite entropy rate.

C Constrained Attack Channels

Next we define a class $\mathcal{A}$ of discrete memoryless channels (DMC) (Def. II.3) and a closely related class $\mathcal{P}_{Y|XS^a} = [A]$ of arbitrarily varying channels (AVC) in which the conditional type of $y$ given $(x, s^a)$ is constrained (Def. II.4). Unlike the classical AVC model [7], the class considered here has arbitrary memory.

Definition II.3 The C-DMC class $\mathcal{A}$ is a subset of $\mathcal{P}_{Y|XS^a}$, the set of all possible DMC’s.

For DMC’s, we have $p_{Y|XS^a}(y|x, s^a) = \prod_{i=1}^N p_{Y|XS^a}(y_i|x_i, s^a_i)$. For simplicity of the exposition, we assume that $\mathcal{A}$ is a closed set. The set $\mathcal{A}$ is defined according to the application.

1. In the case of a known channel [1], $\mathcal{A}$ is a singleton.

2. In information hiding problems [4], $\mathcal{A}$ is the class of DMC’s that introduce expected distortion at most equal to $D_2$:

\begin{equation}
\sum_{s^a, x, y} p_{X,S^a}(x, s^a)p_{Y|XS^a}(y|x, s^a)d(x, y) \leq D_2,
\end{equation}

where $d: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^+$ is a distortion function. $\mathcal{A}$ can also be defined to be a subset of the above class.

3. In some applications, $\mathcal{A}$ could be defined via multiple cost constraints.

Definition II.4 The AVC class $\mathcal{P}_{Y|XS^a} = [A]$ is the set of channels for which the conditional type $p_{Y|XS^a}$ belongs to $\mathcal{A} \cap \mathcal{P}_{Y|XS^a}$ with probability $I$:

\begin{equation}
Pr[p_{Y|XS^a} \in \mathcal{A}] = 1.
\end{equation}

If $\mathcal{A}$ is defined via the distortion constraint (3), condition (4) may be rewritten as

\begin{equation}
Pr[d^N(x, y) \leq D_2] = 0,
\end{equation}

i.e., feasible channels have total distortion bounded by $ND_2$ and arbitrary memory. \footnote{The case of channels with arbitrary memory subject to expected-distortion constraints admits a trivial solution: the adversary substitutes $X$ with a fixed, nonzero probability that depends on $D_2$ but not on $N$, and therefore no reliable communication is possible in the sense of Def. II.5 below.}

D Probability of Error

The average probability of error for a deterministic code $(f_N, g_N)$ when channel $p_{Y|XS^a}$ is in effect is denoted by $P_e(f_N, g_N, p_{Y|XS^a})$. For a randomized code the expression above is averaged with respect to $p(f_N, g_N)$. The minmax probability of error for the class of randomized codes and the class of attack channels considered is given by

\begin{equation}
P_{e,N} = \min_{p(f_N, g_N)} \max_{p_{Y|XS^a}} \sum_{f_N, g_N} p(f_N, g_N)P_e(f_N, g_N, p_{Y|XS^a}).
\end{equation}

There are four combinations of maximum/expected cost constraints for the transmitter and C-DMC/AVC designs for the adversary, and a question is whether same capacity and error exponents will be obtained in all four cases. To simplify the presentation, we only develop the case of maximum cost constraints on the transmitter, in the sense of Def. II.1. We shall see that there is no advantage (in terms of capacity or error probability) to the transmitter in using the weaker expected-cost constraints (Def. II.2).

Definition II.5 A rate $R$ is said to be achievable if $P_{e,N} \rightarrow 0$ as $N \rightarrow \infty$.

Definition II.6 The capacity $C(D_1, \mathcal{A})$ is the supremum of all achievable rates.

Definition II.7 A transmit channel $p_{XU|S^e}$ is feasible if

\begin{equation}
\sum_{u, s^e, x} p_{XU|S^e}(x, u|s^e)p_{S^e}(s^e)\Gamma(s^e, x) \leq D_1
\end{equation}

where $U$ is an auxiliary random variable.
We denote by $P_{XU|S^e}(D_1)$ and $A$ the sets of feasible transmit and memoryless attack channels, respectively. Note that transmit channels have been termed covert channels [4] and watermarking channels [6] in the context of information hiding applications. We retain the terminology “attack channel” in this paper.

**Definition II.8** The reliability function for the class of randomized codes and attack channels considered is defined as

$$E(R) = \lim_{N \to \infty} \sup \left[ -\frac{1}{N} \log P_{e,N}^* \right].$$

(7)

### E Preliminaries

Consider a joint p.m.f. $p_{S^eS^aS^uU^X}(s^e, s^a, S^u, x, y)$. The following difference of mutual informations plays a fundamental role in capacity analyses [1] — [6] of channels with side information. It plays a central role in the analysis of error exponents as well:

$$J(p_{S^eS^aS^uU^X}) = I(U; Y S^d) - I(U; S^e).$$

(8)

All the optimization problems in this paper involve the function $J$. Carathéodory’s theorem can be used to bound the cardinality of $U$ to a finite number, without loss of optimality.

### III Main Results

The main tool used to prove the coding theorems in this paper is the method of types. Coding theorems are derived for fixed type classes, and optimal types for the transmitter and attacker are derived as solutions to maximin problems. A key idea to prove achievability of error exponents is that the worst attacks are uniform over conditional types, as in Somer – Baruch and Merhav’s watermarking capacity game [6]. The same type of attack is used to prove converse theorems as well. Proofs of the theorems appear in [13].

We define the following quantity, which turns out to be a capacity expression for the problems considered in [1] — [6] and in this paper (Theorem III.3):

$$C = C(D_1, A) = \max_{p_{XU|S^e} \in P_{XU|S^e}(D_1)} \min_{p_Y|X \in A} J(p_S p_{XU|S^e} p_Y|X S^a).$$

(9)

In the special case of degenerate $p_S$ (no side information), the maximum above is achieved by $U = X$, and capacity reduces to the standard formula $C = \max_{p_X} \min_{p_Y|X} I_{X,Y}(p_X p_Y|X)$.

**Theorem III.1** For the C-DMC case (Def. II.3), the reliability function is lower-bounded by the random-coding error exponent

$$E_r^{C-DMC}(R) = \frac{1}{N} \log \left[ \min_{p_S \in P_{S^e}} \max_{p_{XU|S^e} \in P_{XU|S^e}(D_1)} \min_{p_Y|X \in A} [D(\hat{p}_S p_{XU|S^e} p_Y|X S^a) + J(\hat{p}_S p_{XU|S^e} p_Y|X S^a) - R[+]^+] \right].$$

(10)

for all $R \leq C$.

**Theorem III.2** For the AVC case (Def. II.4), the reliability function is lower-bounded by the random-coding error exponent

$$E_r^{AVC}(R) = \min_{p_S \in P_{S^e}} \max_{p_{XU|S^e} \in P_{XU|S^e}(D_1)} \min_{p_Y|X \in A} [D(\hat{p}_S p_{XU|S^e} p_Y|X S^a) + J(\hat{p}_S p_{XU|S^e} p_Y|X S^a) - R[+]^+] \right].$$

(11)

Moreover, $E_r^{AVC}(R) = 0$ for $R \geq C$.

The random-coding error exponents (10) and (11) are achieved by conditionally constant-composition codes (using a 3-D binning technique) and a MPMI decoder. The worst attack channel is uniform over a single conditional type.

The following theorem is an extension of previous results in [4, 6], which respectively solved the case of expected distortion constraints (for memoryless attacks) and maximum distortion constraints (for arbitrary attacks).

**Theorem III.3** Capacity is given by (9) for all four combinations of maximum cost constraints (1) (5) and expected-cost constraints (2) (3) on the transmitter and adversary. In (9), $U$ is a random variable defined over an alphabet $|U| = |S^e||X| + 1$.

**Theorem III.4** For the C-DMC case, the reliability function is upper-bounded by the sphere-packing exponent

$$E_{\text{sp}}^{C-DMC}(R) = \min_{\hat{p}_S \in P_{S^e}} \max_{p_{XU|S^e} \in P_{XU|S^e}(D_1)} \min_{p_Y|X \in A} D(\hat{p}_S p_{XU|S^e} p_Y|X S^a) + J(\hat{p}_S p_{XU|S^e} p_Y|X S^a) \leq R$$

(12)

for all $R_{\infty} \leq R \leq C$, where $R_{\infty}$ is the infimum of all $R$ such that $E_{\text{sp}}(R) < \infty$ in (12).
Theorem III.5 For the AVC case, the reliability function is upper-bounded by the sphere-packing exponent

\[
E_{sp}^{AVC}(R) = \min_{p_S \in \mathcal{P}_S} \max_{p_{UX} \in \mathcal{P}_{UX}} \min_{\tilde{p}_Y|S,U \in \mathcal{A}} \sum_{S,U} I_Y;U:S'|X:SU: \left[ D(\tilde{p}_S|S) - \tilde{p}_S - p_{UX} S'|X|S = \tilde{S} \right] + \tilde{I}_Y;U:S'|X:SU: \left[ \tilde{p}_S - p_{UX} S'|X|S = \tilde{S} \right]
\]

for all \( R_\infty \leq R \leq C \), where \( R_\infty \) is the infimum of all \( R \) such that \( E_{sp}(R) < \infty \) in (13).

Remark III.1 Theorems III.1 and III.4 extend known results [7] for DMC's without side information.

Remark III.2 In the case of an AVC channel without side information, (11) reduces to the straight line \( E_r(R) = C - R \) at all rates below capacity.

Remark III.3 Comparing (10) and (11), it can be shown that \( E_r^{AVC}(R) \leq E_r^{DMC}(R) \) at all rates. This is not too surprising because the proof of Theorem III.1 shows there is no loss in optimality in considering AVC's that are uniform over conditional types, and there are more conditional types to choose from under the C-DMC model. Generally, additional flexibility is beneficial for the adversary, and the worst conditional type does not satisfy the hard constraint (4).

Remark III.4 The error exponents for a standard 2-D random binning scheme in the AVC and C-DMC cases are lower than \( E_r^{AVC}(R) \) and \( E_r^{DMC}(R) \), respectively.

It should also be noted that:
1. the worst type classes \( T_{\ast\ast}, T_{\ast\ast}\ast\ast, T_{\ast\ast\ast} \), and best type class \( T_{\ast\ast\ast\ast} \) (in an appropriate min max min sense) determine the error exponents;
2. the straight-line part of \( E_r(R) \) results from the union bound;
3. random codes are generally suboptimal at low rates.

Theorem III.6 For both the C-DMC and AVC cases, there exists a critical rate \( R_{cr} < C \) such that

\[
E_r(R) = \begin{cases} \frac{D_1}{\delta_2} [h(\delta_2) - h(D_2)], & \text{if } 0 \leq D_1 < \delta_2; \\ h(D_1) - h(D_2), & \text{if } \delta_2 \leq D_1 \leq 1/2; \\ 1 - h(D_2), & \text{if } D_1 > 1/2 \end{cases}
\]

IV Binary-Hamming Case

In this section, we consider a problem where \( S^c = \{0,1\} \), \( S^c \) is a Bernoulli sequence with \( Pr[S^c = 1] = p^f = 1 - Pr[S^c = 0] \), transmission is subject to the cost constraint (1) in which \( \Gamma \) is Hamming distance, and the adversary is subject to the maximum-distortion constraint (5), in which \( d \) is also Hamming distance.

Case I: \( p^f = \frac{1}{2}, S_0 = S^d = \emptyset \). This model is analogous to a public watermarking problem studied in recent literature. Capacity for a fixed-DMC problem (adversary implements a binary symmetric channel (BSC) with crossover probability \( D_2 \)) is given by [14]:

\[
C_{pub} = \begin{cases} \frac{D_1}{\delta_2} [h(\delta_2) - h(D_2)], & \text{if } 0 \leq D_1 < \delta_2; \\ h(D_1) - h(D_2), & \text{if } \delta_2 \leq D_1 \leq 1/2; \\ 1 - h(D_2), & \text{if } D_1 > 1/2 \end{cases}
\]

where \( \delta_2 = 1 - 2^{-h(D_2)} \) and \( h(\cdot) \) is the binary entropy function. It can be shown that \( C_{pub} \) is also the capacity of the C-DMC and AVC defined by the distortion constraints (3) and (5), respectively.

Error exponents in the case \( D_1 = 0.4, D_2 = 0.2 \), are given in Fig. 3. We found numerically that at all rates, the worst attack channel \( p_{Y|X} \) is the BSC with crossover probability \( D_2 \). For both the AVC and C-DMC cases, the worst-case \( \tilde{p}_S \) in (10) and (11) coincide with \( p_S \). Therefore the random-coding error exponent is a straight line in the AVC case.

Case II: \( p^f = \frac{1}{2}, S_0 = \emptyset, S^d = S^c \). This is the private watermarking problem of [4]. Capacity is given by

\[
C_{priv} = h(D_1 + D_2) - h(D_2)
\]

where \( D_1 \neq D_2 \neq D_1 = (1 - D_1) D_2 \). Error exponents in the case \( D_1 = 0.4, D_2 = 0.2 \), are shown in Fig. 4. As in Case I, for both the AVC and C-DMC cases, the worst-case \( \tilde{p}_S \) in (10) and (11) coincide with \( p_S \). The random-coding error exponent is nearly a straight line in the AVC case (there is a small elbow near \( R = C \)). The worst attack channel \( p_{Y|X} \) is the BSC with crossover probability \( D_2 \).

The capacity expression (16) was also derived for the AVC problem of Csiszár and Narayan [15], albeit with different assumptions \( p^f = 0 \), i.e., degenerate side information, and channel state \( \theta \) selected independently of \( X \). Error exponents for the latter problem were derived by Hughes and Thomas [9]. They obtained \( E_r^{AVC}(R) = |C - R|^c \) and \( E_r^{AVC}(R) = \infty \) at all rates below capacity.

Case III. Degenerate side information: \( p_k = 0, S^c = S_0 = S^d = \emptyset \). Unlike Case I and Case II, the worst attack is an asymmetric binary channel, favoring outputs with low Hamming weight. Error exponents in the case \( D_1 = 0.4, D_2 = 0.2 \), are given in Fig. 5. In this case, note that Somekh-Baruch and Merhav's formula (50), (25) in [5] for the sphere-packing error exponent reduces to minimizing

3This can be proven analytically in the AVC case.
the function zero over a constrained set. At first sight this leads to a contradiction. However, the feasible set for their minimization problem is empty, in which case the sphere-packing bound is useless. The same phenomenon was observed and discussed in [9].

![Figure 3: Error exponents when $p^c = \frac{1}{2}$, $S^a = S^d = \emptyset$ (public watermarking).](image3.png)

![Figure 4: Error exponents when $p^c = \frac{1}{2}$, $S^a = \emptyset$, $S^c = S^d$ (private watermarking).](image4.png)

In all three cases considered above, the gap between the random-coding and sphere-packing exponents in the C-DMC case is small, except at very low rates. Comparing Figs 3 and 4, we see that error exponents are larger in the latter case (for the C-DMC but not for the AVC scenario). The gap quantifies the cost incurred by the decoder for having no side information available.

**References**


