On Joint Source–Channel Coding for the Wyner–Ziv Source and the Gel'fand–Pinsker Channel *

Neri Merhav and Shlomo Shamai (Shitz) †

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL
{merhav,sshlomo}@ee.technion.ac.il

June 25, 2003

Abstract

We consider the problem of lossy joint source–channel coding in a communication system where the encoder has access to channel state information (CSI) and the decoder has access to side information that is correlated to the source. This configuration combines the Wyner–Ziv model of pure lossy source coding with side information at the decoder and the Shannon/Gel'fand–Pinsker model of pure channel coding with CSI at the encoder. We prove a separation theorem for this communication system, which asserts that there is no loss in asymptotic optimality in applying first, an optimal Wyner-Ziv source code and then, an optimal Gel'fand–Pinsker channel code. We then derive conditions for the optimality of a symbol–by–symbol (scalar) source–channel code, and demonstrate situations where these conditions are met. Finally, we discuss a few practical applications, including overlaid communication where the model under discussion is useful.

Index Terms: Wyner–Ziv coding, Gel'fand–Pinsker coding, side information, channel state information, separation theorem, joint source–channel coding.

*This paper was presented in part at the 40th Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 2002.

†The work of S. Shamai was supported by the Israel Science Foundation (ISF).

1 Introduction

The Wyner–Ziv (W–Z) model of source coding with side information at the decoder (see, e.g., [25], [31], [33], [41], [42], [44]) and the Gel'fand–Pinsker (G–P) model for channel coding with channel state information at the encoder (see, e.g., [1], [2], [5], [12], [15], [18], [19], [22], [23], [24], [26], [32], [36], [37], [38], [40], [43]) as well as the duality between them (see, e.g., [3], [4], [8], [28], [34], [35]) have attracted considerable attention of information theorists over the years.

In this paper, we make one more step in this avenue of the relation and the duality between the two models by combining them, and studying a lossy joint source–channel coding system whose encoder has the channel state information (CSI) available (either causally or non-causally) and whose decoder has access to side information that is correlated to the source (see Fig. 1). This model generalizes also to the setting where both the encoder and decoder have access to different versions of side informations on both the channel state and the source input (see Fig. 2).

Besides this theoretical motivation of enhancing the relation and duality between W–Z source coding and G–P channel coding, it turns out that this combined W–Z/G–P source-channel model has practical applications. One of them is the possible use of systematic codes for writing into a memory device with defects [23],[24],[38], where the systematic part of the code corresponds to uncoded (noisy) side information at the decoder [31]. Another application is related to overlaid back–compatible communication, which will be elaborated on in Section 5, along with a few other examples of applications.

Our main result is a separation theorem for this communication system, which asserts that there is no loss in asymptotic optimality if one applies first, an optimal W–Z source code regardless of the channel, and then, an optimal Shannon/G–P channel code (depending on whether the CSI is causal or not) regardless of the source. It should be noted that the existence of a separation theorem for this system is not a–priori obvious for two reasons, one is related to the operative aspect of the system, and the other is more on the mathematical–technical aspect. As for the operative aspect, it would be natural to expect that the source encoder could benefit from the available CSI (e.g., fading) to adapt the degree of compression to the condition of the channel. Similarly, one might speculate that the channel decoder would take advantage of the available side information correlated to the source. The separation theorem tells us, nonetheless, that there is no loss in optimality if theses informations are ignored altogether by both devices. On the technical aspect, note that the information streams along the main communication link (source → channel input → channel output→ destination) do not admit a standard Markov structure that lends itself to the data processing theorem, like in the classical case. Moreover, both the W–Z rate distortion problem and the G–P channel coding problem include auxiliary random variables which convey strong coupling between the source coding part and the channel coding part, as will be seen in the sequel. This coupling does not exist in the classical setting where no side inforamtion is available. It should be pointed out that for the special case where the main channel (i.e., the upper channel in Fig. 1) is a simple discrete memoryless channel (DMC), a separation theorem was already stated and proved in [31, Theorem 2.1], but the fact that the G–P channel also obeys a separation theorem has not been established before, to the best of our knowledge, even for the ordinary discrete memoryless source (DMS), let alone the DMS with correlated side information at the decoder considered here. In this respect, it should be emphasized that there is a substantial difference between the situation where the CSI is available at both encoder and decoder (in which case, the separation theorem is straightforward) and the Gel'fand–Pinsker setting considered here, where the CSI is available at the encoder only.

Following known results on the forms of capacity–achieving distributions and optimal rate–distortion test channels [14, Chap. 2.3, Theorem 3.4 and Problem 3], (see also Gallager [20, p. 460], as well as Gastpar, Rimoldi, and Vetterli [21]), we then furnish conditions under which a simple, symbol–by–symbol joint source–channel code is optimal in the sense of attaining the joint source–channel distortion bound. We also construct some examples of such systems. This is a point where yet another aspect of the duality between W–Z source coding and G–P channel coding plays a role: one of the conditions for optimality of a joint source–channel code is that the random variable (RV) that represents the source is the optimal auxiliary RV that attains the capacity of the G–P channel, and that the G–P channel output is an optimal auxiliary RV that attains the W–Z rate–distortion function. In other words, the W–Z source and the G–P channel are matched and the auxiliary RV's of both play an operative role.

Finally, as mentioned earlier, we describe, in some detail, a few particular applications that demonstrate the usefulness of combining W–Z coding with G–P coding. These examples include the binary symmetric channel (BSC) with the error state available at the transmitter, systematic coding for defective memories, and overlaid back–compatible communication systems over the binary and the Gaussian channels.

2 Notation and Problem Formulation

Throughout this paper, scalar RVs will be denoted by capital letters, their sample values will be denoted by the respective lower case letters, and their alphabets will be denoted by the respective calligraphic letters. A similar convention will apply to random vectors and their sample values, which will be denoted with same symbols superscripted by the dimension. Thus, for example, W^k will denote a random k-vector $(W_1, ..., W_k)$, and $w^k = (w_1, ..., w_k)$ is a specific vector value in $\mathcal{W}^k$, the k-th Cartesian power of $\mathcal{W}$. The notations w_i^j and W_i^j, where i and j are integers and $i \leq j$, will designate segments $(w_i, \ldots, w_j)$ and $(W_i, \ldots, W_j)$, respectively, where for $i = 1$, the subscript will be omitted (as above). For $i > j$, w_i^j (or W_i^j) will be understood as the null string. Sequences without specifying indices are denoted by $\{\cdot\}$.

Sources and channels will be denoted generically by the letter P subscripted by the name of the RV and its conditioning, if applicable, e.g., $P_U(u)$ is the probability function of U at the point $U = u$, $P_{Z|S}(z|s)$ is the conditional probability of $Z = z$ given $S = s$, and so on. Whenever clear from the context, these subscripts will be omitted. Information theoretic quantities like entropies, divergences, and mutual informations will be denoted following the usual conventions of the information theory literature, e.g., $H(U^N)$, $I(Z^n; W^k)$, $D(P_{Y|XS}\|P_Y)$, and so on.

Consider the communication system depicted in Fig. 1: A source P_{UV}, henceforth referred to as the *W–Z source*, generates independent copies, $\{(U_i, V_i)\}_{i=1}^{\infty}$, of a pair of dependent, finite–alphabet RV's $(U, V) \in \mathcal{U} \times \mathcal{V}$, and operates at the rate of ρ_s symbol pairs per second. Let $N = \rho_s T$ be a positive integer, where T is the duration of the block in seconds. The block $U^N = (U_1, \ldots, U_N)$, of the first component of the source, is fed into a joint source–channel encoder, whereas the corresponding block of the other component, $V^N = (V_1, \ldots, V_N)$, is fed, as side information, into the decoder whose aim is to provide an estimate of U^N, denoted $\hat{U}^N = (\hat{U}_1, \ldots, \hat{U}_N)$, whose components take values in a finite reproduction alphabet $\hat{\mathcal{U}}$. The quality of decoder output, $\hat{U}^N$, is judged with respect to (w.r.t.) the fidelity criterion which is the expectation of

$$d(U^N,\hat{U}^N)=\sum_{i=1}^{N} d(U_i,\hat{U}_i), \tag{1}$$

where $d(u,\hat{u})\geq 0$, $u\in\mathcal{U}$, $\hat{u}\in\hat{\mathcal{U}}$, is a given single–letter distortion function. The conditional probability of V^N given U^N,

$$P_{V^N|U^N}(v^N|u^N)=\prod_{i=1}^{N} P_{V|U}(v_i|u_i), \tag{2}$$

will be referred to as the *W–Z channel* (see Fig. 1).

At the same time duration of T seconds, a memoryless channel, henceforth referred to as the *G–P channel*, which operates at the rate of ρ_c channel uses per second, works as follows: The channel input is a vector pair $(X^n,S^n)=((X_1,S_1),\ldots,(X_n,S_n))$, where $n=\rho_c T$ is a positive integer, and where each X_i and S_i take values in finite sets, $\mathcal{X}$ and $\mathcal{S}$, respectively. The channel output is a vector $Y^n=(Y_1,\ldots,Y_n)$, whose components take values in a finite set $\mathcal{Y}$, and the conditional probability of Y^n given (X^n,S^n) is characterized by

$$P_{Y^n|X^nS^n}(y^n|x^n,s^n)=\prod_{i=1}^{n} P_{Y|XS}(y_i|x_i,s_i). \tag{3}$$

The vector $X^n=(X_1,\ldots,X_n)$ is referred to as the channel input, whereas $S^n=(S_1,\ldots,S_n)$ is referred to as the channel state sequence, which is governed by another discrete memoryless process:

$$P_{S^n}(s^n)=\prod_{i=1}^{n} P_S(s_i), \tag{4}$$

independently of (U^N,V^N). It is also assumed that $V^N\to U^N\to Y^n$ is a Markov chain, guaranteeing independence between the W–Z channel and the G–P channel (see Fig. 1). The channel input may be subjected to a transmission–cost constraint

$$E\left\{\sum_{i=1}^{n}\phi(X_i)\right\}\leq n\Gamma, \tag{5}$$

where ϕ is a given function from $\mathcal{X}$ to $\mathbb{R}^+$ and $\Gamma\geq 0$ is a prescribed value. In the absence of such a constraint, one may simply set $\Gamma=\infty$.

The joint source–channel encoder implements a (possibly randomized[1]) function $x^n=f(u^N,s^n)$. In the case of causal state information, each x_i depends only on u^N, x^{i-1}, and s^i. The decoder is defined by a deterministic function $\hat{u}^N=g(v^N,y^n)$.

[1]A randomized encoder means a conditional distribution $P(x^n|u^N,s^n)$. Due to the transmission constraint, it is not a–priori clear that the optimum encoder would be deterministic in general.

Definition 1 *A distortion level D is said to be achievable if for every $\epsilon > 0$, there exist sufficiently large n and N, with $n/N = \rho_c/\rho_s$, an encoder $f : \mathcal{U}^N \times \mathcal{S}^n \to \mathcal{X}^n$, and a decoder $g : \mathcal{V}^N \times \mathcal{Y}^n \to \hat{\mathcal{U}}^N$ such that eq. (5) is satisfied and*

$$E\left\{\sum_{i=1}^{N} d(U_i, \hat{U}_i)\right\} \leq N(D + \epsilon). \tag{6}$$

The main problem we address, in this paper, is the characterization of the minimum achievable distortion level of this system.

Comment: One might also consider a seemingly more general model where the CSI is partially available at the decoder as well. Specifically, the decoder has additional access to $\tilde{S}^n$, which is generated by yet another DMC fed by S^n. However, this falls within the framework of the current model where the pair $(Y^n, \tilde{S}^n)$ is redefined as the channel output $\tilde{Y}^n$. To symmetrize the model, the encoder may also be assumed to access only a noisy version of the CSI $\hat{S}^n$: Again, this falls again within the framework of our model if the original channel is replaced by

$$P_{\tilde{Y}|X\hat{S}}(\tilde{y}|x, \hat{s}) = \sum_{s} P_{S|\hat{S}}(s|\hat{s}) P_{\tilde{Y}|XS}(\tilde{y}|x, s). \tag{7}$$

By the same token, our model is also general enough to include a situation where the encoder has access to a noisy version $\tilde{V}^N$ of the side information V^N seen by the decoder (created by another, memoryless feedback channel). This is done simply by redefining the source as $(U^N, \tilde{V}^N)$, yet the distortion measure continues to depend only on the first component, i.e., $d'((u, \tilde{v}), \hat{u}) = d(u, \hat{u})$. In summary, our model actually covers a symmetric situation, depicted in Fig. 2, where both encoder and decoder have access to (possibly different) noisy versions of both source–state information and channel–state information.

3 Separation Theorem

In order to state the separation theorem of lossy joint source–channel coding for the W–Z source and the G–P channel defined in Section 2, we will define the following functional of a joint distribution P_{UA} for a generic RV A:

$$\Delta(U|A) = \min_{g:\mathcal{A}\to\hat{\mathcal{U}}} E\{d(U, g(A))\}, \tag{8}$$

and recall that the W–Z rate–distortion function [42] of P_{UV} w.r.t. distortion measure $d(\cdot,\cdot)$, is given by

$$R_{WZ}(D) = \min[I(U;Z) - I(V;Z)] \equiv \min I(U;Z|V) \tag{9}$$

where the minimum is over all auxiliary RV's Z such that $Z \to U \to V$ is a Markov chain and $\Delta(U|V,Z) \leq D$.

As for the channel, we recall that the capacity formula (see, [3][2] and [22]) for the G–P channel $\{P_{Y|XS}, P_S\}$, under the transmission–cost constraint, is given by

$$C_{GP}(\Gamma) = \max[I(W;Y) - I(W;S)], \tag{10}$$

where the maximization is over all pairs of RV's (W,X) such that $W \to (X,S) \to Y$ is a Markov chain and $E\phi(X) \leq \Gamma$. Clearly, as P_S and $P_{Y|XS}$ are given, the degrees of freedom are in the optimal choice of $P_{XW|S} = P_{W|S} \times P_{X|WS} = P_{W|S} \times 1_{\{X=f(W,S)\}}$ subject to the transmission–cost constraint.[3] The capacity for the case where the state sequence S^n is revealed to the encoder causally [32] can be obtained [10],[16] from eq. (10) by imposing the additional constraint that W is independent of S, namely, $P_{XW|S} = P_W \times 1_{\{X=f(W,S)\}}$. In this case, (10) will be denoted by $C_S(\Gamma)$, where the subscript S stands for "Shannon." Note that the term $I(W;S)$ on the right–hand side of (10) vanishes in this case.

Our main result is the following separation theorem for the case where the encoder is noncausal w.r.t. the state sequence. For the causal case, C_{GP} should be replaced by C_S.

Theorem 1 *Under the assumptions described in Section 2, a necessary and sufficient condition for D being an achievable distortion level is*

$$\rho_s R_{WZ}(D) \leq \rho_c C_{GP}(\Gamma).$$

Proof. The proof of the sufficiency part comes, like in the classical case, from considering an asymptotically optimal source code (independent of the channel) followed by a reliable transmission code for the channel (independent of the source) whose rate is close to capacity: If the distortion level of the W-Z source code is chosen such that $\rho_s R_{WZ}(D) < \rho_c C_{GP}(\Gamma)$,

[2] In [3], a somewhat more general result is proved in the context of information embedding, where the encoder is subjected to a distortion constraint $E\{\sum_{i=1}^n d(S_i,X_i)\} \leq nD$. For our purposes, we set $d(s,x) = \phi(x)$.

[3] Here, $1_{\{X=f(W,S)\}}$ means a degenerate conditional distribution that puts all its mass on the point $X = f(W,S)$ for some deterministic function f. In [9, Lemma B.1], it is shown that even in the presence of a transmission–cost constraint, the optimal encoder is deterministic.

one may select two constants R_s and R_c such that $NR_{WZ}(D) < NR_s = nR_c < nG_{GP}(\Gamma)$, compress the source into R_s bits per–symbol within distortion D, and then map the resulting NR_s–bit codeword into a channel codeword of the same number of bits, nR_c. Since $R_c < C_{GP}(\Gamma)$, there exists a reliable G–P channel code which causes asymptotically negligible additional distortion. Since D can be chosen, in this way, such that $\rho_s R_{WZ}(D)$ is arbitrarily close to $\rho_c C_{GP}(\Gamma)$, every distortion level for which $\rho_s R_{WZ}(D) \leq \rho_c C_{GP}(\Gamma)$ is achievable. Obviously, in the causal case, all the above continues to hold provided that C_{GP} is replaced by C_S.

The proof of the necessity part is by a simple fusion of the proofs of the converse theorems in [22] and in [42] (or, more simply, in [31]) with some minor modifications. The idea is to upper bound $I(U^N; Y^n)$ by $nC_{GP}(\Gamma)$ and to lower bound it by $NR_{WZ}(D)$. Clearly, the combined inequality, $NR_{WZ}(D) \leq nC_{GP}(\Gamma)$, with both sides divided by T, is the assertion of Theorem 1.

As for the upper bound to $I(U^N; Y^n)$, we have:

$$I(U^N; Y^n) = I(U^N; Y^n) - I(U^N; S^n) \leq \sum_{i=1}^{n} [I(W_i; Y_i) - I(W_i; S_i)], \qquad (11)$$

where the equality is due to the independence between U^N and S^n, and where the inequality is proved exactly as in [22] with W_i being defined as $(U^N, Y^{i-1}, S_{i+1}^n)$. Thus, the message V, of the proof of the converse theorem of [22], is simply replaced by U^N. Since the remaining part of the proof in [22] uses only the the fact that S^n is drawn by a DMS, and general chain rules of the mutual information, it is general enough to continue to hold in our case. It should be pointed out that the inequality in (11) becomes an equality if and only if the components of Y^n are statistically independent. Now, since $W_i \to (X_i, S_i) \to Y_i$ is a Markov chain, the right–most side of eq. (11) is in turn upper bounded by $nC_{GP}(\Gamma)$, similarly as in [3] and [22]. In the causal case, it should be noted that W_i is independent of S_i. Therefore, the maximization over W is carried out with the additional constraint that W is independent of S, resulting in $nC_S(\Gamma)$.

As for the lower bound to $I(U^N; Y^n)$, we follow the proof of Theorem 2.1 of [31], starting from the second line of the chain of inequalities (2.7) therein. Specifically, defining $Z_i \stackrel{\Delta}{=}$

$(U^{i-1}, V^{i-1}, V_{i+1}^N, Y^n)$, we have the following:

$$\begin{aligned}
I(U^N; Y^n) &\stackrel{\text{(a)}}{=} I(U^N, V^N; Y^n) \\
&\stackrel{\text{(b)}}{\geq} I(U^N; Y^n | V^N) \\
&\stackrel{\text{(c)}}{=} \sum_{i=1}^{N} I(U_i; Y^n | U^{i-1}, V^N) \\
&\stackrel{\text{(d)}}{=} \sum_{i=1}^{N} [I(U_i; Y^n, U^{i-1}, V^{i-1}, V_{i+1}^N | V_i) - I(U_i; U^{i-1}, V^{i-1}, V_{i+1}^N | V_i)] \\
&\stackrel{\text{(e)}}{=} \sum_{i=1}^{N} I(U_i; Z_i | V_i) \\
&\stackrel{\text{(f)}}{\geq} \sum_{i=1}^{N} R_{WZ}(\Delta(U_i | V_i, Z_i)) \\
&\stackrel{\text{(g)}}{\geq} \sum_{i=1}^{N} R_{WZ}(\Delta(U_i | V^N, Y^n)) \\
&\stackrel{\text{(h)}}{\geq} \sum_{i=1}^{N} R_{WZ}(Ed(U_i, \hat{U}_i)) \\
&\stackrel{\text{(i)}}{\geq} N R_{WZ}(D + \epsilon),
\end{aligned} \tag{12}$$

where (a) follows from the Markovity of $V^N \to U^N \to Y^n$, (b),(c), and (d) – from the chain rule of the mutual information, (e) – from the memorylessness of P_{UV}, (f) – from the fact that $Z_i \to U_i \to V_i$ is a Markov chain and from eq. (9), (g) – from the monotonicity of $R_{WZ}(\cdot)$ and eq. (8), (h) – from eq. (8) and the fact that $\hat{U}_i$ is a function of (Y^n, V^N), and (i) – from the convexity and monotonicity of $R_{WZ}(\cdot)$ [42], and the assumption that D is achievable. Since $\epsilon > 0$ can be chosen arbitrarily small, the result follows from the continuity of $R_{WZ}(\cdot)$, which in turn, follows again by convexity for $D > 0$, and established separately for $D = 0$ in [42] (see last paragraph of Subsection 1C, eq. (15b), and Theorem 1 therein). □

Comment 1: Note that in the above chain, (b) becomes equality if Y^n is independent of V^N, and the remaining inequalities become equalities if each Z_i is an optimal auxiliary RV for $R_{WZ}(\cdot)$. The former condition is intuitively appealing because it means that a necessary condition for optimality is that the two information streams that the decoder receives are independent, for otherwise there would be some waste on redundant information.

Comment 2: Though formally we confine attention to finite alphabets and finite sets of

(U, S, X, Y, V), since the basic W–Z rate distortion function and the G–P channel capacity extend, under minor regularity conditions, to general alphabets and sets (see [45] and references therein) and since our proof relies essentially on the very same arguments, Theorem 1 applies to those general settings as well.

Comment 3: As was already pointed out in the Introduction, Theorem 1 has an additional degree of depth relative to the classical separation theorem, and this is because of the role of the side information: It tells us that without loss of optimality, the source encoder ignores the channel state information, and analogously, the channel decoder ignores the partial information on the source.

It is also interesting to put Theorem 1 in the perspective of another recent result [37], where the GP–channel is studied in terms achievable tradeoffs between reliable coding rates (of digital information) and distortion levels in estimating the channel–state sequence. Considering our model for the case $\rho_c = \rho_s$, the combined channel $P(Y, V|X, U, S) = P(Y|X, S)P(V|U)$ can be thought of as a G–P channel whose input is X, its state is (U, S), and its output is (Y, V). Our problem corresponds to the extreme case of the tradeoff investigated in [37], where the desired coding rate is $R = 0$ and the distortion measure regarding the (extended) state (U, S) depends only on the component U. This continues to hold even if the W–Z channel depends on S, namely, it becomes $P(V|U, S)$, which is the case, for example, when the G–P channel and the W–Z channel are physically the same (and the latter corresponds to the systematic part of the transmission). The results in [37], however, are achievability results which are known to be tight only in certain special cases.

4 Symbol–by–Symbol Joint Source–Channel Codes

Even in the classical model, without side information, it is well known that the cost of keeping the optimality of separate source and channel coding is, in general, associated with long blocks, which mean high complexity and long delay. On the other hand, there are also well–known examples of source–channel pairs, operating at the same rate ($\rho_c = \rho_s$), which match each other so well that the joint source–channel distortion bound can be obtained by direct connection of the source to the channel, with no coding at all, or with very simple coding on a scalar, symbol-by-symbol basis ($n = N = 1$). The two classical examples are: (i) the binary memoryless source, with the Hamming distortion measure, and the binary symmetric channel (BSC), where the distortion level equals the crossover probability, and (ii) the Gaussian memoryless source, with the mean–square distortion measure, and the Gaussian power–restricted memoryless channel.

In [14, Chap. 2.3] (see also [21]), this issue of perfect matching between the source and the channel has been studied, and conditions have been furnished for the optimality of a given symbol–by–symbol coding system w.r.t. a given distortion measure and a given transmission–cost function.

In this section, we extend these results from the classical model, without side information, to our model of a W–Z source and a G–P channel. It should be noted that whenever $C_{GP}(\Gamma)$ is strictly larger[4] than $C_S(\Gamma)$, no scalar encoder $X_i = f(U_i, S_i)$ can achieve the former and the only hope is to achieve $C_S(\Gamma)$. Moreover, even if this encoder is allowed to have a *noncausal* access the *entire* state sequence across some block, i.e., $X_i = f_i(U_i, S^n)$, $1 \leq i \leq n$, still $C_S(\Gamma)$ cannot be exceeded since the source is assumed independent of the channel state process, and so, the access to past and future states cannot improve performance. (Another way to see this is to observe that for any realization of (S^{i-1}, S_{i+1}^n), X_i is given by a particular function of (U_i, S_i), whereas the action of the channel at time instant i is insensitive to (S^{i-1}, S_{i+1}^n)).

Consider again the system depicted in Fig. 1, now for the scalar case of block–length $n = N = 1$. The encoder implements a function $x = f(u, s)$ and the decoder is given by $\hat{u} = g(v, y)$. We say that an encoder–decoder pair (f, g) is optimal w.r.t. d if it satisfies the transmission–cost constraint, $E\phi(X) \leq \Gamma$, and it meets the joint source–channel bound with equality, i.e.,

$$R_{WZ}(Ed(U, g(V, Y))) = C_S(\Gamma). \tag{13}$$

Throughout this section, we will always assume the encoder implements an optimal[5] function $x = f^*(u, s)$ that achieves $C_S(\Gamma)$ subject to the constraint $E\phi(X) \leq \Gamma$ because this is obviously a necessary condition for optimality. In the next theorem, we give the additional conditions, which together with the condition $x = f^*(u, s)$, are sufficient for optimality.

Theorem 2 *If all of the following conditions are satisfied, then (f^*, g) is optimal w.r.t. d:*

[4] See [18] for examples where $C_{GP}(\Gamma) = C_S(\Gamma)$.

[5] There is no guarantee that $C_S(\Gamma)$ is attained uniquely by one pair of input distribution and encoding function.

(a) The alphabet $\mathcal{U}$ is large enough to achieve $C_S(\Gamma)$.

(b) Either: (i) $I(U;Y) = C_S(\infty)$ yet $E\phi(X) \leq \Gamma$, or, (ii) $I(U;Y) < C_S(\infty)$ and there exist a positive real α and a constant β such that for every $u \in \mathcal{U}$,

$$D(P_{Y|U=u}\|P_Y) = \alpha E\{\phi(X)|U = u\} + \beta.$$

(c) Y and V are statistically independent.

(d) Either: (i) $I(U;Y|V) = 0$ and g attains $\Delta(U|V)$, or, (ii) $I(U;Y|V) > 0$, g attains $\Delta(U|V,Y)$, and there exist a positive real γ and a function $\delta(u)$ such that for all $(u,y) \in \mathcal{U} \times \mathcal{Y}$:

$$\sum_v P_{V|U}(v|u)d(u, g(v,y)) = -\gamma \log P_{U|Y}(u|y) + \delta(u).$$

Proof. Consider the chains of equalities/inequalities (11) and (12) in the proof of the necessity part of Theorem 1, and confine attention to the degenerate case $n = N = 1$. For a scalar system to be optimal, all inequalities should become equalities. Let us begin with eq. (11). The inequality in eq. (11) boils down to a trivial identity since W_1, in this case, is identical to U. For the right–most side of eq. (11) to achieve $C_S(\Gamma)$, U must be an optimal auxiliary RV for this capacity, which means, first of all, that (a) should be satisfied. Since the constraint of independence between U and S is automatically satisfied by the model assumption, it remains to show that condition (b) guarantees that U maximizes $I(U;Y)$ subject to constraint $E\phi(X) \leq \Gamma$. If $I(U;Y) = C_S(\infty)$ (yet the transmission constraint is not violated), this is obviously the case. If $I(U;Y) < C_S(\infty)$, the alternative condition in (b) can be proved to be sufficient by a straightforward extension of [21, Lemma 3], which for the sake of completeness, will be rederived here. Let $P_{Y|U}$ be the channel from U to Y that is induced by P_S, f^* and $P_{Y|XS}$. Let

$$I_P(u) = D(P_{Y|U=u}\|P_Y), \tag{14}$$

and let $\tilde{P} = \tilde{P}_U$ denote an alternative distribution on U. Denoting by $I_{\tilde{P}}(U;Y)$ the mutual information between U and Y, where U is distributed according to $\tilde{P}$, we have:

$$\begin{aligned}\sum_u \tilde{P}_U(u)I_P(u) - I_{\tilde{P}}(U;Y) &= \sum_u \tilde{P}_U(u)[I_P(u) - I_{\tilde{P}}(u)] \\ &= \sum_u \tilde{P}_U(u)[D(P_{Y|U=u}\|P_Y) - D(P_{Y|U=u}\|\tilde{P}_Y) \\ &= D(\tilde{P}_Y\|P_Y) \geq 0, \end{aligned} \tag{15}$$

where $\tilde{P}_Y$ denotes the channel output marginal induced by $\tilde{P}_U$. Since equality is obviously attained when $\tilde{P}_U = P_U$, it follows that

$$I_{\tilde{P}}(U;Y) - I_P(U;Y) \leq \sum_u [\tilde{P}_U(u) - P_U(u)] I_P(u). \tag{16}$$

Next, assume that $\tilde{P}_U$ is such that

$$E_{\tilde{P}}\phi(X) \leq E_P\phi(X), \tag{17}$$

where E_P and $E_{\tilde{P}}$ are expectations induced by P and $\tilde{P}$, respectively. Then,

$$\begin{aligned}
& I_P(U;Y) - I_{\tilde{P}}(U;Y) \\
\geq\ & \sum_u [P_U(u) - \tilde{P}_U(u)] I_P(u) \\
\geq\ & \sum_{u,s} P_S(s)[P_U(u) - \tilde{P}_U(u)][I_P(u) - \alpha\phi(f^*(u,s))] \\
=\ & \sum_u [P_U(u) - \tilde{P}_U(u)][I_P(u) - \alpha E\{\phi(X)|U=u\}] \\
=\ & \beta \sum_u [P_U(u) - \tilde{P}_U(u)] = 0,
\end{aligned} \tag{18}$$

where the first inequality is (16), the second inequality is (17) and the second equality follows from condition (b). This proves that P_U maximizes $I(U;Y)$ given the transmission constraint.

Consider next the chain of inequalities (12) with $n = N = 1$ and $\epsilon = 0$. For inequality (12-b) to become an equality, Y and V must be independent (cf. the ending comment of Section 3), which is condition (c), (12-f) becomes an equality if $Z = Y$ is an optimal auxiliary RV for the W–Z rate–distortion function,[6] (12-g) is an identity in the case $n = N = 1$, and (12-h) is an equality if the decoder g achieves $\Delta(U|V,Y)$. In the degenerate case of $I(U;Y|V) = 0$, Y trivially minimizes $I(U;Y|Z)$ under this Markovity constraint. In this case, Y is irrelevant to decoding, and the best decoding based on V is, by definition, the one that achieves $\Delta(U|V)$, which is the smallest distortion level at which $R_{WZ}(D) = 0$. For the case $I(U;Y|V) > 0$, we next prove, using the same technique as in [21], that Y is optimal if condition (d) holds. Let $P_{Z|U}$ and $\tilde{P}_{Z|U}$ be two channels from U to Z and let

[6]Notably, $V \to U \to Y$ is always a Markov chain.

$I_P(U;Z|V)$ and $I_{\tilde{P}}(U;Z|V)$ the corresponding induced mutual informations. Then,

$$
\begin{aligned}
& I_{\tilde{P}}(U;Z|V) - \sum_{u,v,z} P_{UV}(u,v)\tilde{P}_{Z|U}(z|u)\log\frac{P_{Z|U}(z|u)}{P_{Z|V}(z|v)} \\
= & \sum_{u,v,z} P_{UV}(u,v)\tilde{P}_{Z|U}(z|u)\left[\log\frac{\tilde{P}_{Z|U}(z|u)}{\tilde{P}_{Z|V}(z|v)} - \log\frac{P_{Z|U}(z|u)}{P_{Z|V}(z|v)}\right] \\
= & \sum_{u,v,z} P_{UV}(u,v)\tilde{P}_{Z|U}(z|u)\left[\log\frac{\tilde{P}_{U|VZ}(u|v,z)}{P_{U|V}(u|v)} - \log\frac{P_{U|VZ}(u|v,z)}{P_{U|V}(u|v)}\right] \\
= & \sum_{u,v,z} P_{UV}(u,v)\tilde{P}_{Z|U}(z|u)\log\frac{\tilde{P}_{U|VZ}(u|v,z)}{P_{U|VZ}(u|v,z)} \\
= & \sum_{v,z} \tilde{P}_{VZ}(v,z)\sum_{u}\tilde{P}_{U|VZ}(u|v,z)\log\frac{\tilde{P}_{U|VZ}(u|v,z)}{P_{U|VZ}(u|v,z)} \geq 0,
\end{aligned}
\tag{19}
$$

with equality whenever $\tilde{P} = P$. Now, let $\tilde{P}_{Z|U}$ be a channel for which

$$
E_{\tilde{P}}\{d(U,g(V,Z))\} \leq E_P\{d(U,g(V,Z))\}, \tag{20}
$$

where E_P and $E_{\tilde{P}}$ are expectations induced by P and $\tilde{P}$, respectively. For any $\gamma > 0$, we have:

$$
\begin{aligned}
& \gamma[I_{\tilde{P}}(U;Z|V) - I_P(U;Z|V)] \\
\geq & \ \gamma\sum_{u,v,z} P_{UV}(u,v)[\tilde{P}_{Z|U}(z|u) - P_{Z|U}(z|u)]\log\frac{P_{Z|U}(z|u)}{P_{Z|V}(z|v)} \\
\geq & \sum_{u,v,z} P_{UV}(u,v)[\tilde{P}_{Z|U}(z|u) - P_{Z|U}(z|u)]\left[\gamma\log\frac{P_{Z|U}(z|u)}{P_{Z|V}(z|v)} + d(u,g(v,z))\right]
\end{aligned}
\tag{21}
$$

where the first inequality follows from (19) and the second inequality follows from (20). A channel $P_{Z|U}$ is then optimal if the last expression is non–negative. In particular, letting $Z = Y$ and choosing γ and $\delta(u,v)$ so as to satisfy condition (d), we have:

$$
\begin{aligned}
& \gamma[I_{\tilde{P}}(U;Y|V) - I_P(U;Y|V)] \\
\geq & \sum_{u,y} P_{UV}(u,v)[\tilde{P}_{Y|U}(y|u) - P_{Y|U}(y|u)][\gamma\log\frac{P_{Y|U}(y|u)}{P_{Y|V}(y|v)} + d(u,g(v,y))] \\
= & \sum_{u,y} P_U(u)[\tilde{P}_{Y|U}(y|u) - P_{Y|U}(y|u)][\gamma\log\frac{P_{Y|U}(y|u)}{P_Y(y)} + \sum_v P_{V|U}(v|u)d(u,g(v,y))] \\
= & \sum_{u,y} P_U(u)[\tilde{P}_{Y|U}(y|u) - P_{Y|U}(y|u)][\gamma\log P_{U|Y}(u|y) - \gamma\log P_U(u) + \sum_v P_{V|U}(v|u)d(u,g(v,y))] \\
= & \sum_{u,y} P_U(u)[\tilde{P}_{Y|U}(y|u) - P_{Y|U}(y|u)][\delta(u) - \gamma\log P_U(u)] = 0,
\end{aligned}
\tag{22}
$$

where the first equality is due to condition (c), the second to the last equality is due to condition (d), and the last equality is due to the fact that $P_{V|U}(v|u)$, $P_{Y|U}(\cdot|u)$ and $\tilde{P}_{Y|U}(\cdot|u)$ are all probability mass functions for every u. This completes the proof of Theorem 2. $\square$

We conclude this section with two examples of simple communication systems and show how their optimality is verified using Theorem 2.

Example 1. Let $U = (U_1, U_2)$ be a pair[7] of independent binary symmetric RV's taking values in $\{0,1\}$ with a distortion measure

$$d(u, \hat{u}) = d_H(u_1, \hat{u}_1) + \theta d_H(u_2, \hat{u}_2), \tag{23}$$

where $\theta > 0$ and d_H designates the Hamming distance. The W–Z channel is given by

$$V = U_1 \oplus W \tag{24}$$

where W is a binary $\{0,1\}$ RV, independent of U, with $P_W(1) = D$, $0 < D < 1/2$. The G–P channel is given by

$$Y = X \oplus S \oplus W', \tag{25}$$

where X, S and W' are binary RV's, W' being independent of S, X, U, and W, and we have $P_S(1) = 1/2$, $P_{W'}(1) = D$, where D is as before. The transmission cost function is $\phi(x) = x$ and its allowable level is $\Gamma = 1/2$.

Consider the encoder $X = U_2 \oplus S$ (which satisfies the transmission constraint) and the decoder $\hat{U} = (V, Y)$. To see that this encoder achieves capacity, we observe that in this case,

$$I(U;Y) = I(U_2;Y) = I(U_2; U_2 \oplus W') = 1 - H_2(D), \tag{26}$$

$H_2(\cdot)$ being the binary entropy function, whereas the capacity cannot exceed this value because it corresponds also to the case where the CSI is available to the decoder as well. As for the conditions stated in Theorem 2, we have the following: Condition (a) is satisfied because when the channel output is binary, the input alphabet $\mathcal{U}$ need not be richer than binary in order to maximize the mutual information. Condition (b) is satisfied since $I(U;Y) = 1 - H_2(D) = C_S(\infty)$. Note also that due to the symmetry, both $D(P_{Y|U=u} \| P_Y)$

[7]This pair may represent the binary expansion of a uniformly distributed, four-valued RV, i.e., $U = 2U_1 + U_2$. Accordingly, the distortion measure in eq. (23) weighs differently errors in the more significant bit and in the less significant bit.

and $E\{\phi(X)|U=u\}$ are independent of u, and so, there exist infinitely many pairs (α,β) that satisfy $D(P_{Y|U=u}\|P_Y)=\alpha E\{\phi(X)|U=u\}+\beta$, including the one where $\alpha=0$, which corresponds to the case where the power constraint is not active [21], i.e., $C_S(\Gamma)=C_S(\infty)$. Condition (c) is satisfied because U_1 and U_2 are independent, and hence so are V and Y. As for condition (d), we have

$$d(u,(v,y))=d_H(u_1,v)+\theta d_H(u_2,y). \tag{27}$$

As for the right–hand side of the equation given in condition (d), first observe that $P_{U|Y}=P_{U_1}\times P_{U_2|Y}$. Now, since U_2 is symmetric and $Y=U_2\oplus W'$ is a BSC, the reverse channel $P_{U_2|Y}$ is also a BSC with crossover probability D, and so,

$$P_{U_2|Y}(u_2|y)=D^{d_H(u_2,y)}(1-D)^{1-d_H(u_2,y)}. \tag{28}$$

It then follows that condition (d) is satisfied by the choice

$$\gamma=\gamma_0\stackrel{\Delta}{=}\frac{\theta}{\log[(1-D)/D]}, \tag{29}$$

and

$$\delta(u)=\sum_v P_{V|U}(v|u)d_H(u_1,v)+\gamma_0\log[P_{U_1}(u_1)(1-D)]. \tag{30}$$

It is also easy to see that $g(v,y)=(v,y)$ achieves $\Delta(U|V,Y)$ in this case, since U_1 and U_2 are independent, they are both symmetric, and $D<1/2$.

To summarize this example, we have shown that the distortion obtained by this simple communication system,

$$E\{d(U,\hat{U})\}=(1+\theta)D, \tag{31}$$

cannot be improved by any other (more complicated) coding scheme. This example can be extended to allow for (symmetric) sources, G–P channels, W–Z channels of larger alphabets provided that the $\oplus$ operation is more generally understood as addition/subtraction modulo the alphabet size.

Example 2. Let U be uniformly distributed over the interval $[-2,2]$. The W–Z channel accepts an input $f(U)$, where f is subjected to design, and generates

$$V=\mathcal{Q}(f(U)+W)-W \tag{32}$$

where $\mathcal{Q}(\cdot)$ is a uniform quantizer of stepsize ϵ (with infinitely many levels) and W is an arbitrary RV, independent of U. The Shannon/G–P channel is given by

$$Y = S \cdot X + W' \tag{33}$$

where X is a continuous–valued RV, whose support is limited to $[-1,1]$ (i.e., $\phi(x)=\infty$ for $|x|>1$), $S\in\{-1,+1\}$, and W' is uniform over $[-1,1]$, independently of all other RV's. The distortion measure is

$$d(u,\hat{u}) = \begin{cases} 0 & |u-\hat{u}|\leq \epsilon/2 \\ \infty & |u-\hat{u}| > \epsilon/2 \end{cases} \tag{34}$$

Let $U_1 \stackrel{\Delta}{=} f(U) = U - \mathrm{sgn}(U) \stackrel{\Delta}{=} U - U_2$ and consider the encoder $X = S\cdot U_2$ (which satisfies the channel input constraint) and decoder $\hat{U} = V + \mathrm{sgn}(Y)$. Since $\mathrm{sgn}(Y) = U_2$ and $|U_1 - V|\leq \epsilon/2$ with probability one, the distortion is zero, and so, this communication system is trivially optimal. We would like to demonstrate, nevertheless, that Theorem 2 indeed tells that as well.

As for condition (a), we have

$$I(U_2;Y) = I(U_2;U_2+W') = h(U_2+W') - h(W') = \log 4 - \log 2 = 1, \tag{35}$$

where h denotes the differential entropy. Obviously, no input can achieve larger mutual information since Y is binary. This also implies that condition (b) is satisfied. Condition (c) is satisfied since U_1 and U_2 are independent. To see why this is true, note that U_1 is uniform over $[-1,1]$ regardless of whether $U_2 = 1$ or $U_2 = -1$. As for condition (d), we note that the distortion measure (34) can be formally decomposed as

$$d(u,g(v,y)) = d(u_1+u_2, v+\mathrm{sgn}(y)) = d(u_1,v) + d'(u_2,\mathrm{sgn}(y)) \tag{36}$$

where $d(u_1,v)$ is as in (34) and $d'(a,b)$ is 0 if its binary arguments are equal and infinite otherwise. The rationale behind this decomposition is that the distortion between the source and the reproduction is zero if and only if both $|u_1-v|\leq\epsilon/2$ and $u_2=\mathrm{sgn}(y)$, and if at least one of the conditions is violated, the distortion is infinite. Now, the conditional expectation of the first term given $U=u$, on the r.h.s. of eq. (36), is absorbed in $\delta(u)$ of condition (d) and the second term is proportional to $-\log P(u_2|y)$ because $P(u_2|y) = 1\{u_2=\mathrm{sgn}(y)\}$. The overall distortion, in this example, is always zero as mentioned earlier, thus g trivially achieves $\Delta(U|V,Y)$.

5 Applications

In this section, we outline several applications of our theory where a G–P, or a Shannon, channel emerges naturally in combination with a W–Z channel. In subsection 5.1, we examine a BSC with the error state available to the transmitter, that :w is subjected to a Hamming weight constraint Γ. We proceed to discuss defected memory channels, and close this section with two models of an overlaid, back–compatible communication system. The general setting motivated by the underlying framework here facilitates quantative assessment of the degradation in combining raw data (that is, systematic transmission) in the communication system and in that respect, this section extends the treatment of [31] and enhances the insight into this communications setting of practical importance.

5.1 The BSC with CSI at the Transmitter

Consider a binary symmetric source (BSS) $\{U_i\}$, operating at the rate of ρ_s bits per second, which is to be transmitted, within a prescribed Hamming distortion D, through a BSC with crossover probability $\epsilon = 1/2$, which operates at the rate of ρ_c channel uses per second. More explicitly, the channel output is given by

$$Y_i = X_i \oplus S_i, \tag{37}$$

where $\oplus$, as before, designates the XOR operation, $\{S_i\}$ – the noise sequence, is drawn by a BSS, independent of $\{U_i\}$, and $\{X_i\}$ is the binary channel input. We consider here the case where $\{S_i\}$ is available to the transmitter either in a causal or a non–causal manner, corresponding to the Shannon and the G–P settings, respectively. The channel input $\{X_i\}$ is subjected to a weight constraint:

$$E\left\{\frac{1}{n}\sum_{i=1}^{n} X_i\right\} \leq \Gamma. \tag{38}$$

We enquire, what is the maximal source rate, ρ_s, that can be conveyed subject to these constraints.

Theorem 1 tells us that for the G–P channel:

$$\rho_s \leq \frac{\rho_c \cdot H_2(\Gamma)}{1 - H_2(D)} \tag{39}$$

and, for the Shannon channel:

$$\rho_s \leq \frac{\rho_c \cdot 2\Gamma}{1 - H_2(D)}, \tag{40}$$

where we have used the facts that $C_{GP}(\Gamma) = H_2(\Gamma)$, $C_S(\Gamma) = 2\Gamma$ [4],[45] and $R(D) = 1 - H_2(D)$ [13] in this case.

Next, consider systematic coding. Since the systematic part has an average weight 1/2, it cannot endure for more than a fraction of 2Γ of the channel uses as to conform with the weight constraint (38). Now, observe that in this case, still the weight constraint is satisfied if $X_i = U_i \oplus S_i$, which is also of average weight 1/2. This yields a clean output $Y_i = U_i$ for those systematic bits that can be accommodated. Note that the transmitted bits are perfectly recovered whereas those that could not be transmitted are fully distorted ($D = 1/2$), so for $D \leq 1/2$, a fraction of $1-2D$ of the systematic bits should be transmitted. Thus, the systematic option yields:

$$\rho_s \leq \frac{\rho_c \cdot 2\Gamma}{1-2D}. \tag{41}$$

Comparing to (40), we see that the systematic approach is optimal in the lossless case ($D = 0$) of the Shannon causal setting. Note that this optimality is also achieved in the sense of Section 4, that is, symbol–by–symbol encoding.

The actual coding over these G–P and Shannon channels, in terms of random jointly typical codes or constructive algebraic binning, are explained in [45] and references therein.

5.2 Defected Memory

Consider the defective memory channel model described in [24]. According to this model, the channel output is, with probability $p_s/2$ – stuck at '0', with probability $p_s/2$ – stuck at '1', and with probability $(1-p_s)$ – behaves like a BSC with crossover probability ϵ. The capacity of this channel, with the stuck cells known non–causally at the transmitter (the G–P model), has been determined to be

$$C_{GP} = (1-p_s)[1 - H_2(\epsilon)] \quad \text{bits/memory-cell}. \tag{42}$$

If we wish to record a BSS of N bits ($N \to \infty$) within Hamming distortion D, Theorem 1 guarantees that this is possible provided that

$$1 - H_2(D) \leq \frac{n}{N}(1-p_s)[1 - H_2(\epsilon)]. \tag{43}$$

We now envisage a practical scenario where the data is recorded unaltered (that is, systematically). This is done to facilitate either simple reading or, alternatively, fast memory access, tolerating no delay, which otherwise is inherently associated with decoding.

Evidently, the distortion associated with this simple procedure is

$$D_u = \epsilon(1 - p_s) + \frac{p_s}{2}, \tag{44}$$

where the subscript u stands for "uncoded." Yet, the remaining $n-N$ memory cells are used to record optimally coded information in an effort to reduce the distortion (as compared with D_u) and that is at a cost of a more complicated, and hence slower, reader which performs the decoding.

To determine the optimal distortion in this case, designated by D_c, where subscript c stands for "coded," we interpret the uncoded recorded data as a W–Z channel. This W–Z channel is, in fact, a BSC with crossover probability D_u. The associated W–Z rate–distortion function $R_{WZ}(D; D_u)$ (where the second argument, with a slight abuse of notation, designates the crossover probability of this W–Z channel), has been established in [42] and it equals

$$R_{WZ}(D; D_u) = \begin{cases} g(D), & 0 \leq D \leq D_t \\ g(D_c)\left(1 - \frac{D-D_t}{D_u-D_t}\right), & D_t < D \leq D_u \end{cases} \tag{45}$$

where

$$g(D) = \begin{cases} H_2(D_u * D) - H_2(D), & 0 \leq D \leq D_u \\ 0, & D = D_u \end{cases} \tag{46}$$

where $*$ designates the binary convolution operation: $\alpha * \beta = \alpha(1-\beta) + \beta(1-\alpha)$, $0 \leq \alpha, \beta \leq 1$, and where the threshold distortion D_t is the solution to the equation

$$\frac{g(D_t)}{D_t - D_u} = g'(D_t). \tag{47}$$

Evidently, by Theorem 1, which involves here the W–Z and G–P components, the reduced distortion associated with the addition of the coded part, D_c is given by the solution to the equation

$$R_{WZ}(D_c; D_u) = \left(\frac{n}{N} - 1\right)(1 - p_s)[1 - H_2(\epsilon)]. \tag{48}$$

This expression describes the best possible trade-off between the excess memory $(n - N)$ versus distortion reduction (from D_u to D_c). Note that due to the general suboptimality of the W–Z coding as compared to the case where the W–Z side information is available to the encoder in this binary regime [31],[42], there is a cost associated with imposing the recording of raw data. This cost, in general, manifests itself in $D_c \geq D$, cf. (48) and (43), respectively.

5.3 Overlaid Back–Compatible Communication

Within this framework of overlaid back–compatible communication, we shall consider the binary case and the Gaussian case. Both cases invoke the model where common information is transmitted to users who employ standard (old) equipment. New users, who may use more advanced equipment, are interested in better performance. However, service to old users must be maintained, with a prescribed level of allowed degradation. We investigate the trade–off between improvement in performance enjoyed by the new users and the degradation inflicted on the old ones. In the binary case, we shall focus on providing unequal distortion levels, that is, the improvement in performance refers to part of the information which is considered to be more important. We envisage the raw data being composed of two parts. The first part, of fraction q, is the important part, the performance of which we are to enhance, at the expense of large distortion in the remaining part (a fraction of $(1-q)$) of the data.

These models can be cast also within the general framework of lossy transmission over broadcast channels, see [7],[27],[29] and references therein, and the technique here might be useful in devising achievable regions for other communications configurations. We emphasize that the theory developed in the previous sections is merely used *to motivate* the communication schemes discussed below, but by no means, these schemes will be claimed to be optimal.

For simplicity, we shall use, in this subsection and in the figures involved, scalar notation, conforming with the single–letter expressions of the associated information–theoretic expressions.

5.3.1 The Binary Case

Consider a BSC with crossover probability p over which the common data is transmitted. Evidently, this BSC may model a full–fledged communication system, where p characterizes its error probability. Thus, p is the Hamming distortion, or the bit–error rate (BER) of the raw information transmitted over this channel. The new user, who employs advanced detection methods, is interested in improving on the distortion of the more important part of the data, that is a fraction of q of the total data, while in doing so, penalizing to an allowable degree the rest of the data, a fraction of $(1-q)$. In this case, the upgrade of the important part of the data is due to the more involved coded communication technique used, taking advantage of the superimposed coded part on the less important fraction of the data.

We would like then to provide a quantitative assessment of the trade–off between the BER reduction of the important part of the data, and the performance degradation of the less important part. We emphasize that back compatability with respect to old receivers is preserved in terms of the important part being absolutely unaffected, that is subjected to the channel distortion (bit error probability) p, while the less important part suffers the very same degradation, for new and old receivers, namley,

$$D_b = p * e \tag{49}$$

where $0 \leq e \leq 1/2$ is thought of as a parameter. Our encoding system, described in Fig. 3, overlays communication only on the less important part of the data (that is, a fraction of $(1-q)$ of the channel uses), while the important data (fraction of q) is transmitted unaltered. We will refer now to the altered part, which appears in Fig. 3, and produce then the channel input

$$X' = U_b \oplus X, \tag{50}$$

where U_b stands for the symmetric $\{0,1\}$ less important fraction of the data and X the encoded part. The received signals associated with the less important (U_b) and the more important (U_g) parts of the data, designated by Y_g and Y_b, respectively, are given by

$$\begin{aligned} Y_b &= X' \oplus W_b \\ Y_g &= U_g \oplus W_g, \end{aligned} \tag{51}$$

where W_b and W_g stand for the independent binary noise components with $P_{W_b}(1) = \Pr\{W_b = 1\} = P_{W_g}(1) = p$. The reconstructed information of the two parts (U_g, U_b) of the data associated with the old receiver is designated by $(\hat{U}_g, \hat{U}_b)$ while the corresponding reconstruction associated with the new receiver is denoted by $(\hat{\hat{U}}_g, \hat{\hat{U}}_b)$ (see Fig. 3). Clearly, the additional transmitted part X should be constrained such that the combined binary noise $X \oplus W_b = Y_b \oplus U_b$, which impact the less important part of the data, does not inflict more degradation than allowed. In other words, this imposes the constraint $P_X(1) \leq e$ (i.e., $\Gamma = e$).

The encoded message is constructed by concatenating W–Z and G–P encoders, where the W–Z encoder accounts for the side information channel between U_g and Y_g, where

$$E\{U_g \oplus Y_g\} = p. \tag{52}$$

The associated rate (per information bit) is given by $R_{WZ}(D_g; p)$ (cf. eq. (45)). Now, the channel from X to Y_b is viewed as a G–P channel where U_b plays the role of a state sequence given non–causally to the transmitter,[8] whose capacity (per channel use) [4] is given by

$$C_{GP}(e) = \text{u.c.e}\left\{H_2(e) - H_2(p), (0,0)\right\}, \tag{53}$$

where u.c.e$\{\cdot\}$ denotes the upper concave envelope as a function of e, which includes the origin $(0, 0)$. Theorem 1 characterizes then the trade-off between D_g, the distortion associated with the important part of the data and the new user and D_b, the distortion associated with the less important part in a parametric set of equations. Namely, the solution for D_g of the equation

$$qR_{WZ}(D_g; p) = (1 - q)C_{GP}(e), \tag{54}$$

together with (49), where $0 \leq e \leq 1/2$ is the parameter. The q and $1 - q$ factors in the equation above account for the fact that the G–P channel is used in a fraction of $(1 - q)$ of the channel uses (associated with the less important part of the data), while the important part of the data is transmitted unaltered over the fraction q of the channel uses. Note that G–P encoding [4] guarantees that X is independent of U_b, yielding that $X \oplus W_b$ in (50),(51), are noises that are independent of U_b. Practical concepts of G–P and W–Z coding within the context of either random or nested linear codes are reviewed in [45]. Evidently, $qD_g + (1-q)D_b \geq p$ since uncoded transmission, which achieves average Hamming distortion p, is optimal in this setting [21].

Note this approach can also be viewed as non-equal error protection implemented within back compatability constraints. The issue of non-equal error protection is also closely related to the framework of broadcast channels [39], which highlights again the general suboptimality of source/channel separation in the realm of a broadcast setting, see also [7],[27],[29] and references therein.

[8] This is implemented by delaying the raw data before transmission.

5.3.2 The Gaussian Case

Here, we address the setting introduced first in [31], and is depicted in Fig. 4. Here U is the binary memoryless source transmitted uncodedly and antipodally, with values $\pm\sqrt{P_I}$. Evidently, the Hamming distortion associated with this transmission is

$$p_b = Q\left(\sqrt{\frac{P_I}{\sigma^2}}\right) = \frac{1}{\sqrt{2\pi}}\int_{\sqrt{P_I/\sigma^2}}^{\infty} e^{-x^2/2}dx, \tag{55}$$

where $Q(\cdot)$ designates the complementary error function and where we have assumed an additive Gaussian noise with variance $EW^2 = \sigma^2$. Again, in order to enhance performance for new users, a certain fraction of the power P_D $(0 \leq P_D \leq P_I)$ is taken away and assigned to the coded transmission X as shown in Fig. 4. The achievable distortion D_u for the uncoded user is[9] then

$$D_u = E\{U \oplus \hat{U}_u\} = Q\left(\sqrt{\frac{P_I - P_D}{\sigma^2 + P_D}}\right), \tag{56}$$

where $\hat{U}_u$ stands for the uncoded demodulated information based on the channel output Y. The strategy introduced here can be viewed as complementary to Strategy 1 described in [31], and is uniformly advantageous over Strategy 2 therein, as is shown next.

The model in Fig. 4 can readily be generalized to the case where the two decoders, producing $\hat{U}_u$ and $\hat{U}_c$, have different ambient Gaussian noise conditions, in parallel to the binary setting as described in Fig. 3. For the sake of simplicity and comparison w ith the strategies in [31], we adhere to the special case as described in Fig. 4.

The encoding procedure is motivated by Theorem 1, combining W–Z source coding and G–P channel coding. First, observe that the W–Z channel, as seen from U to Y, is equivalent to a binary input Gaussian channel with input power $P_I - P_D$ and additive Gaussian[10] noise of power $P_D + \sigma^2$. The W–Z rate–distortion function of a BSS with Hamming distortion D and a Gaussian W–Z channel having signal–to–noise ratio SNR, designated by $R_{WZ}(D; SNR)$, is given [31] by

$$R_{WZ}(D; SNR) = \text{l.c.e}\{F(D; SNR)\}, \tag{57}$$

where l.c.e$\{\cdot\}$ stands for a lower convex envelope as a function of D, and where $F(D; SNR)$, is given parametrically by two functions, $D(\alpha)$ and $F(\alpha)$, of in independent parameter

[9] It will be argued that the coded part X is Gaussian and independent of U.

$0 \leq \alpha \leq 1/2$, which for a given SNR are defined as follows:

$$\begin{aligned} D(\alpha) &= \alpha Q\left(\sqrt{SNR} - \frac{1}{2\sqrt{SNR}} \log \frac{1-\alpha}{\alpha}\right) \\ &\quad +(1-\alpha) Q\left(\sqrt{SNR} + \frac{1}{2\sqrt{SNR}} \log \frac{1-\alpha}{\alpha}\right), \end{aligned} \tag{58}$$

$$F(\alpha) = \int_{-\infty}^{\infty} \frac{du}{\sqrt{2\pi}} e^{-u^2/2} H_2\left(\frac{(1-\alpha) + \alpha e^{-2(u\sqrt{SNR}+SNR)}}{1 + e^{-2(u\sqrt{SNR}+SNR)}}\right) - H_2(\alpha). \tag{59}$$

Now, for the G–P part, the uncoded antipodal signaling (of power $P_I - P_D$) serves as a state known in advance to the transmitter[9]. Invoking the generalization[10] [11],[17] of the classical setting of Costa [12], the capacity of this channel yields

$$C_{GP}(P_D) = \frac{1}{2} \log\left(1 + \frac{P_D}{\sigma^2}\right), \tag{60}$$

eliminating absolutely the state inflicted interference. The trade–off between the uncoded distortion $D_u = E\{U \oplus \hat{U}_u\}$ and the distortion for the coded part $D_c = E\{U \oplus \hat{U}_c\}$, is then given by the pair of equations, namely, (56) and the solution D_c to the equation

$$R_{WZ}\left(D_c; \frac{P_I - P_D}{\sigma^2 + P_D}\right) = C_{GP}(P_D). \tag{61}$$

Note that also in the generalized Costa setting[10] [11], X is independent of U and is Gaussian with variance P_D. Hence, the overall noise $X + W$ w.r.t. the uncoded part is also viewed as Gaussian, whose power is $P_D + \sigma^2$, thus giving rise to equation (56), as well as to the Gaussian W–Z side information channel.

Observe that the overlaid communication strategy proposed here favors the coded part in the sense that it suffers no interference from the uncoded part via the G–P setting. In this respect, our strategy is uniformly better than Strategy 2 of [31] opting for the same goal, namely, reducing the interference to the coded part. This can be verified by noting that the performance here equals to that of [31] eq. (54) with q therein (commensurates with the residual interferences of the uncoded part to the coded one) set to zero. The complement Strategy 1 of [31], which favors the W–Z channel, while absorbing the full interference of the uncoded part to the coded signal, exhibits the trade-off characterized by eq. (56) and equation

$$R_{WZ}\left(D_c; \frac{P_I - P_D}{\sigma^2}\right) = \frac{1}{2} \log\left(1 + \frac{P_D}{\sigma^2 + P_I - P_D}\right), \tag{62}$$

[10]The state sequence here is binary and memoryless, which differs from the Gaussian memoryless state in Costa's setting [12].

solved for D_c.

In this model of the overlaid communication system, the very same channel output is used to produce the W–Z and the G–P channels, which are evidently correlated being affected by the very same noise component (available neither to the transmitter nor to the receiver). No use of this feature was attempted and hence no optimality is claimed. Note, however, that the coded overlaid part is independent of the uncoded part, where the latter is interpreted as the known state sequence in the G–P channel model.

Specific coding strategies either random or structured (based on nested lattice for the W–Z and G–P problems as specialized here) are overviewed in [45]. Further note that no look–ahead techniques are needed in the preferred strategy of Fig. 4 as compared to Strategy 2 of [31], but a simple delay[9].

6 Conclusions

We have addressed a communication framework which combines two basic ingredients of side information, namely, side information about the source provided to the decoder, which give rise to the W–Z rate distortion [42] and side information about a communication channel provided to the transmitter only which matches the Gel'fand–Pinsker [22] or Shannon [32] models. A separation theorem is shown to exist also in this general communication framework where the standard Markov structure, necessary for the application of the classical data processing theorem [13], is not maintained. This implies that no advantage can be gained by incorporating the channel state information within the source encoder.

Within this framework, the conditions that guarantee the optimality of scalar uncoded communication are determined extending earlier classical results to this setting.

Various application of communications systems where the W–Z and G–P models emerge in a natural way are discussed. In particular, this setting facilitates quantitive assessment of the degradation associated with systematic coding, where the W–Z channel is structured by those channel outputs that correspond to the systematic transmitted part (raw data). This treatment extends previous analyses [30], [31], which have considered standard channels with no side information. Motivated by the theory here, we have advocated a combined W–Z and G–P approach to overlaid, back compatible, communication systems. In the binary setting (Fig. 3) the enhancement is associated with the important part of the data, and it is achieved at the expense of larger distortion in the less important data. Still, back compatibility is guaranteed and no degradation is inflicted on the important part of the data. In the setting of Fig. 4, both the W–Z and G–P channel are correlated in terms of the ambient noise, and hence no optimality claims of our overlaid communication structures are made. Yet, as demonstrated, the current approach is uniformly better than previous treatments of similar models [31, Section V, Strategy 2]. This combination of the W–Z and G–P models can also be used as to gain further insight to the general problem of lossy transmission over a broadcast channel [27].

The overall setting of our model, as depicted in Fig. 2 is general enough to allow distorted versions of side information about the source and channel to be available to both transmitter and receiver, though we do demand that the G–P and W–Z channels are statistically independent given the source input. As mentioned, extension to general alphabets and sets follows directly by classical extensions of the W–Z and G–P stand alone results, and facilitate the application of the results here to a wide classes of sources and channels with side information.

Acknowledgement

The authors are grateful to the anonymous reviewers and the Associate Editor for their useful comments. They also thank Yossi Steinberg for an interesting observation that lead to a more general (and hence weaker) form of condition (d) in Theorem 2, relative to the one of the original version of this paper.

References

[1] R. Ahlswede, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. Inform. Theory*, vol. IT-32, no. 5, pp. 621–629, September 1986.

[2] E. A. Aroutunian and M. E. Aroutunian, "E-capacity upper bound for a channel with random parameter," *Probl. Contr. Inform. Theory*, vol. 17, no. 2, pp. 99–105, 1988.

[3] R. J. Barron, B. Chen, and G. W. Wornell, "The duality between information embedding and source coding with side information and its implications/applications," preprint, January 2000, revised, June 2001 (available at: `http://allegro.mit.edu/dspg/publications/Journals/pdf/00Barron.pdf`).

[4] R. J. Barron, B. Chen, and G. W. Wornell, "On the duality between information embedding and source coding with side information and some applications," in *Proc. Int. Symp. Information Theory (ISIT2001)*, Washington, DC, June 2001, p.300. Full version in *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1159–1180, May 2003.

[5] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading Channels: Information–Theoretic and Communications Aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, October 1998 (special commemorative issue 1948–1998).

[6] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2007–2019, September 1999.

[7] B. Chen, S.C. Draper and G.W. Wornell, "Information Embedding and Related Problems: Recent Results and Applications," *39th Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, Illinois, October 3-5, 2001.

[8] M. Chiang and T. M. Cover, "Unified duality between channel capacity and rate distortion with state information," *Proc. ISIT 2001*, p. 301, Washington, D.C., June 2001.

[9] A. S. Cohen, *Information Theoretic Analysis of Watermarking Systems*, Ph.D. ddissertation, MIT, Cambridge, MA, September 2001 (available at: `http://www.mit.edu:8001/people/acohen/Pubs/index.html`).

[10] A. S. Cohen, "Communication with side information," graduate seminar 6.962, MIT, Cambridge, MA, Spring 2001 (available at: `http://web.mit.edu/6.962/www/www_spring_2001/schedule.html`).

[11] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639–1667, June 2002.

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

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

[14] I. Csisz'ar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, 1981.

[15] A. Das and P. Narayan, "Capacities of time–varying multiple–access channels with side information," *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 4–25, January 2002.

[16] U. Erez, "Noise prediction for channel coding with side information at the transmitter," M.Sc. dissertation, Tel Aviv University, December 1998.

[17] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," *Proc. ISITA 2000*, Honolulu, Hawaii, USA, November 2000.

[18] U. Erez and R. Zamir, "Noise prediction for channel coding with side information at the transmitter," *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 1610–1617, July 2000.

[19] U. Erez and R. Zamir, "Error exponents of modulo–additive noise channels with side information at the transmitter," *IEEE Trans. Inform. Theory*, vol. 47, no. 1, pp. 210–218, January 2001.

[20] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, 1968.

[21] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source–channel communication revisited," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003. (See also, R. Rimoldi, "Beyond the separation principle: a broader approach to source–channel coding," *Fourth Int'l. ITG Conf. on Source & Channel Coding*, pp. 233–238, January 28–30, 2002, Berlin, Germany.)

[22] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Information and Control*, pp. 19-31, 1980.

[23] C. Heegard, "On the capacity of permanent memory," *IEEE Trans. Inform. Theory*, vol. IT–31, no. 1, pp. 34–42, January 1985.

[24] C. Heegard and A. A. El–Gammal, "On the capacity of commuter memory with defects," *IEEE Trans. Inform. Theory*, vol. IT–29, no. 5, pp. 731–739, September 1983.

[25] A. H. Kaspi, "Rate–distortion function when side information may be present at the decoder," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 2031–2034, November 1994.

[26] A. V. Kuznetsov and A. J. Han Vinck, "On the general defective channel with informed encoder and capacities of some constrained memories," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1866–1871, November 1994.

[27] U. Mittal and N. Phamdo, "Joint source–channel codes for broadcasting and robust communication," to appear in *IEEE Trans. Inform. Theory*.

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

[29] Z. Reznic, R. Zamir and M. Feder, "Joint source–channel coding of a Gaussian mixture source over the Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, Vol. 48, No. 3, pp. 776–781, March 2002.

[30] S. Shamai (Shitz), S. Verdú and R. Zamir, "Information Theoretic Aspects of Systematic Coding," *Proc. Int. Symp. on Turbo Codes & Related Topics*, pp. 40–46, Brest France, September 3–5, 1997.

[31] S. Shamai (Shitz), S. Verdú and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 564–579, March 1998.

[32] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, October 1958.

[33] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT–19, no. 4, pp. 471–480, July 1973.

[34] J. K. Su, J. J. Eggers, and B. Girod, "Illustration of the duality between channel coding and rate distortion with side information," *Proc. 34th Asimolar Conference on Signals, Systems and Computers*, October–November, 2000, Asimolar, CA, U.S.A.

[35] J. K. Su, J. J. Eggers, and B. Girod, "Channel coding and rate distortion with side information: geometric interpretation and illustration of duality," preprint 2000 (revised 2001), submitted to *IEEE Trans. Inform. Theory.*

[36] A. Sutivong, T. M. Cover, and M. Chiang, "Tradeoff between message and state information rates," *Proc. ISIT 2001*, p. 303, Washington, D.C., June 2001.

[37] A. Sutivong, T. M. Cover, M. Chiang, and Y.-H. Kim, "Rate vs. distortion trade-off for channels with state information," *Proc. ISIT 2002*, p. 226, Lausanne, Switzerland, June–July, 2002 (full version to appear in *IEEE Trans. Inform. Theory*).

[38] B. van Thanh, "Storage capacity of computer memories with defects," *Probl. Contr. Inform. Theory*, vol. 19, no. 5–6, pp. 423–434, 1990.

[39] M. D. Trott, "Unequal error protection codes: theory and practice," *Proc. 1996 IEEE Information Theory Workshop*, p. 11, Haifa, Israel, June 1996.

[40] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inform. Theory*, vol. 45, no. 2 pp. 761–771, March 1999.

[41] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT–21, no. 3, pp. 294–300, May 1975.

[42] A. D. Wyner and J. Ziv, "The rate–distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT–22, no. 1, pp. 1–10, January 1976.

[43] W. Yu, A. Sutivong, D. Julian, T. M. Cover, and M. Chiang, "Writing on colored paper," *Proc. ISIT 2001*, p. 302, Washington, D.C., June 2001.

[44] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 2073–2084, November 1996.

[45] R. Zamir, S. Shamai (Shitz) and U. Erez, "Nested linear/lattice codes for structured multi–terminal binning," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

Neri Merhav (S'86–M'87–SM'93–F'99) was born in Haifa, Israel, on March 16, 1957. He received the B.Sc., M.Sc., and D.Sc. degrees from the Technion, Israel Institute of Technology, in 1982, 1985, and 1988, respectively, all in electrical engineering.

From 1982 to 1985 he was a research associate with the Israel IBM Scientific Center in Haifa, where he developed algorithms for speech coding, speech synthesis and adaptive filtering of speech signals in array sensors. From 1988 to 1990 he was with AT&T Bell Laboratories, Murray Hill, NJ, USA, where he investigated and developed algorithms for speech recognition. Since 1990 he has been with the Electrical Engineering Department of the Technion - Israel Institute of Technology, where he is a Professor. During 1994–1999 he was also serving as a consultant to the Hewlett-Packard Laboratories – Israel (HPL-I). During 1995–1996 he was on sabbatical leave at Hewlett-Packard Laboratories, where he was invloved in research topics associated with image and video compression. His current research interests are statistical signal processing, information theory, and statistical communications.

Dr. Merhav was a co-recepient of the 1993 Paper Award of the IEEE Information Theory Society and he is a Fellow of the IEEE since 1999. He also received the 2002 Technion Henry Taub Prize for Excellence in Research. From 1996 until 1999 he served as an Associate Editor for Source Coding of the IEEE TRANSACTIONS ON INFORMATION THEORY. He also served as a co-chairman of the Program Committee of the 2001 IEEE International Symposium on Information Theory.

Shlomo Shamai (Shitz) (S'80–M'82–SM'89–F'94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975-1985 he was with the Communications Research Labs in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now the William Fondiller Professor of Telecommunications. His research interests include topics in information theory and statistical communications. He is especially interested in theoretical limits in communication with practical constraints, multi-user information theory and spread spectrum systems, multiple-input-multiple-output communications systems, information theoretic models for wireless networks and systems, information theoretic aspects of magnetic recording, channel coding, combined modulation and coding, turbo codes and LDPC, in channel, source, and combined source-channel applications, iterative detection and decoding algorithms, coherent and noncoherent detection and information theoretic aspects of digital communication in optical channels.

Dr. Shamai (Shitz) is a member of the Union Radio Scientifique Internationale (URSI). He is the recipient of the 1999 van der Pol Gold Medal of URSI, and a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, and the 2002 joint IT/COM societies paper award. He is also the recipient of the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE Transactions on Information Theory, and also serves on the Board of Governors of the Information Theory Society.

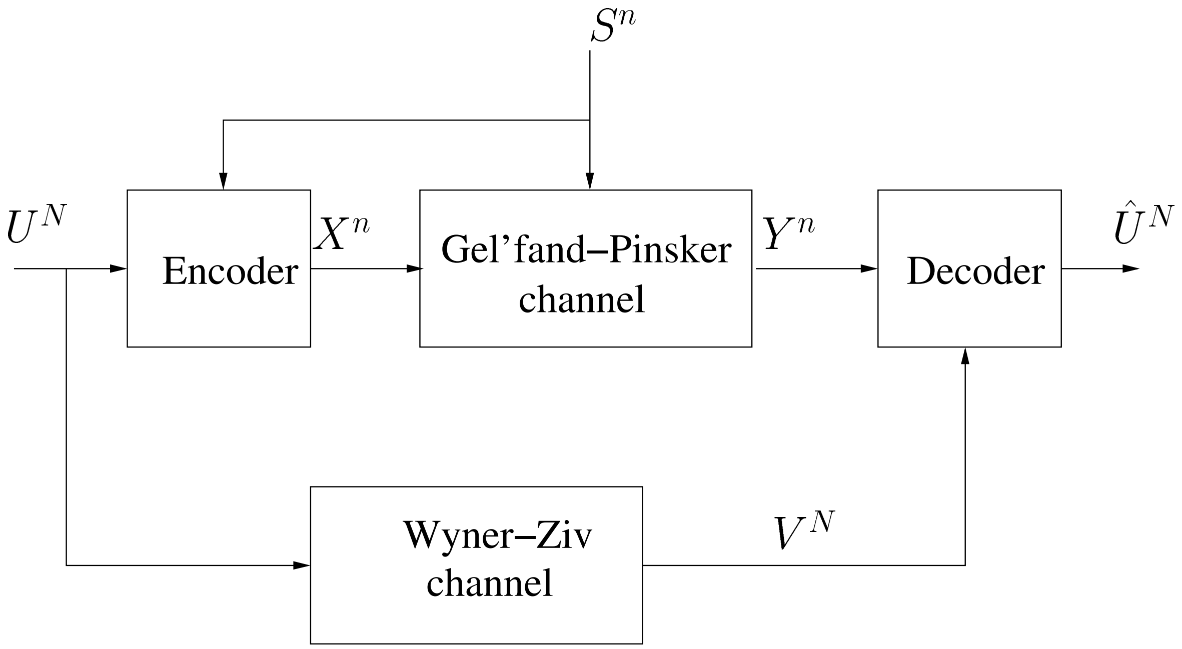

Figure 1: Source–channel coding for the combined model of Wyner–Ziv and Gel'fand–Pinsker.

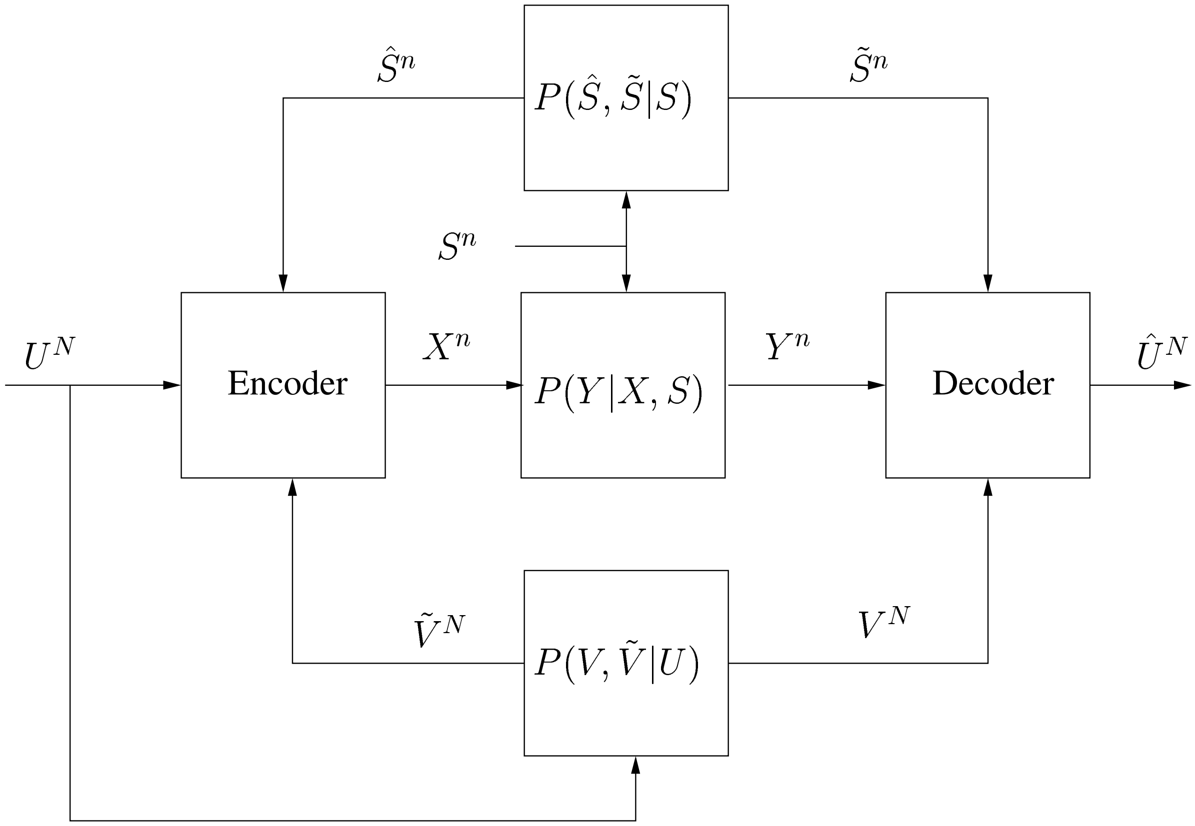

Figure 2: Symmetric side information model.

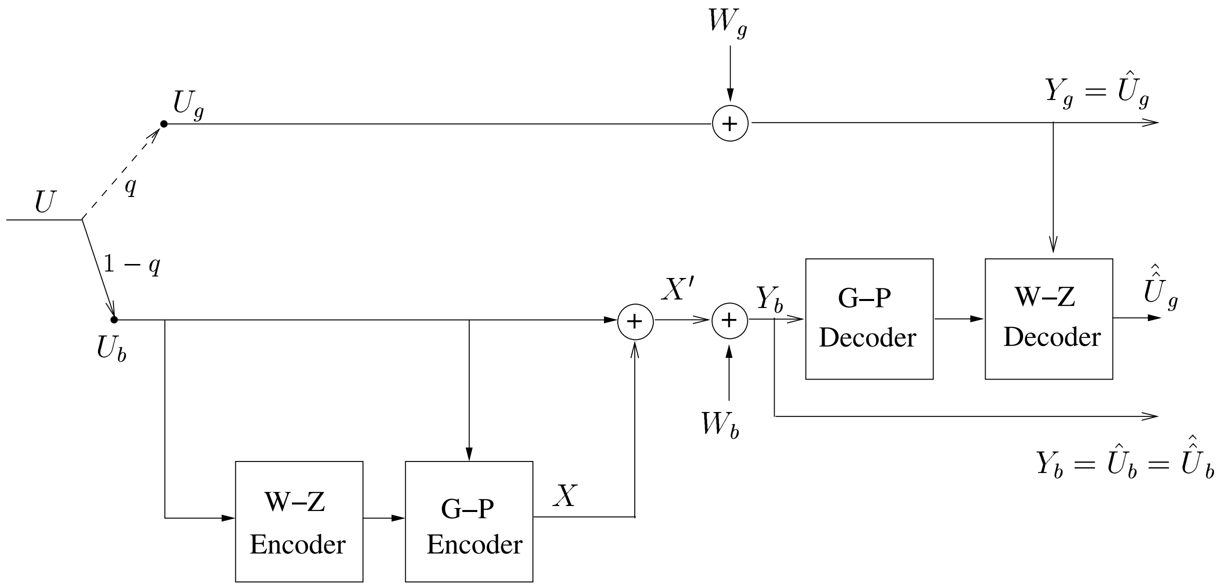

Figure 3: A binary hierarchical overlaid communication system.

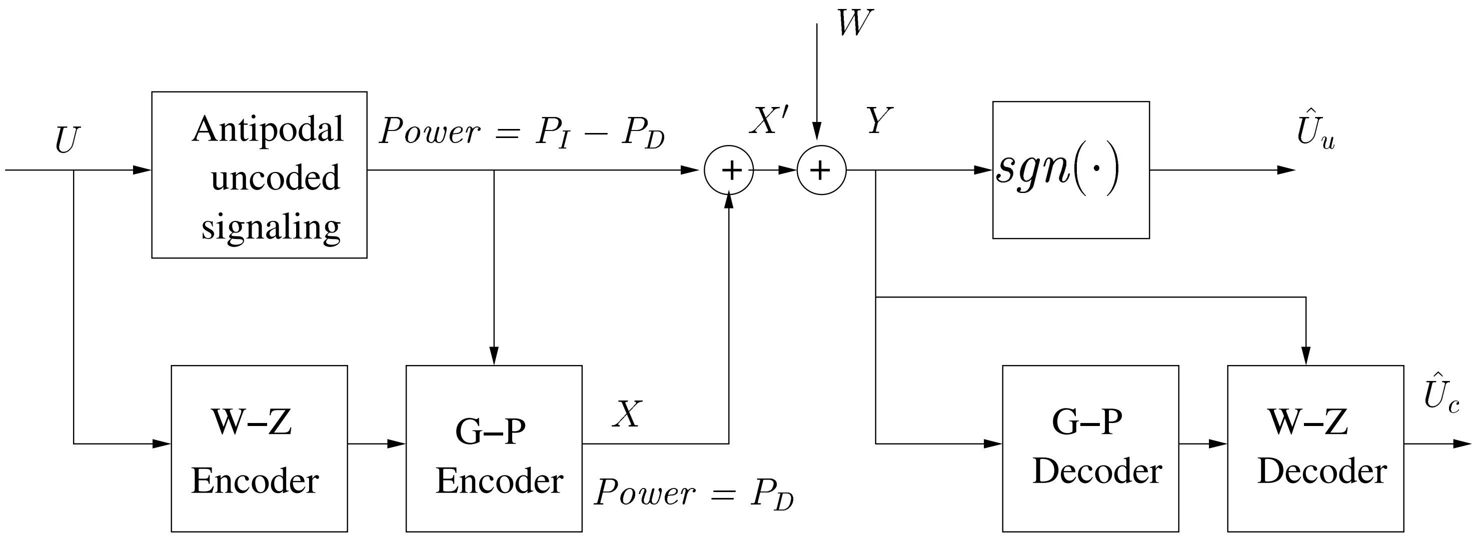

Figure 4: A Gaussian overlaid communication system.