Forty-Fourth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
Sept 27-29, 2006

Joint Source-Channel Coding for the Arbitrarily Varying Wyner-Ziv Source and Gel'fand-Pinsker Channel

Amir Winshtok and Yossef Steinberg

***Abstract*— In this work we characterize the rate-distortion function of a Wyner-Ziv source, depending on an arbitrarily varying state, known non-causally at the encoder. We then consider the problem of joint source-channel coding of this source transmitted over an arbitrarily varying Gel'fand-Pinsker channel, where we assume that one jammer controls both, the source state sequence and the channel state sequence. Utilizing our result, and a previous result by Ahlswede on the capacity of the arbitrarily varying Gel'fand-Pinsker channel, we show that a separation principle for this setup holds. A by product of this result is that the best strategy for the jammer, is to choose the source state and the channel state sequences in such a manner, that the Wyner-Ziv source and the Gel'fand-Pinsker channel look like independent of each other. This implies that a separation holds also with respect to the operation of the jammer, as the jammer employing the best strategy (from his point of view) can be split into two non cooperating jammers, one of which controls only the source state sequence, and the other controls only the channel state sequence.**

I. INTRODUCTION

Source and channel models that depend on state variables have been studied intensively with various assumptions on the way the state parameter is selected and on the knowledge of the state parameter in the encoder/decoder. In this work we consider a source where no statistical model of the way the state variables are selected is available, and under the assumption that they are known in a non-causal manner at the encoder. We consider a rate-distortion problem of a source that produces a pair (U, V) of random variables with probability depending on the state variable. The variable U is considered as the source we would like to compress and reconstruct at the decoder with a given distortion level, and V serves as side information at the decoder. We refer to this source as an arbitrarily varying Wyner-Ziv source. The rate-distortion result we prove states that the minimum rate needed to achieve a distortion level D is the maximum of the Wyner-Ziv rate distortion function, over all possible memoryless state distributions.

We then consider the problem of joint source-channel coding of this source transmitted over an arbitrarily varying Gel'fand-Pinsker channel. Utilizing our result, and a previous result by Ahlswede [4] on the capacity of the arbitrarily varying Gel'fand-Pinsker channel, we show that a separation principle for this setup holds, that is, a distortion level D is achievable if and only if the source rate-distortion function evaluated at D is not larger than the capacity of the arbitrarily varying Gel'fand-Pinsker channel. In other words, no loss in optimality is incurred if we separate the source coding form the channel coding.

This work was supported by The Israel Science Foundation under grant 61/03.

The authors are with the Faculty of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel (e-mail: wamir@tx.technion.ac.il; ysteinbe@ee.technion.ac.il).

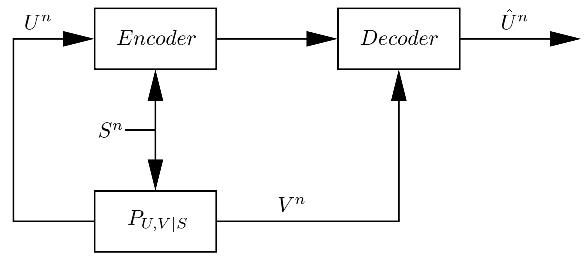

Fig. 1. An arbitrarily varying Wyner-Ziv source with side information at the encoder

II. ARBITRARILY VARYING WYNER-ZIV SOURCE

Let $\{(U_i, V_i)\}_{i=1}^{\infty}$ denote a sequence of independent random variables (RVs) U and V, taking values in the finite sets $\mathcal{U}$ and $\mathcal{V}$, respectively, with joint probability $P_{U,V|S}$. s is a state variable, taking values in the finite set $\mathcal{S}$. The state sequence S^n is drawn according to some distribution, or arbitrarily, as will be defined later on.

Consider the system described in Fig. 1. We would like to encode a sequence of source symbols of length n, U^n, into codewords of rate R bits per source symbol. We assume that the encoder has an access to the state sequence s^n. At the decoder, an approximation of the source is constructed. The quality of reconstruction (approximation) is measured with respect to (w.r.t.) the distortion measure

$$d(u^n, \hat{u}^n) = \frac{1}{n}\sum_{i=1}^{n} d(u_i, \hat{u}_i) \qquad (1)$$

where $\hat{u}^n$ is the reconstruction sequence, and the letters $\hat{u}_i$ take values in the finite set $\hat{\mathcal{U}}$, and where $d(u, \hat{u})$ is a single letter distortion measure, bounded by D_{max}. The rate-distortion function $R(D)$ is the minimal rate for which an encoder and decoder system described above exists for sufficiently large n, with distortion arbitrary close to the given distortion level D. In this section we characterize the rate-distortion function of the described source, when the state sequence is chosen in an arbitrary manner from the set of all available state sequences.

A. *Definitions and known results*

We start with definitions of three source models, distinguished by the way the state sequence is selected. For a given state sequence s^n, the probability of the sequences (u^n, v^n) is given by:

$$P_{U^n,V^n|S^n}(u^n, v^n|s^n) = \prod_{i=1}^{n} P_{U,V|S}(u_i, v_i|s_i). \quad (2)$$

For shorthand notation, we will denote this probability by $\psi(u^n, v^n|s^n)$. We define a *Wyner-Ziv source with random parameters (RP-WZ)* as a source such that s^n is chosen iid with known probability q, i.e., $P_{S^n}(s^n) = \prod_{i=1}^{n} q(s_i)$. We denoted this source by $\{q, P_{U,V|S}\}$. Note that the RP-WZ source is actually a regular WZ source, where the S component, available at the encoder, is not part of the reconstruction. Thus it can be expressed as a WZ source (U, S), with side information V, where the distortion measure is insensitive to the S component.

For any finite set M, let $\mathcal{P}(M)$ stand for the set of all probability distributions on M. Let $\mathcal{Q}$ be a subset of $\mathcal{P}(\mathcal{S})$. A *Compound WZ (C-WZ)* source is a RP-WZ source where s^n is chosen in an iid manner with unknown probability from the set $\mathcal{Q}$. We denote this source by $\{\mathcal{Q}, P_{U,V|S}\}$. A $\{q, P_{U,V|S}\}$ RP-WZ source is a special case of a $\{\mathcal{Q}, P_{U,V|S}\}$ C-WZ source with $\mathcal{Q} = \{q\}$.

Let $\mathcal{Q}^n$ be a subset of $\mathcal{P}(\mathcal{S}^n)$. Note that it may contain atoms (i.e., probability distributions that put all the mass on a single sequence), as well as memoryless state distributions. Let $\mathbb{Q}$ be a sequence in n of $\mathcal{Q}^n$, i.e., $\mathbb{Q} = \{\mathcal{Q}^n\}_{n=1}^{\infty}$. The state sequence of length n in an *Arbitrarily Varying Wyner-Ziv (AV-WZ)* source model is chosen with unknown probability from the set $\mathcal{Q}^n$. We denote this source by $\{\mathbb{Q}, P_{U,V|S}\}$. If for every n, $\mathcal{Q}^n = \mathcal{P}(\mathcal{S}^n)$, the AV-WZ source is denoted by $\{P_{U,V|S}\}$. A $\{\mathcal{Q}, P_{U,V|S}\}$ C-WZ source is a special case of an $\{\mathbb{Q}, P_{U,V|S}\}$ AV-WZ source with $\mathcal{Q}^n = \{\prod_{i=1}^{n} q, \, q \in \mathcal{Q}\}$, for every n.

We proceed to define *rate-distortion codes* (or simply codes) for the AV-WZ source. Since the RP-WZ and C-WZ sources are spacial cases of an AV-WZ source, the definitions of rate-distortion codes hold for all types of sources that we mentioned thus far. We distinguish between three classes of such codes. The first class is the class of *deterministic rate-distortion codes*. An (R, n, D) deterministic rate-distortion code consists of a deterministic encoding function

$$f : \mathcal{U}^n \times \mathcal{S}^n \to \{1, \dots, 2^{nR}\} \quad (3)$$

and a deterministic decoding function

$$\varphi : \{1, \dots, 2^{nR}\} \times \mathcal{V}^n \to \hat{\mathcal{U}}^n, \quad (4)$$

such that the maximal distortion (over all state distributions in $\mathcal{Q}^n$) is bounded by D:

$$D \geq \sup_{q \in \mathcal{Q}^n} \sum_{s^n} \sum_{u^n, v^n} q(s^n) \psi(u^n, v^n|s^n) \; d\big(u^n, \varphi(f(u^n, s^n), v^n)\big). \quad (5)$$

The rate of the code is R. We denote this code by (f, φ).

Fix a state sequence s^n. Let $\Delta_\delta(s^n)$ be the set of all pairs of sequences $(u^n, v^n) \in \mathcal{U}^n \times \mathcal{V}^n$ with distortion greater than $D + \delta$, i.e.,

$$\Delta_\delta(s^n) \triangleq \big\{(u^n, v^n) \in \mathcal{U}^n \times \mathcal{V}^n : d(u^n, \varphi(f(u^n, s^n), v^n)) > D + \delta\big\}. \quad (6)$$

We refer to $\Delta_\delta(s^n)$ as the *miss represented set, given* s^n. For notational convenience we will omit the δ and simply denote the miss represented set by $\Delta(s^n)$.

The second class is the class of *randomized-encoder rate-distortion codes*. An (R, n, D) randomized-encoder rate-distortion code consists of an encoder probability distribution over the set $\{1, \dots, 2^{nR}\}$, depending on the source sequence u^n and the source state sequence s^n, $r(\cdot|u^n, s^n)$, and a deterministic decoding function φ (4), such that the maximal distortion (over all state distributions in $\mathcal{Q}^n$) is bounded by D:

$$D \geq \sup_{q \in \mathcal{Q}^n} \sum_{s^n} \sum_{u^n, v^n} \sum_{l=1}^{2^{nR}} q(s^n) \psi(u^n, v^n|s^n) \; r(l|u^n, s^n) d\big(u^n, \varphi(l, v^n)\big). \quad (7)$$

We denote this code by (r, φ). A deterministic code is a special case of a randomized-encoder code where $r(\cdot|u^n, s^n)$ puts all the mass on $f(u^n, s^n)$.

The third class is the class of *correlated rate-distortion codes*. An (R, n, D) correlated rate-distortion code consists of a collection of deterministic (R, n, D) codes, $\{(f^\gamma, \varphi^\gamma)\}_{\gamma \in \Gamma}$, together with a probability distribution μ on the codes collection Γ. Before encoding begins, one code out of the collection is chosen at random, according to the distribution μ. We assume both the encoder and the decoder know the randomly selected code. The maximal distortion of the correlated code (over all state distributions in $\mathcal{Q}^n$) is bounded by D:

$$D \geq \sup_{q \in \mathcal{Q}^n} \sum_{\gamma \in \Gamma} \sum_{s^n} \sum_{u^n, v^n} \mu(\gamma) q(s^n) \psi(u^n, v^n|s^n) \; d\big(u^n, \varphi^\gamma(f^\gamma(u^n, s^n), v^n)\big). \quad (8)$$

We denote this code by $(\mu, \Gamma, \{(f^\gamma, \varphi^\gamma)\}_{\gamma \in \Gamma})$. A randomized-encoder code is a special case of a correlated code if all the deterministic codes in the collection have the same decoding function, $\varphi^\gamma = \varphi$. Note that the miss represented set is defined currently only for deterministic codes. This notion can be extended to the randomized-encoder and correlated codes classes, by defining it properly as random sets. However, for the purpose of the analysis presented in this work, the notion of miss represented sets for deterministic codes suffices.

Fix a source model and coding class. A rate distortion pair (R, D) is said to be *achievable*, for the selected source model with the selected coding class, if for every $\delta > 0$ and sufficiently large n, there exists an $(R, n, D + \delta)$ code from the selected class. The closure of the achievable rate-distortion pairs is called the *rate-distortion region*. The *rate-distortion function* is the boundary of the rate distortion region, i.e., the infimum of the achievable rates with distortion not more than D.

We denote by $R^{\mathcal{Q}}(D)$ (resp. $R^{\mathbb{Q}}(D)$, $R(D)$) the rate distortion function of the source $\{\mathcal{Q}, P_{U,V|S}\}$ (resp. $\{\mathbb{Q}, P_{U,V|S}\}$, $\{P_{U,V|S}\}$), under deterministic coding. Note that $\mathcal{Q}$ can contain a single distribution, in which case it is the random-parameter WZ source, with rate distortion function $R^q(D)$. The functions $R^q_{rnd}(D)$, $R^{\mathcal{Q}}_{rnd}(D)$, $R^{\mathbb{Q}}_{rnd}(D)$ and $R_{rnd}(D)$ are defined similarly, with random encoding, and the functions $R^q_{corr}(D)$, $R^{\mathcal{Q}}_{corr}(D)$, $R^{\mathbb{Q}}_{corr}(D)$ and $R_{corr}(D)$ with correlated encoding.

Cover and Chiang [3] characterized the rate-distortion function of a RP-WZ source, using deterministic codes. An application of their result to our model is stated next. It should be noted that this result can also be deduced directly from [6].

Theorem 1 ([3], [6]): The rate-distortion function of a $\{q, P_{U,V|S}\}$ RP-WZ source using deterministic codes is given by:

$$R^q(D) = \min_{\mathcal{M}(D)} \big[I(Z;S,U) - I(Z;V)\big] \tag{9}$$

where $\mathcal{M}(D)$ is the set of all random variables $\{S, U, V, Z, \hat{U}\}$ with joint probability distribution $P_{S,U,V,Z,\hat{U}} = qP_{U,V|S}P_{Z|U,S}P_{\hat{U}|Z,V}$, such that the average distortion between U and $\hat{U}$ is not more than D, i.e., $\mathbb{E}d(U,\hat{U}) \leq D$. The external random variable Z takes values in the finite set $\mathcal{Z}$. The distributions $P_{Z|U,S}$ and $P_{\hat{U}|Z,V}$ are subject to optimization. We remark that Theorem 1 can be derived directly from [6], by defining $\tilde{U} \triangleq (U,S)$, $P_{\tilde{U},V} \triangleq P_{U,V|S}P_S$ and considering the distortion measure $d(\tilde{U},\hat{U}) \triangleq d(U,\hat{U})$.

Inspection of the proof of the Theorem allows one to show that for every rate greater than the rate-distortion function, there exists a sequence (in n) of (R,n,D) codes such that the probability of the miss represented set is exponentially small with the blocklength n, i.e.,

$$P(\Delta(S^n)) = \sum_{s^n} q(s^n)\psi(\Delta(s^n)|s^n) \leq e^{-n\alpha}. \tag{10}$$

The exponential coefficient α is positive, and depends on $P_{S,U,V}$ and on δ. For later use, we denote by $\Delta^q(s^n)$ and by α_q, the miss represented set and its exponential coefficient, respectively, of a codes sequence that achieves the rate-distortion function $R^q(D)$ and (10) for the $\{q, P_{U,V|S}\}$ RP-WZ source. Note that the miss represented set depends only on the code (f,φ), and does not depend on the distribution q. We superscript Δ by q for notational convenience, to emphasize that it is the miss represented set of a code that achieves R and D with the state distribution q.

B. The rate-distortion function

Our first result is a characterization of an AV-WZ source rate-distortion function, under deterministic coding. It is stated in the next Theorem.

Theorem 2: The deterministic code rate-distortion function of the AV-WZ source $\{P_{U,V|S}\}$, is given by:

$$R(D) = \max_{q \in \mathcal{P}(\mathcal{S})} R^q(D). \tag{11}$$

We start the proof with some definitions. Let s^n be a sequence from $\mathcal{S}^n$. We denote its type by $\mathcal{T}(s^n)$. For a given type t, $\mathcal{S}(t,n)$ is the set of all n-length sequences with elements in $\mathcal{S}$ of type t. Finally, $\mathcal{T}(n,\mathcal{S})$ is the set of all types of n-length sequences with elements in $\mathcal{S}$.

The converse part of Theorem 2 follows almost immediately from Theorem 1 and the definition of an AV-WZ source. If an achievable rate-distortion pair (R,D) satisfies $R < \max_{q\in\mathcal{P}(\mathcal{S})} R^q(D)$, then there exists a probability distribution $q^* \in \mathcal{P}(\mathcal{S})$ such that $R < R^{q^*}(D)$. From the assumption that R is achievable, there exists an (n,R,D) deterministic code, with distortion not more than D for every state sequence s^n. Averaging the distortion with respect to q^* results in a contradiction to Theorem 1.

The achievability part of the Theorem is shown using arguments dual to those of Ahlswede [2]. The first step is to extend the RP-WZ source rate-distortion function (9) to the compound setup, namely to a $\{\mathcal{P}(\mathcal{S}), P_{U,V|S}\}$ source. As in [2], we preassign a probability distribution t' on $\mathcal{S}$, to every possible type on $\mathcal{S}$. The choice of t' is made so that the resulting performance with codes designed for t' is "close" (in a sense that will be made precise later) to the best possible performance with codes designed for the distributions in the class $\mathcal{Q}$. The actual type of S^n is passed to the decoder. Since the number of different types is only polynomial in n, this can be done with a code of negligible rate, not affecting the overall code rate. With the concatenated structure, every rate $R \geq \sup_{q\in\mathcal{Q}} R^q(D)$ can be achieved with distortion level not higher than D, and therefore $R^{\mathcal{Q}}(D) = \sup_{q\in\mathcal{Q}} R^q(D)$. In particular, it also follows that $R^{\mathcal{P}(\mathcal{S})}(D) = \max_{q\in\mathcal{P}(\mathcal{S})} R^q(D)$, as $\mathcal{P}(\mathcal{S})$ is closed.

Next, applying the robustification technique (RT), as stated in [2], allows one to show that given a deterministic code for a $\{\mathcal{P}(\mathcal{S}), P_{U,V|S}\}$ C-WZ source with distortion level D, a correlated code with the same distortion level can be constructed for the $\{P_{U,V|S}\}$ AV-WZ source. This implies that $R_{corr}(D) \leq R^{\mathcal{P}(\mathcal{S})}(D)$.

The elimination technique (ET) [1] is then used to show that the size of the codes space Γ (corresponding to the amount of correlation) does not have to be too large: In fact, only polynomial number (in blocklength) of codes in the set is sufficient. The polynomial correlation can be passed to the decoder, with a negligible increase in the rate. Thus we conclude that for that case, any rate achievable with a correlated code, is also achieved by a randomized encoder code. In other words, $R_{rnd}(D) = R_{corr}(D)$. Combining the ET result with the RT result yield that $R_{rnd}(D) \leq R^{\mathcal{P}(\mathcal{S})}(D)$.

Generally speaking, in an informed encoder scenario, a randomized-encoder code can not outperform a deterministic code, because for every fixed s^n and u^n

$$\sum_{i=1}^{2^{nR}} r(i|u^n,s^n)\mathbb{E}\big[d(u^n,\varphi(i,V^n))\big|u^n,s^n\big]$$

$$\geq \min_{i\in\{1,\ldots,2^{nR}\}} \Big[\mathbb{E}\big[d\left(u^n,\varphi(i,V^n)\right)\big|u^n,s^n\big]\Big], \quad (12)$$

where the expectation is w.r.t the RV V^n. Therefore, given a randomized-encoder code, one can define a deterministic code, by:

$$f(u^n,s^n) = \operatorname*{argmin}_{i\in\{1,\ldots,2^{nR}\}} \Big[\mathbb{E}\left[d\left(u^n,\varphi(i,V^n)\right)\big|u^n,s^n\right]\Big]. \quad (13)$$

A deterministic code defined in this way has a distortion level which is not higher than the original randomized-encoder code distortion level. Thus, the rate-distortion region achieved by deterministic codes is the same as the region achieved with randomized-encoder codes. In other words, $R(D) = R_{rnd}(D)$. Combining this observation with the ET and RT results yield that $R(D) \leq R^{\mathcal{P}(\mathcal{S})}(D)$. Since a $\{\mathcal{P}(\mathcal{S}), P_{U,V|S}\}$ C-WZ source is a special case case of an $\{P_{U,V|S}\}$ AV-WZ source, the opposite inequality, $R(D) \geq R^{\mathcal{P}(\mathcal{S})}(D)$, is obvious. This will conclude the proof of Theorem 2.

In the following subsections these results are proved formally.

1) C-WZ source rate-distortion function:

Theorem 3: Let $\{\mathcal{Q}, P_{U,V|S}\}$ be a C-WZ source. The deterministic code rate-distortion function is given by:

$$R^{\mathcal{Q}}(D) = \sup_{q\in\mathcal{Q}} R^q(D). \quad (14)$$

Proof: Theorem 1 asserts that for every $q\in\mathcal{Q}$ and for every rate $R > \sup_{q\in\mathcal{Q}} R^q(D)$, there exists a deterministic codes (f^q,φ^q) with distortion level not higher than D, with exponentially small miss represented set probability. Preassign for every type of state sequence $t\in\mathcal{T}(n,\mathcal{S})$, a unique index $i_t \in \{1,\ldots,|\mathcal{T}(n,\mathcal{S})|\}$ and a state distribution $t'\in\mathcal{Q}$ such that:

$$\sum_{s^n\in\mathcal{S}(t,n)} \psi(\Delta^{t'}(s^n)|s^n) \leq \inf_{q\in\mathcal{Q}} \sum_{s^n\in\mathcal{S}(t,n)} \psi(\Delta^q(s^n)|s^n) + e^{-\epsilon n} \quad (15)$$

for some fixed $\epsilon > 0$. Since the number of types is only polynomial in n, $|\mathcal{T}(n,\mathcal{S})| \leq (n+1)^{|\mathcal{S}|}$, the encoder can inform the type t to the decoder using a concatenation of the index i_t, without increasing the asymptotic rate. Define a concatenated deterministic code for every sequence $s^n\in\mathcal{S}(t,n)$ by:

$$f(u^n,s^n) = (f^{t'}(u^n,s^n), i_t) \quad (16)$$

$$\varphi((i,i_t),v^n) = \varphi^{t'}(i,v^n) \quad (17)$$

where $i\in\{1,\ldots,2^{nR}\}$. That is, the encoder concatenated the type index of s^n, i_t, with the index given by the appropriate preassigned encoding function $f^{t'}(u^n,s^n)$. The decoder uses the index i_t in order to approximate u^n using the appropriate decoder $\varphi^{t'}$. Let $\Delta^{\mathcal{Q}}(s^n)$ be the miss represented set of the suggested concatenated code. The following equation shows that for every given state probability $q\in\mathcal{Q}$, the miss represented set probability is exponentially small with blocklength.

$$\begin{aligned} & P(\Delta^{\mathcal{Q}}(S^n)) \\ &= \sum_{s^n} q(s^n)\psi(\Delta^{\mathcal{Q}}(s^n)|s^n) && (18)\\ &\overset{(a)}{=} \sum_{t\in\mathcal{T}(n,\mathcal{S})} q(\mathcal{S}(t,n)) \sum_{s^n\in\mathcal{S}(t,n)} \psi(\Delta^{t'}(s^n)|s^n) && (19)\\ &\overset{(b)}{\leq} \sum_{t\in\mathcal{T}(n,\mathcal{S})} q(\mathcal{S}(t,n)) \inf_{\tilde{q}\in\mathcal{Q}} \sum_{s^n\in\mathcal{S}(t,n)} \psi(\Delta^{\tilde{q}}(s^n)|s^n) + e^{-\epsilon n} && (20)\\ &\overset{(c)}{\leq} \inf_{\tilde{q}\in\mathcal{Q}} \sum_{s^n} q(s^n)\psi(\Delta^{\tilde{q}}(s^n)|s^n) + e^{-\epsilon n} && (21)\\ &\leq e^{-\alpha_q n} + e^{-\epsilon n} && (22)\end{aligned}$$

where (a) holds because all sequences in $\mathcal{S}(t,n)$ have the same probability, (b) holds by equation (15), and (c) holds because summation of infimas is less than infimum of the summation. This analysis shows that $R^{\mathcal{Q}}(D) \leq \sup_{q\in\mathcal{Q}} R^q(D)$. The opposite inequality is obvious. This completes the proof of Theorem 3. ∎

In the proof of Theorem 2 we will interested in the set $\mathcal{Q} \triangleq \mathcal{P}(\mathcal{S})$. Application of Theorem 3 to $\mathcal{P}(\mathcal{S})$ is stated in the next corollary.

Corollary 4: Let $\{\mathcal{P}(\mathcal{S}), P_{U,V|S}\}$ be a C-WZ source. The deterministic code rate-distortion function is given by:

$$R^{\mathcal{P}(\mathcal{S})}(D) = \max_{q\in\mathcal{P}(\mathcal{S})} R^q(D). \quad (23)$$

Proof: The corollary is a simple application of Theorem 3, the compactness of the set $\mathcal{P}(\mathcal{S})$ and the continuity of $R^q(D)$ as a function of q. ∎

2) Application of the Robustification Technique: Let Π_n be the set of all permutations of length n. We denote members of Π_n simply by π. We next quote the RT, as stated in [2].

Let $h: \mathcal{S}^n \to [0,1]$ be a given function. If, for some fixed $\alpha\in(0,1)$, and for every memoryless distribution $q\in\mathcal{P}(\mathcal{S})$, the function h satisfies:

$$\sum_{s^n} q(s^n)h(s^n) \leq \alpha, \quad (24)$$

then h satisfies for every $s^n\in\mathcal{S}^n$

$$\frac{1}{n!}\sum_{\pi\in\Pi_n} h(\pi s^n) \leq \alpha_n, \quad (25)$$

where $\alpha_n \triangleq \alpha(n+1)^{|\mathcal{S}|}$.

We use the RT following [2]. By corollary 4, every rate $R > R^{\mathcal{P}(\mathcal{S})}(D)$ can be achieved with a deterministic code such that the miss represented set $\Delta^{\mathcal{P}(\mathcal{S})}(s^n)$ has a probability not more than $\alpha = e^{-\epsilon_1 n}$, $\epsilon_1 > 0$. In other words, for every $q\in\mathcal{P}(\mathcal{S})$

$$\sum_{s^n} q(s^n)\psi(\Delta^{\mathcal{P}(\mathcal{S})}(s^n)|s^n) \leq \alpha. \quad (26)$$

Define $h(s^n)$ by:

$$h(s^n) = \psi(\Delta^{\mathcal{P}(\mathcal{S})}(s^n)|s^n). \quad (27)$$

$h(s^n)$ fulfills the RT hypothesis (24). Therefore, for every $s^n \in \mathcal{S}^n$:

$$\frac{1}{n!}\sum_{\pi\in\Pi_n}\psi(\Delta^{\mathcal{P}(\mathcal{S})}(\pi s^n)|\pi s^n) \le \alpha_n \tag{28}$$

were $\alpha_n = e^{-\epsilon_1 n}(n+1)^{|\mathcal{S}|} = exp\left\{-n(\epsilon_1 - |\mathcal{S}|\frac{ln(n+1)}{n})\right\}$.

Next, let us define a collection of deterministic codes as follows. For every $\pi \in \Pi_n$ define:

$$f^\pi(u^n, s^n) = f(\pi u^n, \pi s^n) \tag{29}$$

$$\varphi^\pi(i, v^n) = \pi^{-1}\varphi(i, \pi v^n) \tag{30}$$

Let $\Delta^\pi(s^n)$ be the miss represented set of the deterministic code (f^π, φ^π). Observe that:

$$\begin{aligned}
&\Delta^\pi(s^n)\\
&\stackrel{(a)}{=} \{(u^n,v^n) : d(u^n, \varphi^\pi(f^\pi(u^n,s^n),v^n)) > D+\delta\} && (31)\\
&\stackrel{(b)}{=} \{(u^n,v^n) : d(u^n, \pi^{-1}\varphi(f(\pi u^n,\pi s^n),\pi v^n)) > D+\delta\} && (32)\\
&\stackrel{(c)}{=} \{(u^n,v^n) : d(\pi u^n, \varphi(f(\pi u^n,\pi s^n),\pi v^n)) > D+\delta\} && (33)\\
&\stackrel{(d)}{=} \pi^{-1}\{(u^n,v^n) : d(u^n, \varphi(f(u^n,\pi s^n),v^n)) > D+\delta\} && (34)\\
&= \pi^{-1}\Delta^{\mathcal{P}(\mathcal{S})}(\pi s^n) && (35)
\end{aligned}$$

where (a) is by the definition of $\Delta^\pi(s^n)$, (b) is by the definition of the code (f^π, φ^π) (29) and (30), (c) is a result of the definition of the distortion measure (1), and (d) is obtained by the substitution of variables $\pi u^n \to u^n$ and $\pi v^n \to v^n$.

Define a correlated code $(\mu, \Pi_n, \{(f^\pi, \varphi^\pi)\}_{\pi\in\Pi_n})$ by letting μ be a uniform distribution on the set Π_n. Let $\Delta(s^n)$ be the miss represented set of this code. We will upper bound the miss represented set probability for every fix $s^n \in \mathcal{S}^n$ by:

$$\begin{aligned}
&\psi(\Delta(s^n)|s^n)\\
&\stackrel{(a)}{=} \frac{1}{n!}\sum_{\pi\in\Pi_n}\psi(\Delta^\pi(s^n)|s^n) && (36)\\
&\stackrel{(b)}{=} \frac{1}{n!}\sum_{\pi\in\Pi_n}\psi(\pi^{-1}\Delta^{\mathcal{P}(\mathcal{S})}(\pi s^n)|s^n) && (37)\\
&\stackrel{(c)}{=} \frac{1}{n!}\sum_{\pi\in\Pi_n}\psi(\Delta^{\mathcal{P}(\mathcal{S})}(\pi s^n)|\pi s^n) && (38)\\
&\stackrel{(d)}{\le} \alpha_n && (39)
\end{aligned}$$

where (a) is from the definition of the correlated code, (b) is from (35), (c) is from the memoryless structure of the conditional probability of the WZ source (2) and (d) is the result of the RT (28). Therefore, we can conclude that every rate $R > R^{\mathcal{P}(\mathcal{S})}(D)$ can be achieved with a correlated code. In other words, $R_{corr}(D) \le R^{\mathcal{P}(\mathcal{S})}(D)$.

3) Elimination Technique for rate-distortion: The application of the RT in subsection II-B.2 shows that the rate-distortion function of an AV-WZ source under correlated coding is not more than that of a C-WZ source rate-distortion function under deterministic coding. The common randomness used in a correlated codes is their major drawback. We will state and prove next a Theorem based on Ahlswede's Elimination Technique (ET) [1] showing that the size of the codes space Γ (corresponding to the amount of correlation) does not have to be too large: In fact, only polynomial large size in blocklength n is sufficient. Thus, the polynomial correlation can be passed to the decoder using a concatenation of codes, concluding that any rate achievable with a correlated code, can also be achieved by a randomized-encoder code. The Theorem is stated and proved next.

Theorem 5: Let $\{P_{U,V|S}\}$ be an AV-WZ source and let $(\mu, \Gamma, \{f^\gamma, g^\gamma\}_{\gamma\in\Gamma})$ be a sequence of correlated codes with miss represented set probability $e^{-\epsilon n}$, $\epsilon > 0$. Then, for every fixed $\lambda \in (0,1)$ and every sufficiently large n there exists a correlated code $(\mu^*, \Gamma^*, \{f^\gamma, g^\gamma\}_{\gamma\in\Gamma^*})$ such that the code collection is a subset of the original code collection with $|\Gamma^*| \le n^2$, and the miss represented set probability is not more than λ.

Proof: Let L be an integer to be chosen later, and let L_j, $j = 1, \ldots, L$ be a sequence of iid RVs with probability μ. Fix s^n and define $K_j(s^n)$ by:

$$K_j(s^n) = \psi(\Delta^{L_j}(s^n)|s^n). \tag{40}$$

$\Delta^{L_j}(s^n)$ is the miss represented set of the deterministic code (f^{L_j}, φ^{L_j}). The expectation of $K_j(s^n)$, with respect to L_j satisfies $\mathbb{E}K_j(s^n) \le e^{-\epsilon n}$. Note that the expectation is w.r.t. L_j only, and s^n is a specific sequence held fixed. Moreover, $K_j(s^n)$, $j = 1, 2, \ldots, L$, are independent of each other.

Fix n sufficiently large such that $\lambda > e^{-\epsilon n}$. Then, for every $\beta > 0$:

$$\begin{aligned}
&P\left(\sum_{j=1}^{L} K_j(s^n) \ge L\lambda\right) \le\\
&\qquad \mathbb{E}exp\left\{\beta\left(\sum_{j=1}^{L} K_j(s^n) - L\lambda\right)\right\} \le\\
&\qquad\qquad e^{-\beta L\lambda}\left(1 + e^{\beta}e^{-\epsilon n}\right)^L. && (41)
\end{aligned}$$

The last inequality holds since for $x \in [0,1]$ and $\beta > 0$, we have $e^{\beta x} \le 1 + xe^{\beta}$. Choosing $\beta = 2$, $L = n^2$ and upper bounding $\left(1 + e^{\beta}e^{-\epsilon n}\right)$ (for sufficiently large n) by e^{λ} produce:

$$P\left(\frac{1}{n^2}\sum_{j=1}^{n^2} K_j(s^n) \ge \lambda\right) \le e^{-\lambda n^2}. \tag{42}$$

Applying the union bound gives:

$$P\left(\max_{s^n}\left[\frac{1}{n^2}\sum_{j=1}^{n^2} K_j(s^n) \ge \lambda\right]\right) \le |\mathcal{S}|^n e^{-\lambda n^2}. \tag{43}$$

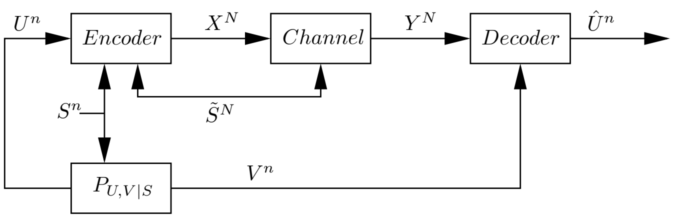

Fig. 2. Joint source-channel coding for the arbitrarily varying Wyner-Ziv source and Gel'fand-Pinsker channel, with side information at the encoder.

Since $|\mathcal{S}|^n$ grows (only) exponentially with n, we obtain a super exponential bound for the average miss represented set probability to exceed λ, for any $s^n \in \mathcal{S}^n$. Therefore, random selection argument guaranties the exists of a correlated code $(\mu^*, \Gamma^*, \{f^\gamma, g^\gamma\}_{\gamma\in\Gamma^*})$, as in the Theorem. This completes the proof. ∎

III. JOINT AV-WZ SOURCE AND AV-GP CHANNEL

In this section we will study the problem of joint source-channel coding for the arbitrarily varying WZ source and the Gel'fand-Pinsker (GP) channel [4].

A GP channel is a memoryless channel given by the transition probability $w(y|x,\tilde{s})$, where x takes values in the channel finite input alphabet $\mathcal{X}$, and y takes values in the channel finite output alphabet $\mathcal{Y}$. $\tilde{s}$ is a state variable, takes values in the finite set $\tilde{\mathcal{S}}$, and changes according to some distribution, or arbitrarily, as will be defined later on.

Consider the system described in Fig. 2. The WZ source operates at a rate of ρ_s symbols per second. We would like to encode a sequence of n source symbols u^n, into a sequence of N channel input symbols, x^N, that will be transmitted over the channel operating at a rate of ρ_c channel uses per second. The integers n and N are selected such that $\frac{n}{N} = \frac{\rho_s}{\rho_c}$. ρ_s and ρ_c are sometimes referred to as the bandwidth expansion ratios of the source and the channel, respectively, or just rates. We will assume that the encoder has an access to the source state sequence, s^n, and the channel state sequence, $\tilde{s}^N$. The aim of the decoder is to use the channel output sequence y^N and the SI sequence v^n, in order to approximate the source sequence, with the sequence $\hat{u}^n$, where $\hat{u}_i$ takes values in the finite reproduction alphabet $\hat{\mathcal{U}}$. The quality of the joint source-channel coding is evaluated w.r.t. the fidelity criterion defined in (1). We are interested in the minimal achievable distortion level, for given pair of rates ρ_s, ρ_c.

A. Definitions and known results

We first introduce the notation of the channel, and briefly describe known results on the channel capacity. Consider the channel $w(y|x,\tilde{s})$ described in Fig. 3. If the state sequence is chosen in an iid manner with known probability $P_{\tilde{S}} = \tilde{q}$, we will regard the channel as a *Gel'fand-Pinsker (GP)* channel, and will denote it by $\{\tilde{q}, w(y|x,\tilde{s})\}$. We will denote a code of length N, rate R and average error probability P_e by (N, R, P_e), and the channel capacity by $C^{\tilde{q}}$.

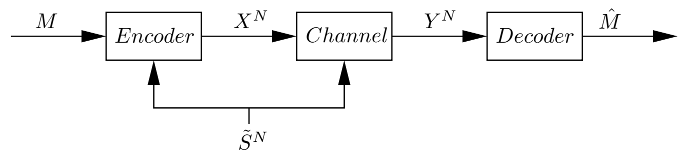

Fig. 3. An arbitrarily varying Gel'fand-Pinsker channel

The capacity formula for this channel was found by Gel'fand and Pinsker [4], and is given by:

$$C^{\tilde{q}} = \max_{\mathcal{M}} \left[I(Z;Y) - I(Z;\tilde{S}) \right] \tag{44}$$

where $\mathcal{M}$ is the set of all probability distribution on the quartet of random variables $\{\tilde{S}, X, Y, Z\}$, with joint probability distribution $P_{\tilde{S},X,Y,Z} = \tilde{q} \cdot w \cdot P_{Z|\tilde{S}} \cdot P_{X|Z,\tilde{S}}$. The external random variable Z takes values in the finite set $\mathcal{Z}$. The distributions $P_{Z|\tilde{S}}$ and $P_{X|Z,\tilde{S}}$ are subject to optimization.

If the state sequence of length n is chosen with unknown probability from the set $\mathcal{P}(\tilde{S}^n)$, we will regard the channel as an *Arbitrarily Varying GP (AV-GP)* channel, and will denote it by $\{w(y|x,\tilde{s})\}$. Exactly like for the AV-WZ source, one can talk about various coding schemes – deterministic, randomized-encoder, and correlated. Here, we will consider only deterministic coding scheme. The (deterministic code) capacity of an $\{w(y|x,\tilde{s})\}$ AV-GP, denoted by C, was found by Ahlswede [2]. His result is stated in the next Theorem.

Theorem 6 ([2]): Let $\{w(y|x,\tilde{s})\}$ be an AV-GP. The (deterministic code) capacity is positive if and only if for every state $\tilde{s}$ the discrete memoryless channel $w(y|x,\tilde{s})$ has a positive capacity. If the capacity is positive, then it is given by:

$$C = \min_{\tilde{q}\in\mathcal{P}(\tilde{\mathcal{S}})} C^{\tilde{q}}. \tag{45}$$

We return to the communication system described in Fig. 2. Let n and N be integers such that $n/N = \rho_s/\rho_c$. An (n, N, D) joint source-channel code is defined by an encoder function $f : \mathcal{U}^n \times \mathcal{S}^n \times \tilde{\mathcal{S}}^N \to \mathcal{X}^N$ and a decoder function $\varphi : \mathcal{Y}^N \times \mathcal{V}^n \to \hat{\mathcal{U}}^n$, such that the maximal distortion (over all state distributions in $\mathcal{P}(\mathcal{S}^n)$ and in $\mathcal{P}(\tilde{\mathcal{S}}^n)$) is bounded by D:

$$\begin{aligned} D \geq \max_{q\in\mathcal{P}(\mathcal{S}^n)} \max_{\tilde{q}\in\mathcal{P}(\tilde{\mathcal{S}}^N)} \sum_{s^n,\tilde{s}^N} \sum_{u^n,v^n,y^N} q(s^n)\tilde{q}(\tilde{s}^N) \\ \psi(u^n, v^n|s^n) w(y^N|f(u^n, s^n, \tilde{s}^N), \tilde{s}^N) \\ d(u^n, \varphi(y^N, v^n)). \end{aligned} \tag{46}$$

A rate-distortion level D is said to be *achievable* if for every $\delta > 0$ and every sufficiently large n and N there exists an $(n, N, D + \delta)$ code.

A separation Theorem for a random parameters version of the described communication system, that is, a RP-WZ source transmitted over a GP channel, was stated by Merhav and Shamai [5]. An application of their result to our model is stated in the next Theorem.

Theorem 7: ([5]) Let $\{q, P_{U,V|S}\}$ be a RP-WZ source transmitted over a $\{\tilde{q}, w(y|x,s)\}$ GP channel. The distortion level D is achievable if and only if:

$$\rho_s R^q(D) \leq \rho_c C^{\tilde{q}}. \quad (47)$$

That is, for any given pair of rates ρ_c, ρ_s, no loss in asymptotic distortion is incurred by separating the source coding from the channel coding.

B. A Separation Theorem for the arbitrarily varying model

Our result in this section is a separation Theorem for the communication model described above. It is stated and proved next.

Theorem 8: Let $\{P_{U,V|S}\}$ and $\{w(y|x,s)\}$ be arbitrarily varying Gel'fand-Pinsker channel, and Wyner-Ziv source, respectively. A distortion level D with given rates ρ_s, ρ_c is achievable if and only if

$$\rho_s R(D) \leq \rho_c C. \quad (48)$$

Theorem 8 asserts that no loss is caused when the joint source-channel encoder is constructed in two independent stages, as described in Fig. 4. First, the encoder encode the source with an AV-WZ source encoder, using s^n and ignoring $\tilde{s}^N$. Then, the result is encoded with an AV-GP encoder, using $\tilde{s}^N$ and ignoring s^n. Similarly, the decoder first decode the channel output with an AV-GP decoder, ignoring v^n. Then, the source sequence is approximated with an AV-WZ source decoder using v^n.

Proof: We start with the sufficiently part. Select a distortion level D such that $\rho_s R(D) < \rho_c C$. Fix a source rate R_s and channel rate R_c satisfying $\rho_s R(D) < \rho_s R_s = \rho_c R_c < \rho_c C$. Fix some $\delta > 0$. Theorem 2 guaranties that for every $\delta' > 0$ there exists a source code with rate R_s and distortion level not more than $D + \delta'$. Theorem 6 guaranties that for every $\epsilon > 0$ there exists a channel code with rate R_c and error probability not more than ϵ. Create a joint source-channel code by concatenating the source code with the channel code as in Fig. 4. The distortion level of the concatenated code for every given s^n and $\tilde{s}^N$ can be upper bound by:

$$\mathbb{E}\left[d(U^n, \hat{U}^n)\big|s^n, \tilde{s}^N\right] \leq (1-\epsilon)(D+\delta') + \epsilon D_{max} \quad (49)$$

where D_{max} is an upper bound to the (single letter) distortion measure. Selecting δ' and ϵ such that $\delta > (1-\epsilon)\delta' + \epsilon(D_{max} - D)$ completes the suffieciency part.

The necessity part is shown by a contradiction to Merhav-Shamai Theorem [5]. Assume that there exists an achievable distortion level D such that $\rho_s R(D) > \rho_c C$. Therefore, there

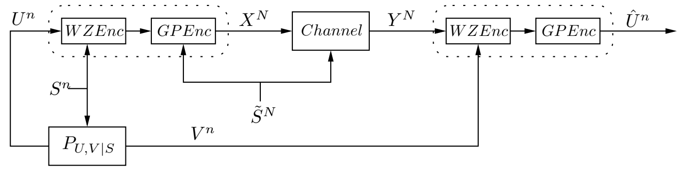

Fig. 4. An optimal separated scheme of a joint source-channel coding for the arbitrarily varying Wyner-Ziv source and Gel'fand-Pinsker channel, with side information at the encoder.

exist $q \in \mathcal{P}(\mathcal{S})$ and $\tilde{q} \in \mathcal{P}(\tilde{\mathcal{S}})$ such that $\rho_s R^q(D) > \rho_c C^{\tilde{q}}$. Since the distortion level D is achievable, there exists a joint source-channel code with distortion not more than D for every source and channel state sequences, s^n and $\tilde{s}^N$. In particular, averaging the distortion level of this code with respect to q and $\tilde{q}$ has a distortion level not more than D, which contradicts Theorem 7. This completes the proof of Theorem 8. ∎

It is interesting to examine the result of Theorem 8 when $\rho_s = \rho_c = 1$. In this case, the two state sequences have the same rate, and are synchronized. The result implies that although the state sequences are synchronized, the best strategy for the jammer is to choose the states so that they look like independent of each other. Another interpretation of this result is that a separation principle holds also with respect to the operation of the jammer: the optimal jammer can be split into two non cooperating jammers, one of which has access only to the state sequence of the source, and the other has access only to the state sequence of the channel.

REFERENCES

[1] R. F. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 44, pp. 159–175, 1978.

[2] ——, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. Inform. Theory*, vol. 32, no. 5, pp. 621–629, Sept. 1986.

[3] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1629–1638, June 2002.

[4] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[5] N. Merhav and S. Shamai, "On joint source-channel coding for the Wyner-Ziv source and the Gel'fand-Pinsker channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 2844–2855, Nov. 2003.

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