IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Joint Source Channel Coding for Fading Channel: The Broadcast Approach-Expected Distortion, Gaussian Source and Channel

C. Tian, A. Steiner, S. Shamai (Shitz) and S. Diggavi

CCIT Report #709
November 2008

Electronics
Computers
Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT REPORT #709 November 2008

Joint Source Channel Coding for Fading Channel: The Broadcast Approach-Expected Distortion, Gaussian Source and Channel

Chao Tian*, Avi Steiner†, Shlomo Shamai (Shitz)† and Suhas Diggavi*
*School of Computer and Communication Sciences
Laboratory for Information and Communication Systems (LICOS), EPFL, Lausanne, CH1015, Switzerland.
†Dept. of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel.
Email: *{chao.tian, suhas.diggavi}@epfl.ch, †savi@tx.technion.ac.il, sshlomo@ee.technion.ac.il

***Abstract*— We consider the problem of transmitting a Gaussian source on a slowly fading Gaussian channel, subject to the mean squared error distortion measure. The channel state information is known only at the receiver but not the transmitter. The source is assumed to be encoded in a successive refinement manner, and then transmitted over the channel using the broadcast strategy. In order to minimize the expected distortion at the receiver, optimal power allocation is essential. We propose an efficient algorithm to compute the optimal solution in linear time $O(M)$. Moreover, we provide a derivation of the optimal power allocation when the fading state is a continuum, using the classical variational method. The proposed algorithm as well as the continuous solution is based on an alternative representation of the capacity region of the Gaussian broadcast channel.**

I. Introduction

Fading channel occurs naturally as a model in wireless communications. For slow fading, the receiver can usually recover the channel state information (CSI) accurately, however the transmitter only knows the probability distribution of CSI, but not the realization. Such uncertainty can cause significant performance degradation, and the broadcast strategy was used in [1] as an approach to combat this effect. In this strategy, some information can only be decoded when the fading is less severe, which is superimposed on the information that can be decoded under more severe fading. Thus the receiver can decode the information adaptively, according to the realization of the channel state. The similarity to the degraded broadcast channel [2] is clear in this context, particularly for channels with finite number of fading states. Generalizing this view, when the fading gain can take continuous values, the receiver can be taken as a continuum of users in a broadcast channel.

The broadcast strategy naturally matches the successive refinement (SR) source coding framework [3], as the information decodable under the strongest fading is protected the most, and should be used to convey the base layer information in the SR coding. As more information can be decoded when the channel is subject to less fading, more SR coded layers can be decoded, and the reconstruction quality improves. In this work, we consider this scenario for a quadratic Gaussian source. In order to minimize the expected distortion at the receiver, it is essential to find the optimal power allocation in the broadcast strategy, and this is indeed our focus. This cross layer design approach was in fact already suggested in [1].

Initial effort on this problem was made by Sesia *et al.* in [4], where the broadcast strategy coupled with SR source coding was compared with several other schemes. Etemadi and Jafarkhani also considered this problem in [5], and provided an iterative algorithm, by separating the optimization problem into two sub-problems. In two more recent works [6] [7], Ng *et al.* provided a recursive algorithm to compute the optimal power allocation for M fading states, with worst case complexity of $O(2^M)$; moreover, by directly taking the limit of the optimal solution for the discrete case, a solution was given for the continuous case optimal power allocation, under the assumption that the optimal power allocation is concentrated in a single interval. Similar problems were considered in [8] [9] in the high SNR regime from the perspective of distortion exponent.

Our contribution in the present work is two-fold: firstly, we propose a new algorithm that computes the optimal power allocation for M fading states with $O(M)$ complexity, i.e., in linear time; secondly, we provide a derivation of the continuous case optimal power allocation solution by the classical variational method [10]. Both the algorithm and the derivation rely on an alternative representation of the Gaussian broadcast channel capacity, which appeared in [11].

The rest of the paper is organized as follows. In Section II we give the system model, and in Section III the new algorithm is provided and its optimality is proved. In Section IV we give the derivation for the continuous case solution, and Section V concludes the paper.

II. System model and Preliminaries

We assume the memoryless source $\{X_i\}_{i=1}^{\infty}$ is generated independently and identically according to a zero-mean unit variance Gaussian distribution. To simplify the notation, we directly assume the channel in the real domain as

$$Y_c = \sqrt{s}X_c + N, \quad (1)$$

where X_c is the real-valued channel input and Y_c is the channel output, $s \in \mathbb{R}$ is the (random) channel power gain, and N is the zero-mean unit variance Gaussian additive noise in the channel. Extension to complex system with circular symmetric complex noise is straightforward.

We consider a slowly fading channel model, where each channel codeword consists of a length-l_c channel symbol block. Source symbols of block length-l_s is encoded into a single channel codeword, and there is a source channel mismatch factor $b = l_c/l_s$; see also Fig. 1. Each channel block is assumed to be sufficiently long to approach channel

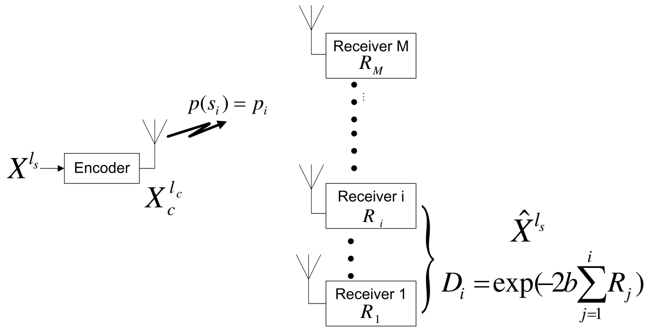

Fig. 1. The broadcast approach for minimizing the expected distortion.

capacity, as well as the rate-distortion limit, however still much shorter than the dynamics of the slowly fading process.

For the case with a finite number of fading states, the M possible power gains in an increasing order $s_1 < s_2 < ... < s_M$ are distributed according to a probability mass function p_i such that $\sum_{i=1}^{M} p_i = 1$. The transmitter has a power constraint P, and if power P_i is allocated to the i-th layer in the broadcast strategy, the i-th layer channel rate R_i is given by

$$R_i = \frac{1}{2}\log(1 + \frac{P_i}{1/s_i + \sum_{j=i+1}^{M} P_j}), \tag{2}$$

where we use natural log. The equivalence to broadcast on a set of channels with different noise variances is clear. Let $n_i \triangleq 1/s_i$, which implies $n_1 > n_2 > ... > n_M$ are the equivalent noise power on the channels.

Since the Gaussian source is successively refinable [3], the receiver with power gain s_i can thus achieve the distortion

$$D_i = \exp(-2b \sum_{j=1}^{i} R_j). \tag{3}$$

Combining (2) and (3), the problem is essentially the following minimization over the power allocation $(P_1, P_2, ..., P_M)$,

$$\min \sum_{i=1}^{M} p_i \left(\prod_{j=1}^{i} (1 + \frac{P_i}{1/s_i + \sum_{k=i+1}^{M} P_k}) \right)^{-b} \tag{4}$$

$$\text{subject to:} \quad P_i \geq 0, \quad i = 1, 2, ..., M, \quad \sum_{i=1}^{M} P_i \leq P.$$

When the fading state is continuous, the density of the power gain distribution is then given by $f(s)$, which is assumed to be continuously differentiable. The goal is then to find a power allocation density function $P(s)$, or its cumulative function, which minimizes the expected distortion.

III. New Algorithm and Its Optimality

The problem in (4) is not convenient for optimization due to its complicated form. In fact it is not immediately clear that the function being optimized is a convex function of $(P_1, P_2, ..., P_M)$, but it is clear that the rate-region of the broadcast channel is convex. Thus we seek another characterization of $\boldsymbol{R} = (R_1, R_2, ..., R_M)$ for the broadcast channel capacity. An alternative characterization was given in [11], and we now translate it into our notation.

A. Rederivation of the equivalent representation

The rate vector $\boldsymbol{R}$ on the boundary region corresponds to a power allocation $P_1, P_2, ..., P_M$. Solving the value of P_i in terms of the rate vector $\boldsymbol{R}$ gives an equivalent system of equations, and through further simplification, we have[1]

$$\sum_{i \geq m} P_i = \sum_{i \geq m} (n_i - n_{i+1}) \exp\left(2 \sum_{j=m}^{i} R_j\right) - n_m, \quad m = 1, 2, 3, ..., M, \tag{5}$$

where we define $n_{M+1} \triangleq 0$. The RHS of the above equation is monotonically decreasing in m, hence for any rate vector $\boldsymbol{R}$, provided that (5) is satisfied for $m = 1$, i.e.

$$P = \sum_{i \geq 1} (n_i - n_{i+1}) \exp\left(2 \sum_{j=1}^{i} R_j\right) - n_1,$$

there must be a power allocation such that (5) is satisfied for all m, and it is on the boundary of the capacity region. Thus we can alternatively characterize the capacity region as

$$\mathcal{C} = \left\{ (R_1, R_2, ..., R_M) : R_m \geq 0, \; m = 1, 2, ..., M, \; \sum_{i=1}^{M} (n_i - n_{i+1}) \exp\left(2 \sum_{j=1}^{i} R_j\right) - n_1 \leq P \right\}.$$

Moreover, the above function is convex in $(R_1, R_2, ..., R_M)$.

We can now reformulate the optimization problem as a convex programming problem:

$$\min \sum_{i=1}^{M} p_i \exp(-2b \sum_{j=1}^{i} R_j)$$

$$\text{subject to:} \quad R_i \geq 0, \quad i = 1, 2, ..., M, \quad \sum_{i=1}^{M} (n_i - n_{i+1}) \exp\left(2 \sum_{j=1}^{i} R_j\right) - n_1 \leq P.$$

From the optimal rate vector $\boldsymbol{R}$ we can then recover the corresponding power allocation.

B. The Lagrangian formulation and the algorithm

Now consider the Lagrangian form

$$L = \sum_{i=1}^{M} p_i \exp(-2b \sum_{j=1}^{i} R_j) - \sum_{i=1}^{M} \nu_i R_i + \lambda \left(\sum_{i=1}^{M} (n_i - n_{i+1}) \exp\left(2 \sum_{j=1}^{i} R_j\right) - n_1 - P \right).$$

The Karush-Kuhn-Tucker (KKT) condition requires that $\frac{\partial L}{\partial R_m} = 0$ for the optimal solution, and taking the difference between $\frac{\partial L}{\partial R_m} = 0$ and $\frac{\partial L}{\partial R_{m+1}} = 0$ gives

$$-2bp_m \exp(-2b \sum_{j=1}^{m} R_j) + 2\lambda(n_m - n_{m+1}) \exp\left(2 \sum_{j=1}^{m} R_j\right) = \nu_m - \nu_{m+1}, \quad m = 1, 2, ..., M, \tag{6}$$

[1]The equations given in [11] appears to have a minor mistake that the inner sum was given as $\sum_{j=1}^{i} R_j$.

where $\nu_{M+1} \triangleq 0$, for some $\nu_m \geq 0$, $m = 1, 2, ..., M$. Furthermore, the complementary slackness requires $\nu_m R_m = 0$, $m = 1, 2, ..., M$; the power constraint should be satisfied with equility, or $\lambda = 0$.

Since the optimization problem is a convex programming problem, the KKT condition is both necessary and sufficient for an optimal solution. Clearly if the quantity

$$\kappa_m \triangleq \frac{bp_m}{n_m - n_{m+1}}$$

is monotonically increasing, we can set $\nu_i = 0$ for $i = 1, 2, ..., M$ and find an explicit solution, provided the power constraint is not violated; however this is not true in general.

We can use the following algorithm to find the optimal rate allocation. When a layer is assigned zero rate, it will be called "ineffective"; otherwise it will be called "effective". For simplicity, define $\kappa_{M+1} = \infty$. A layer is labeled "active" if it is a result of combination of layers in the previous loop. An intuitive explanation is given in the next subsection.

1) Combination of layers to reach a monotonic κ sequence.
   a) Assign $\Delta n_m = n_m - n_{m+1}$ and calculate κ_m for $m = 1, 2, ..., M$. Label all the layers effective and active. Let $r = 1$.
   b) Denote the lower effective neighbor of layer i as i^-, and its upper effective neighbor layer as i^+. Start from $i_k = i_1$, for all the a_r active layers $i_1, i_2, ..., i_{a_r}$:
      i) If $i_k > 1$ and $\kappa_{i_k^-} \geq \kappa_{i_k}$: label layer i_k ineffective and combine it with its **current** lower effective neighbor layer j. Update $p_j = p_j + p_{i_k}$, $\Delta n_j = \Delta n_j + \Delta n_{i_k}$, as well as κ_j values accordingly. Label j as active.
      ii) If $\kappa_{i_k} \geq \kappa_{i_k^+}$: label layer i_k^+ ineffective and combine it with its **current** lower effective neighbor layer j. Update $p_j, \Delta n_j$ and κ_j values accordingly. Label j as active.
      iii) If $k < a_r$, increment k by 1 and return to step (1(b)i).
   c) If after the above loop, any layer remains active: increment r by 1 and return to step (1b).
2) Denote the number of effective layers by K. For all the effective layers i_k, $k = 1, 2, ..., K$, let $\exp(2R_{i_k}) = \kappa_{i_k}^{1/(b+1)} / \kappa_{i_{k-1}}^{1/(b+1)}$; in other words for all the effective layers we have $\exp(2\sum_{j=1}^{i_k} R_j) = \kappa_{i_k}^{1/(b+1)}$. Assign the ineffective layers rate zero.
3) Check power consumption.
   a) Let i_{k_o} be the lowest effective layer, and define $P_n = P + n_{i_{k_o}}$.
   b) Let
$$\lambda^{1/(b+1)} = \frac{\sum_{k=k_o}^{k_K} (n_{i_k} - n_{i_{k+1}}) \exp\left(2\sum_{j=1}^{i_k} R_j\right)}{P_n}.$$
   c) If
$$\frac{bp_{i_{k_o}}}{\lambda(n_{i_{k_o}} - n_{i_{k_o+1}})} \geq 1, \quad (7)$$
   then reduce $R_{i_{k_o}}$ by $\frac{1}{2(b+1)} \log \lambda$; otherwise, label i_{k_o} ineffective ($R_{i_{k_o}} = 0$), increment k_o by 1, update $P_n = P + n_{i_{k_o}}$, and return to step (3b).

C. The correctness of the algorithm and its complexity

Intuitively speaking, we classify the layers into two kinds: those with κ value lower than or equal to its lower neighbor (*the first kind*), and those with κ value higher than its lower neighbor (*the second kind*). The algorithm combines the first kind layers in each step, and then continues this operation until no layers of the first kind exist in the resulting sequence. The resulting rate allocation is valid, if the κ sequence is indeed monotonically increasing and the power constraint is satisfied.

A few more comments are in order: 1) In step 1, we want to form a monotonic sequence of κ_i by combining consecutive layers, such that step 2 can provide meaningful rates. To do this, we combine (remove) all the layers that are monotonically non-increasing. Only the neighbors of those whose κ value were updated in the previous loop need to be considered, because this is the only case that a change of classification may occur. 2) In step 3, we need to assure the total power is used up by adjusting the value of λ. However, this has to be done such that the lower layer still has positive rate, which is the condition in (7) If this is not possible, the lowest effective layer is eliminated; this condition is checked repeatedly for the reduced layers until it is satisfied. 3) In the loop of step (1b), we emphasize the layer is combined with its current effective lower layer, because the layer i_k^- (or i_k) may become ineffective in the previous steps.

The complexity of the algorithm is $O(M)$. Step 2 is clearly of $O(M)$ complexity. In Step 3, λ is updated less than M times. A close inspection of the summation in the numerator reveals that each time it can be done with $O(1)$ complexity, and thus Step 3 is of $O(M)$ complexity. The complexity of Step 1 is more subtle. The value of κ_i can be computed in $O(M)$. Denote the number of loops in Step 1 as r_0; denote the number of layers with κ value lower than or equal to its lower neighbor (*the first kind*) in the r-th loop as b_r, and the number of other effective layers (*the second kind*) as c_r. The complexity in the r-th loop is bounded by linear term of b_r, however $b_r \leq 2a_r$, because only the active layers and their lower effective neighbor layers can be of the first kind. Moreover, notice that $\sum_2^{r_0} a_r \leq M$, since it is upper bounded by the total number of layers made ineffective, further implied by the fact that a layer is active only when an ineffective layer is combined into it. Clearly we have $a_1 = M$, and thus $\sum_1^{r_0} a_r \leq 2M$. The overall complexity is thus $O(M)$, and then the conversion into power allocation is of $O(M)$ complexity.

We note that in order to achieve the $O(M)$ complexity of the given algorithm, fairly involved data structure is needed. More precisely, a doubly-linked list to update effective layers, coupled together with a singly-linked list to update the active layers (which can be combined into one linked-list) appears most appropriate. However, even a naive implementation without such data structure is of $O(M^2)$ complexity.

D. The optimality of the algorithm

Since the problem is a convex optimization, and it obviously satisfies Slater's condition, the KKT conditions are sufficient for optimality. Thus the proof for optimality reduces to to find $\nu_i \geq 0$, that satisfy the complementary slackness condition $\nu_i R_i = 0$, $i = 1, 2, ..., M$, which also satisfy (6), because the power constraint is already satisfied with equality.

Theorem 1: The algorithm given above finds the optimal rate allocation.

Proof: Since for the effective layers, $R_{i_k} > 0$, we may set $\nu_{i_k} = 0$ by the complementary slackness condition. There are several cases that we need to consider:

1) The ineffective layers above the lowest effective layer.
2) The original effective layers which are rendered ineffective by the power constraint, i.e., the layers that become ineffective in step 3.
3) The (original) ineffective layers below the lowest effective layers.

For the first case, suppose these layers are between two effective layers I and J, $I \leq J$. If there are ineffective layers above the highest effective layer, we take $J = M + 1$. From step 3 of the algorithm we can essentially assign the value of $\rho \triangleq \exp(\sum_{j=1}^{I} 2R_j)$ such that

$$-2b[p_I + p_{I+1} + ... + p_{J-2} + p_{J-1}]\rho^{-b} + 2\lambda(n_I - n_J)\rho = 0. \tag{8}$$

Since layer I and J are effective, we set $\nu_I = \nu_J = 0$. Expand the condition in (6), we have

$$-2bp_k\rho^{-b} + 2\lambda(n_k - n_{k+1})\rho - \nu_k + \nu_{k+1} = 0, \quad k = I, I+1, ..., J-1.$$

Though the above equations (under the solution found by the algorithm) uniquely specifies ν_i, $I < i < J$, it is not obvious that those values are indeed non-negative. We need the following lemma to proceed.

Lemma 1: For the combined layers between layer-I and layer-J, given any j^* such that $I \leq j^* \leq J - 2$, we have

$$\kappa^- \triangleq \frac{b\sum_{i=I}^{j^*} p_i}{\sum_{i=I}^{j^*} \Delta n_i} \geq \frac{b\sum_{i=j^*+1}^{J-1} p_i}{\sum_{i=j^*+1}^{J-1} \Delta n_i} \triangleq \kappa^+. \tag{9}$$

The proof of the lemma is in [12], and is omitted here.

Now we are ready to prove the existence of non-negative ν value for the first kind of ineffective layers. The value of ν_i, where $I < i^* < J$ has to satisfy

$$-2b\sum_{i=i^*}^{J-1} p_i\rho^{-b} + 2\lambda\sum_{i=i^*}^{J-1} \Delta n_i\rho - \nu_{i^*} = 0, \tag{10}$$

$$-2b\sum_{i=I}^{i^*-1} p_i\rho^{-b} + 2\lambda\sum_{i=I}^{i^*-1} \Delta n_i\rho + \nu_{i^*} = 0. \tag{11}$$

From (8), we see that ν_{i^*} is indeed non-negative, because Lemma 1 asserts that for i^* we have

$$\kappa^- = \frac{b\sum_{i=I}^{i^*-1} p_i}{\sum_{i=I}^{i^*-1} \Delta n_i} \geq \kappa^+ = \frac{b\sum_{i=i^*}^{J-1} p_i}{\sum_{i=i^*}^{J-1} \Delta n_i}. \tag{12}$$

If $\nu < 0$ was true, then

$$-2b\sum_{i=I}^{J} p_{i^*-1}\rho^{-b} + 2\lambda\sum_{i=I}^{J} p_{i^*-1}\Delta n_i\rho < 0, \tag{13}$$

which would contradict (8).

We next consider the second kind of ineffective layers. Suppose the original effective layers i_k, i_{k+1},...,i_{k+h} becomes ineffective due to the power constraint. Since they are all effective originally, we have by the monotonicity of the κ factor

$$\frac{bp^*_{i_k}}{n_{i_k} - n_{i_{k+1}}} \leq \frac{bp^*_{i_{k+1}}}{n_{i_{k+1}} - n_{i_{k+2}}} \leq ... \leq \frac{bp^*_{i_{k+h}}}{n_{i_{k+h}} - n_{i_{k+h+1}}}, \tag{14}$$

where we used p^* to stand for the accumulated probability after the combining of layers in Step 1 but before Step 3. By step 3 of the algorithm we have

$$\lambda > \frac{bp^*_{i_{k+h}}}{n_{i_{k+h}} - n_{i_{k+h+1}}}. \tag{15}$$

Thus we only need to show the following equations specify a set of non-negative ν_i for $i = i_k, i_{k+1}, ..., i_{k+h}$:

$$\begin{aligned}
-2bp^*_{i_{k+h}} &+ 2\lambda(n_{i_{k+h}} - n_{i_{k+h+1}}) &- \nu_{i_{k+h}} = 0,\\
-2bp^*_{i_{k+h-1}} &+ 2\lambda(n_{i_{k+h-1}} - n_{i_{k+h}}) &- \nu_{i_{k+h-1}} + \nu_{i_{k+h}} = 0,\\
& \quad ... & ...\\
-2bp^*_{i_k} &+ 2\lambda(n_{i_k} - n_{i_{k+1}}) &- \nu_{i_k} + \nu_{i_{k+1}} = 0.
\end{aligned} \tag{16}$$

From the first equation, we get

$$\nu_{i_{k+h}} = -2bp^*_{i_{k+h}} + 2\lambda(n_{i_{k+h}} - n_{i_{k+h+1}}),$$

and it is non-negative because of (15). From the second equation, we have

$$\nu_{i_{k+h-1}} = -2bp^*_{i_{k+h-1}} + 2\lambda(n_{i_{k+h-1}} - n_{i_{k+h}}) + \nu_{i_{k+h}}, \tag{17}$$

which is also non-negative, because $-2bp^*_{i_{k+h-1}} + 2\lambda(n_{i_{k+h-1}} - n_{i_{k+h}}) \geq 0$ due to (15) and (14), and the last term is non-negative from the proceeding argument. Continue this line of argument, it is clear

$$\nu_{i_k} \geq \nu_{i_{k+1}} \geq ... \geq \nu_{i_{k+h}} \geq 0. \tag{18}$$

By this we found the proper non-negative ν_i for the second kind of ineffective layers.

For the third kind of ineffective layers, a proof can be found following the same line as the first kind, and the details are omitted due to space constraint. ■

IV. Variational derivation of the continuous case solution

We now consider the case with continuum of layers, starting with a reformulated optimization problem. Define

$$I(i) = \exp(\sum_{j=1}^{i} 2R_j). \tag{19}$$

We take the number of layers to infinity and the constraint becomes an integral equation, where we convert back to the power gain s instead of noise power n, and it is clear we can replace the inequality by equality without loss of optimality

$$\int_0^\infty I(s)\frac{1}{s^2}ds = P \tag{20}$$

The term to be optimized is given by

$$\int_0^\infty \frac{f(s)}{I(s)^b} ds. \quad (21)$$

Note the additional condition that $I(s)$ has to be monotonically non-decreasing, and the the boundary conditions $I(0) = 1$. Due to the limited space, we next focus on the case when only a single positive power allocation interval exists.

Ignoring the positivity constraint $I'(s) \geq 0$ for now, take

$$J(s, I, I') = \frac{f(s)}{I^b(s)}, \quad G(s, I, I') = \frac{I(s)}{s^2},$$

the problem is thus

$$\text{minimize} \quad \int_0^\infty J(s, I, I') ds, \quad (22)$$

$$\text{subject to} \quad \int_0^\infty G(s, I, I') ds = P. \quad (23)$$

Next we assume there is a unique interval $[s_1, s_2]$ for which power allocation is non-zero. Under this assumption, the objective function reduces to

$$D(I) = \int_{s_1}^{s_2} \frac{f(s)}{I(s)^b} ds + F(s_1) + \frac{1 - F(s_2)}{I(s_2)^b}, \quad (24)$$

where $F(s) = \int_0^s f(r)dr$, and the constraint becomes

$$P(I) = \int_{s_1}^{s_2} I(s) \frac{1}{s^2} ds + \frac{I(s_2)}{s_2} - \frac{1}{s_1} - P. \quad (25)$$

Then we can write the Lagrangian form $L(I) = D(I) + \lambda P(I)$.

To find the extremum of $I(s)$, we consider an increment $q(s)$, and thus the increment of the Lagrangian functional is given by $\Delta(q) = L(I + q) - L(I)$.

The Euler-Lagrange equation (pp. 42-50 [10]) then requires

$$J_I + \lambda G_I - \frac{d}{dh}[J_{I'} + \lambda G_{I'}] = 0, \quad (26)$$

which further simplifies to

$$I(s) = \left(\frac{bf(s)s^2}{\lambda} \right)^{1/(b+1)}. \quad (27)$$

It is clear that for $I'(s) \geq 0$ to be true, which is necessary for $I(s)$ to be a valid solution, $f(s)s^2$ should have non-negative derivative in any interval such that (27) holds; if there is only one interval over the support of $f(s)$ where such property is satisfied, then the single interval solution assumption is indeed true. Now since $q(s_2)$ can be arbitrary, at this variable end (pages 25-29 [10]) a necessary condition for an extremum is

$$\frac{-b(1 - F(s_2))}{I(s_2)^{b+1}} + \lambda \frac{1}{s_2} = 0, \quad (28)$$

which gives

$$\lambda = \frac{bs_2(1 - F(s_2))}{I(s_2)^{b+1}}. \quad (29)$$

With $I(s_1) = 1$, we have $\lambda = bf(s_1)s_1^2$, which now with the expression of $I(s)$ gives one boundary condition

$$1 - F(s_2) = f(s_2)s_2. \quad (30)$$

The lower bound s_1 is determined by the power constraint, from which we have

$$\int_{s_1}^\infty \frac{I(s)}{s^2} ds = P + \frac{1}{s_1}. \quad (31)$$

We have thus find the unique extremal solution.

To find the corresponding power allocation, define $T(s) = \int_s^\infty P(r)dr$. We thus derive from (5) that

$$\begin{aligned} T(s) &= \left(\frac{f(s_2)s_2^2}{f(s)s^2} \right)^{1/(b+1)} \frac{1}{s_2} \\ &+ \int_s^{s_2} \left(\frac{f(r)r^2}{f(s)s^2} \right)^{1/(b+1)} \frac{1}{r^2} dr - \frac{1}{s}. \quad (32) \end{aligned}$$

This is in fact the same solution as that in [7], and indeed the derivation in [7] matches the optimal solution derived through the classical variational method. Furthermore, the variational method derivation directly asserts that $f(s)s^2$ has non-negative derivative is necessary for any positive power allocation interval; this condition was however lacking in [7]. In the more general case without the assumption of only one effective interval, a solution can also be derived from the variational method through a more involved route (given in [12]), however it is not clear the method in [7] can be generalized to this case.

V. Conclusion

We considered the optimal power/rate allocation in the broadcast strategy, in order to minimize the expected distortion of a quadratic Gaussian source transmitted over a fading channel. A linear complexity algorithm is proposed, and its correctness and optimality are proved. Moreover, a derivation for the optimal allocation with a continuum of layers is given, using the classical variational method.

Acknowledgement

This research work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++.

References

[1] S. Shamai and A. Steiner, "A broadcast approach for single-user slowly fading MIMO channel," *IEEE Trans. Information Theory*, vol. 49, No. 10, pp. 2617–2635, Oct. 2003.

[2] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 1991.

[3] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Information Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.

[4] S. Sesia, G. Caire, and G. Vivier, "Lossy transmission over slow-fading awgn channels: a comparison of progressive, superposition and hybrid approaches," in *Proc. IEEE International Symposium on Information Theory*, Adelaide, Australia, Sep. 2005.

[5] F. Etemadi and H. Jafarkhani, "Optimal layered transmission over quasi-static fading channels," in *Proc. IEEE Symposium on Information Theory*, Seattle, WA, Jul., 2006.

[6] C. Ng, D. Gündüz, A. Goldsmith and E. Erkip, "Recursive power allocation in Gaussian layered broadcast coding with successive refinement," To appear, *IEEE International Conference on Communications (ICC)*, Glasgow, Scotland, Jun., 2007.

[7] C. Ng, D. Gündüz, A. Goldsmith and E. Erkip, "Minimum Expected Distortion in Gaussian Layered Broadcast Coding with Successive Refinement," To appear, *IEEE International Symposium on Information Theory*, Nice, France, Jun., 2007.

[8] G. Caire and K. Narayanan, "On the SNR exponent of hybrid digital-analog space time coding," in *Proc. Allerton Conf. Commun., Contr. and Comp.*, 2005.

[9] D. Gündüz and E. Erkip, "Source and channel coding for quasi-static fading channels," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov., 2005.

[10] I. M. Gelfand and S. V. Fomin, *Calculus of variations*, Englewood Cliffs, NJ: Prentice-Hall, 1963.

[11] D. N. C. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," U.C. Berkeley technical report UCB/ERL M99/7. http://www.eecs.berkeley.edu/Pubs/TechRpts/1999/3578.html,

[12] C. Tian, A. Steiner, S. Shamai and S. Diggavi, "Expected Distortion for Gaussian Source with a Broadcast Transmission Strategy over Fading Channels," In preparation.