On opportunistic codes and broadcast codes with degraded message sets

Suhas N. Diggavi
School of Computer and
Communication Sciences, EPFL,
Lausanne, Switzerland.
Email: suhas.diggavi@epfl.ch

David N C. Tse
Department of EECS,
UC Berkeley,
Berkeley, California, USA.
Email: dtse@eecs.berkeley.edu

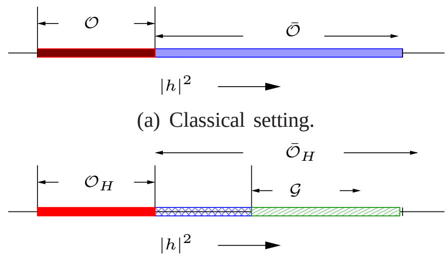
***Abstract*— Diversity embedded codes are opportunistic codes which take advantage of good channel realizations while ensuring at least part of the information is received reliably for bad channels. We establish a connection between these codes and degraded message set broadcast codes. We characterize the achievable rate region for the parallel Gaussian degraded message set broadcast problem, when only the strongest user needs the private information. Using this, we partially characterize the set of achievable rate-diversity tuples for the diversity embedded problem for parallel fading channels.**

I. INTRODUCTION

Randomness, inherent in wireless communication, causes uncertainty at the transmitter about the communication channel. Therefore the classical approach is to ensure a desired level of reliability against adverse channels while maximizing the transmission rate. This tradeoff between rate and reliability is fundamental [8].

However, the classical approach, while ensuring reliability against deep fades, does not take advantage of good channel realizations. An alternate strategy was proposed in [2], where a high-reliability code was embedded within a higher rate code. This allowed opportunistic communication where the high rate code opportunistically takes advantage of good channel realizations whereas the embedded high-diversity code ensures that at least part of the information is received reliably.

A natural framework to address the question of performance limits of diversity-embedded codes is using the outage formulation. First recall the outage formulation for the classical single target rate setting. Given the space of channel realizations, we divide it into outage set $\mathcal{O}$ and a non-outage set $\bar{\mathcal{O}}$. We require a code which can be decoded reliably (with arbitrarily small error probability) for all channels in the set $\bar{\mathcal{O}}$. Therefore the "worst" channel in the non-outage set $\bar{\mathcal{O}}$ limits the rate of transmission. For a given outage probability p, the performance limit is the outage capacity C_p, which is the maximum achievable such worst-case rate, maximum over all non-outage set with probability of at least $1-p$. Inverting the mapping from p to C_p, we get the outage probability $p(R)$ for a given target rate R. In terms of outages, diversity embedded codes (or opportunistic codes) can be viewed as sending two messages m_H (higher priority) and m_L (lower priority) such that m_H is recovered reliably for all channels in the non-outage set $\bar{\mathcal{O}}_H$, but in addition m_L is to be recovered if we get a better channel in $\mathcal{G} \subset \bar{\mathcal{O}}_H$. See Figures 1(a) and 1(b). The streams m_H and m_L have reliability $p_H = \mathbb{P}\{\mathcal{O}_H\}$ and $p_L = \mathbb{P}\{\bar{\mathcal{G}}\}$ respectively; the low-priority stream m_L has lower reliability but it represents additional information that can be pumped through the channel when it is good. The performance limit is then characterized by the set of all achievable 4-tuples (R_H, p_H, R_L, p_L), where R_H and R_L are the rates supportable by the worst channels in $\bar{\mathcal{O}}_H$ and $\mathcal{G}$ respectively.



(a) Classical setting.

(b) Diversity embedded setting.

Fig. 1. Outage events in the classical setting and for diversity-embedded coding.

We can connect this idea to broadcast codes with degraded message sets [5]. We can think of the set of users as identified by the different channel realizations in $\bar{\mathcal{O}}_H$. We identify m_H as the common message to be delivered to all the users and m_L as a private message to be delivered to a subset of the users $\mathcal{G} \subset \bar{\mathcal{O}}_H$. This point of view was in fact proposed by Cover [1] in his original broadcast paper. Shamai [6] applied this approach to the SISO fading channel but his performance objective was the transmission rate averaged over the channels rather than the outage probabilities. In contrast, we are not interested in maximizing the average rate, but characterizing the achievable rate-diversity tuple for diversity embedded codes. Moreover, [6] focuses on the performance of one specific coding strategy (Gaussian superposition) while we will also characterize the optimal performance achievable by *any* scheme.

In [3], we focus on the cases of single transmit antenna (SIMO) and single receive antenna (MISO) under Rayleigh fading and characterize the high SNR optimal diversity-multiplexing tradeoff of diversity-embedded codes.

We showed a *successive refinable* property: the high-priority stream can attain the optimal diversity performance as though the low-priority stream were not there and yet the diversity performance of the low priority stream is the same as that optimal diversity of a stream with the aggregate rate of the two streams. The commonality of these cases is that there is only a single degree of freedom. In this paper, we consider the simplest scenario with multiple degrees of freedom: parallel fading channels. We provide a characterization for the rate-diversity tradeoff when it is required that the higher priority stream gets the same performance as if the resources were exclusively devoted to it. We compute the optimal diversity multiplexing tradeoff of the opportunistic lower priority stream. It is shown that though the performance is not as optimistic as in the single degree of freedom case, there is still some opportunism that can be utilized. The connection between diversity embedding and broadcast codes allows us to establish the desired bounds on the performance of diversity embedded codes through characterizing the rate regions of certain parallel broadcast channels with degraded message sets.

The paper is organized as follows. In Section II, we formally state the problem, establish notation and review some previous results. Section III is devoted to some characterizations for the degraded message set problem, both for the parallel discrete memoryless channel as well as the parallel Gaussian channel. In Section IV we use the connection with the degraded message set problem to provide a characterization for the rate-diversity tuple, when the higher priority stream gets maximal performance. We conclude in Section V with a brief discussion. Some of the proof details are given in the appendices.

II. Problem Statement and preliminaries

Our focus in this paper is on the quasi-static flat-fading channel where we transmit information coded over M_t transmit antennas and have M_r antennas at the receiver:

$$\mathbf{Y}(i) = \mathbf{H}\mathbf{X}(i) + \mathbf{Z}(i), \tag{1}$$

where $\{\mathbf{X}(i)\}$ is the input sequence with transmit power constraint P, and $\{\mathbf{Z}(i)\}$ is assumed to be additive white Gaussian noise with variance σ^2. Furthermore, the assumption is that the transmitter has no channel state information, whereas the receiver is able to perfectly track the channel (a common assumption, see for example [8]). More specifically we focus on the parallel channel where the channel matrix $\mathbf{H}$ is diagonal with i.i.d. Rayleigh distributed gains and we have $M_t = M_r = K$.

In the classical single-stream scenario, [8] formulated the diversity-multiplexing tradeoff $d^*(r)$ as the high SNR version of the outage probability curve $p(R)$, scaled as follows:

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log p(r\log\mathsf{SNR}))}{\log\mathsf{SNR}} = -d^*(r).$$

We use the notation:

$$p(r\log\mathsf{SNR}) \doteq \mathsf{SNR}^{-d^*(r)}$$

to denote this. We similarly define $\dot{\geq}, \dot{\leq}$ as well.

Now consider the scenario when we have two information streams with different reliability levels. We can define (r_H, d_L, r_L, d_L) as an achievable rate-diversity tuple if there exists a sequence $\{(R_H(\mathsf{SNR}), p_H(\mathsf{SNR}), r_L(\mathsf{SNR}), p_L(\mathsf{SNR}))\}$ such that $(R_H(\mathsf{SNR}), p_H(\mathsf{SNR}), r_L(\mathsf{SNR}), p_L(\mathsf{SNR}))$ is an achievable outage performance at each finite SNR and

$$R_H = r_H \log\mathsf{SNR}, \quad p_H \doteq \mathsf{SNR}^{-d_H},$$

$$R_L = r_L \log\mathsf{SNR}, \quad p_L \doteq \mathsf{SNR}^{-d_L},$$

where $p_H = \mathbb{P}\{\mathcal{O}_H\}$ and $p_L = \mathbb{P}\{\bar{\mathcal{G}}\}$.

If viewed as a single layer code, the diversity embedded code achieves rate-diversity pairs (r_H, d_H) and $(r_H + r_L, d_L))$, where we have assumed that $d_H > d_L$, since naturally we want higher reliability for the higher priority stream. Since we cannot beat the single-layer rate-diversity trade-off, we see that necessarily $d_H \leq d^*(r_H)$ and $d_L \leq d^*(r_H + r_L)$ where $(r, d^*(r))$ is the optimal rate-diversity point predicted by Theorem 2 in [8]. We established the following result [3].

Theorem 2.1: When $\min(M_t, M_r) = 1$ and the entries of $\mathbf{H}$ are i.i.d. Rayleigh, then the diversity-multiplexing tradeoff curve is successively refinable, *i.e.*, for any multiplexing gains r_H and r_L such that $r_H + r_L \leq 1$, the diversity orders

$$d_H = d^*(r_H), \; d_L = d^*(r_H + r_L)$$

are achievable, where $d^*(r)$ is the optimal diversity order given in Theorem 2 in [8]. Hence, the best possible performance can be achieved.

A natural question is whether this property carries over when we have more than one degree of freedom. We investigate this question for parallel channels, which form the simplest example of channels with multiple degrees of freedom.

III. Degraded message sets over parallel channels

In Section III-A, we take a brief detour to examine the degraded message set broadcast channel (DMSBC) over discrete memoryless channels (DMC). We do this since the result for DMCs can be used to prove the result for the parallel Gaussian channel as done in Section III-B.

A. Some characterizations for DMCs

All the channels considered in this subsection are DMCs. We first start with a finite set of users $Y_1, \ldots, Y_N$ and a set of broadcast channels defined by $(\mathcal{X}, p(y_1, \ldots, y_N|x), \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_N)$, where the input $X \in \mathcal{X}$ and the outputs $Y_n \in \mathcal{Y}_n, n = 1, \ldots, N$ come from a finite alphabet. We assume that the common message is required by all the users, but the private message is required only for the subset of users $\{Y_n : n \in \mathcal{Q}_p\}$. We denote by R_c, R_p as the rate per transmission for the common message and the private message respectively. The goal is to characterize all the achievable pairs $(R_c, R_p) \in \mathcal{C}_{dms}$.

Since all the users require to decode the common message and a subset of the users require to decode the private message, a superposition strategy is natural. Such a strategy achieves the following rate region $\mathcal{R}_s$ defined using auxiliary random variable U and a distribution, $p(u,x,y_1,\ldots,y) = p(u)p(x|u)p(y_1,\ldots,y_N|x)$ as:

$$\begin{aligned}\mathcal{R}_s = \text{conv}\,\Cup_U \{(R_c,R_p) : R_c &\leq I(U;Y_n),\ n=1,\ldots,N\\ R_p &\leq I(X;Y_n|U), n\in\mathcal{Q}_p.\}\end{aligned} \tag{2}$$

When the number of users $N=2$, Korner and Marton [5] have shown that this provided a complete characterization of the rate region. It is quite intriguing that the above region can be proved to be a complete characterization for just some sporadic examples when $N>2$.

We denote a Markov chain as $U \leftrightarrow X \leftrightarrow Y$ when $p(u,x,y) = p(u)p(x|u)p(y|x)$. We give a characterization for some special cases of the DMSBC when the set of channels have a Markov chain structure.

Theorem 3.1: If $X \leftrightarrow Y_1 \leftrightarrow \{Y_2,\ldots,Y_N\}$ and message W_c is to be delivered to all users, but $\mathcal{Q}_p = \{1\}$, then the capacity region $\mathcal{C}_{dms} = \mathcal{R}_s$.

Theorem 3.2: If $X \leftrightarrow \{Y_1,\ldots,Y_{N-1}\} \leftrightarrow Y_N$ and message W_c is to be delivered to all users, but $\mathcal{Q}_p = \{1,\ldots,N-1\}$, then the capacity region $\mathcal{C}_{dms} = \mathcal{R}_s$.

Consider a parallel DMC broadcast channel with N users $\mathbf{Y}_1,\ldots,\mathbf{Y}_N$ such that,

$$p(\mathbf{Y}_1,\ldots,\mathbf{Y}_N) = \prod_{k=1}^{K}\{p(y_{1,k}|x_k)p(y_{2,k},\ldots,y_{N,k}|y_{1,k})\}, \tag{3}$$

i.e., we have a set of independent parallel channels such that $X_k \leftrightarrow Y_{1,k} \leftrightarrow \{Y_{2,k},\ldots,Y_{N,k}\},\ k=1,\ldots,K$.

Theorem 3.3: For the parallel discrete memoryless broadcast channel defined in (3), and given arbitrary random variables $\mathbf{U} = \{U_1,\ldots,U_K\}$, and a distribution

$$p(\mathbf{u},\mathbf{X},\mathbf{Y}_1,\ldots,\mathbf{Y}_N) = \left\{\prod_{k=1}^{K} p(u_k)p(x_k|u_k)\right\} p(\mathbf{Y}_1,\ldots,\mathbf{Y}_N|\mathbf{X})$$

the capacity region for the DMSBC for $\mathcal{Q}_p = \{1\}$ is given by

$$\begin{aligned}\mathcal{C}_d = \text{conv}\bigcup_{\mathbf{U}}\{(R_c,R_p) : R_c &\leq \sum_{k=1}^{K} I(U_k;Y_{n,k}),\ n=1,\ldots,N\\ R_p &\leq \sum_{k=1}^{K} I(X_k;Y_{1,k}|U_k)\}.\end{aligned} \tag{4}$$

In [4], El-Gamal characterizes the capacity region of the parallel broadcast channel with degraded sub-channels for $N=2$ users and $K=2$ sub-channels. The above result holds for arbitrary number of users and sub-channels but with a specific degradedness structure.

B. A characterization for parallel Gaussian channels

We define DMSBC with parallel Gaussian channels for a set of N users in the following way. For user $n \in \{1,\ldots,N\}$, we define a channel $\mathbf{h}_n = \{h_{n,1},\ldots,h_{n,K}\}$ as,

$$Y_{n,k} = h_{n,k}X_k + Z_{n,k},\ k=1,\ldots,K, \tag{5}$$

where $Z_{n,k} \sim \mathbb{C}\mathcal{N}(0,\sigma^2)$ and independent across n,k and we have an input power constraint $\sum_{k=1}^{K}\mathbb{E}[X_k|^2 \leq P$. We also define the special structure for the channel gains such that

$$|h_{1,k}|^2 \geq |h_{n,k}|^2,\ k=1,\ldots,K,\ \forall n, \tag{6}$$

i.e., user $\mathbf{Y}_1$ has the strongest channel in *each* of the sub-channels. Therefore, this is stochastically equivalent to channels which have the Markovian structure needed in Theorem 3.3. Using Theorem 3.3, we can apply the scalar entropy-power inequality in order to be able to write the capacity region for the parallel Gaussian problem defined in (5). The proof outline of the following result is given in the Appendix III.

Theorem 3.4: For the parallel Gaussian defined in (5)-(6), the capacity region is given by,

$$\mathcal{C}(P) = \bigcup_{\mathbf{P}_H,\mathbf{P}_L\geq 0:\sum_{k=1}^{K}(P_{H,k}+P_{L,k})\leq P} \mathcal{C}_p(\mathbf{P}_H,\mathbf{P}_L) \tag{7}$$

where $\mathbf{P}_H = \{P_{H,1},\ldots,P_{H,K}\}, \mathbf{P}_L = \{P_{L,1},\ldots,P_{L,K}\}$ and

$$\begin{aligned}\mathcal{C}_p(\mathbf{P}_H,\mathbf{P}_L) &= \{(R_c,R_p) : \text{for } n=1,\ldots,N\\ R_H &\leq \sum_{k=1}^{K}\log\left(1+\frac{P_{H,k}|h_{n,k}|^2}{\sigma^2+P_{L,k}|h_{n,k}|^2}\right),\\ R_L &\leq \sum_{k=1}^{K}\log\left(1+\frac{P_{L,k}|h_{1,k}|^2}{\sigma^2}\right)\}\end{aligned} \tag{8}$$

Therefore, this shows that in the parallel Gaussian degraded message set broadcast channel, when the private message is required *only* by the user with the largest channel gain in *each* sub-channel, then a Gaussian superposition coding strategy with successive decoding is optimal.

IV. CHARACTERIZATION FOR DIVERSITY EMBEDDING

In this section we use the connection between the diversity embedding codes and the degraded message set broadcast codes to characterize a particular region of the diversity-rate tuple for the embedding problem. As mentioned in Section II, we focus on the parallel fading channel model where the kth sub-channel faces a Rayleigh distributed fading gain of h_k and the fading gains are assumed to be independent across the K sub-channels. For this case, the diversity-multiplexing tradeoff achieved in a single stream setting is $d^*(r) = K-r$ [8]. Our central question is the performance in a diversity-embedding setting when we require that $d_H = d^*(r_H)$. The main result of this paper is summarized in the following theorem.

Theorem 4.1: For the parallel fading channel defined in (1), given r_H, r_L, if $d_H = d^*(r_H)$, then the optimal achievable d_L is given by

$$d_L^* = \{K-(r_H+r_L)-(K-1)r_H\}^+,$$

where $\{\cdot\}^+ = \max\{0,\cdot\}$.

Note that a single stream at multiplexing gain r_H+r_L would enjoy a diversity gain of $K-(r_H+r_L)$. Thus there is a loss of diversity of $(K-1)r_H$ due to having a high-priority layer embedded.

For achievability of Theorem 4.1 we design a Gaussian superposition strategy, with successive decoding. In particular we use i.i.d. Gaussian codes for each stream, allocating a power of P_H, P_L respectively to the high and low-priority streams respectively. Moreover, we split the power equally across the sub-channels for both streams. Furthermore we design the power allocation to the streams such that $\frac{P_H}{P_L} = (P/\sigma^2)^\beta = \mathsf{SNR}^\beta,\ \beta \in [0,1]$. Then, at high SNR, we have $\mathsf{SNR_H} \doteq \mathsf{SNR},\ \mathsf{SNR_L} \doteq \mathsf{SNR}^{1-\beta}$. Therefore, for given target rates R_H and R_L for the two streams, and a successive decoding strategy, the outage probabilities are given by:

$$\begin{aligned}\mathcal{O}_H(\mathsf{SNR}) &= \left\{\mathbf{h} : \sum_{k=1}^{K} \log\left(1 + \frac{|h_k|^2\mathsf{SNR_H}}{1+|h_k|^2\mathsf{SNR_L}}\right) < R_H\right\}\\ \tilde{\mathcal{O}}_L(\mathsf{SNR}) &= \left\{\mathbf{h} : \sum_{k=1}^{K} \log\left(1 + |h_k|^2\mathsf{SNR_L}\right) < R_L\right\}, \quad (9)\\ p_H(\mathsf{SNR}) &= \mathbb{P}\left\{\mathcal{O}_H(\mathsf{SNR})\right\}, p_L(\mathsf{SNR}) = \mathbb{P}\left\{\mathcal{O}_L(\mathsf{SNR})\right\},\end{aligned}$$

where $\mathcal{O}_L(\mathsf{SNR}) = \mathcal{O}_H(\mathsf{SNR}) \bigcup \tilde{\mathcal{O}}_L(\mathsf{SNR})$, $\mathsf{SNR_H} = \frac{P_H}{K\sigma^2}$, $\mathsf{SNR_L} = \frac{P_L}{K\sigma^2}$.

For a given value of β, the diversity exponents can be calculated based on appealing to Laplace's principle [7]. We can reparameterize $|h_k|^2 := \mathsf{SNR}^{-\alpha_k}$ for all k, and substituting into the outage condition for the high-priority layer in (9) yields the asymptotic outage condition for m_H on $\alpha := (\alpha_1, \ldots, \alpha_K)$:

$$\tilde{\mathcal{A}}_H(\beta) = \{\alpha : \sum_{k=1}^{K} \min(\beta, 1-\alpha_k) < r_H\} \qquad (10)$$

Since the h_k's are i.i.d. Rayleigh, the outage probability is asymptotically $\mathsf{SNR}^{-\min_{\alpha_k \in \tilde{\mathcal{A}}_H(\beta)} \sum_{k=1}^{K} \alpha_k}$. A direct calculation shows that the desired diversity exponent of $K - r_H$ is achievable if and only if $\beta > r_H$, in which case the set of exponents of the typical channels that cause outage is:

$$\mathcal{A}_H := \{\alpha : \sum_k \alpha_k = K - r_H, \alpha_k \in [0,1], \forall k\}, \qquad (11)$$

all of which have the same asymptotic probability $\mathsf{SNR}^{-(K-r_H)}$. See Figure 2.

Similarly, the diversity exponent of the low-priority layer can be calculated to be $[(1-\beta)K - r_L]^+$ with the set of the exponents of the typical channels that cause outage for m_L being:

$$\mathcal{A}_L(\beta) := \{\alpha : \sum_k \alpha_k = [(1-\beta)K - r_L]^+, \alpha_k \in [0,1], \forall k\} \qquad (12)$$

Thus, by choosing β arbitrarily close to but larger than r_H, the diversity of the low-priority layer can be arbitrarily close to: $[(1-r_H)K - r_L]^+$, which is as claimed in Theorem 4.1. Moreover, it can be shown from symmetrical considerations that splitting the power unequally across the sub-channel cannot yield better diversity performance.

Next, we will provide a proof outline for the converse part of this theorem. Suppose for the sake of contradiction the existence of a small $\epsilon > 0$ and a coding scheme that yields a

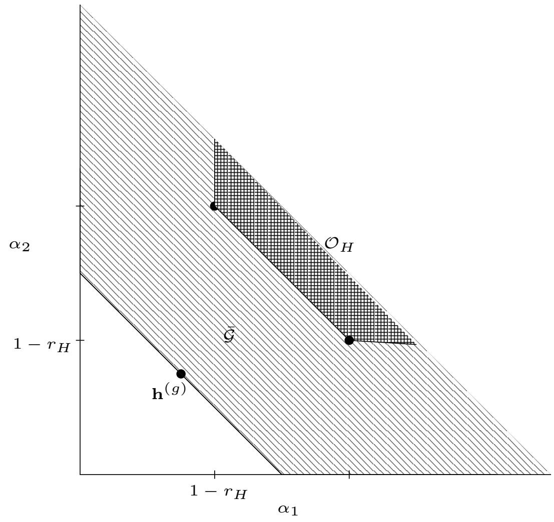


Fig. 2. Outage events $\mathcal{O}_H, \bar{\mathcal{G}}$ for messages m_H, m_L respectively in parallel fading channel for $K = 2$ and $|h_k|^2 = \mathsf{SNR}^{-\alpha_k}, k = 1,2$. The 45 degree line lower bounding $\mathcal{O}_H$ is $\mathcal{A}_H$.

diversity of $d_H = K - r_H$ for the high-priority layer and a diversity of $d_L = [(1 - r_H + \epsilon)K - r_L]^+$ for the low-priority layer, better than claimed. Then it must be that we can deliver the message m_H to all the typical channels in

$$\mathcal{A}_H := \{\alpha : \sum_k \alpha_k = K - r_H, \alpha_k \in [0,1], \forall k\}$$

and deliver m_L to all channels in

$$\mathcal{A}_L(r_H - \epsilon) = \{\alpha : \sum_k \alpha_k = [(1 - r_H + \epsilon)K - r_L]^+, \alpha_k \in [0,1], \forall k\}$$

In particular it can deliver the message m_L to the symmetrical channel $\mathbf{h}^{(g)} \in \mathcal{A}_L$ with $|h_k^{(g)}|^2 = \mathsf{SNR}^{-(1-r_H-r_L/K+\epsilon)}, \forall k$. Note that channel $\mathbf{h}^{(g)}$ is stronger than all the channels in $\mathcal{A}_H$, as can be seen in Figure 2. This is a degraded message set problem with the user requiring the private message stronger than all other users in all sub-channels. Hence, we can apply Theorem 3.4, which shows that Gaussian superposition with appropriate power allocation is optimal. But this is precisely our achievability strategy above and we have already seen that this strategy cannot simultaneously deliver the message m_H to all users in $\mathcal{A}_H$ and to the user with channel $\mathbf{h}^{(g)}$ while delivering message m_L to $\mathbf{h}^{(g)}$. Hence we have a contradiction. Also, since the diversity order is non-negative we can see that $d_L \leq \{K - (r_H + r_L) - (K-1)r_H\}^+$.

Even though one may not achieve successive refinability on the parallel channel, the diversity-embedded coding system should still be considered as strictly better than the single-stream system. Consider the rate-diversity tuple $(r_H, K - r_H, r_L, K(1 - r_H) - r_L)$, achievable by the strategy above for $r_H < 1, Kr_H + r_L < K$. A single-stream system with multiplexing gain r_H achieves a diversity of $K - r_H$, even with an optimal code. A diversity-embedded system can provide that performance (in the high-priority stream) but can in addition pump extra bits through when the channel is good. This means that there is slack in the original single-stream system and the diversity-embedded system intelligently exploits that slack.

V. Discussion

By establishing a connection between diversity embedding and broadcast channels with degraded message sets, we were able to partially characterize the achievable rate tuples (r_H, d_H, r_L, d_L) for the parallel fading channel, for the case when the high-priority stream enjoys a diversity as though the entire resource of the channel were devoted to it. The main difficulty in the general characterization appears to be in solving the general DMSBC problem for more than 2 users.

Appendix I
Proof of Theorems 3.1 & 3.2

First we prove the statement of Theorem 3.1, and the proof of Theorem 3.2 follows quite similarly. If the decoders decode message to be $\hat{W}_c(n), n = 1,\ldots,N$ and $\hat{W}_p$, then we define $\mathcal{P}_e^{(m)} = \max\{\mathbb{P}\left\{\hat{W}_c(n) \neq W_c\right\}, n = 1,\ldots,N, \mathbb{P}\left\{\hat{W}_p \neq W_p\right\}\}$. Using Fano's inequality, we can write

$$mR_c \overset{(a)}{\leq} \sum_{i=1}^{m} \left\{ I(W_c, Y_1^{i-1}; Y_n(i)) \right\} + mR_c\mathcal{P}_e^{(m)} + 1$$

where (a) follows for $n \geq 2$ due to the Markov chain relationship $Y_n^{i-1} \leftrightarrow Y_1^{i-1} \leftrightarrow X(i) \leftrightarrow Y_n(i)$ and trivially for $n = 1$. Similarly, we have

$$mR_p \overset{(b)}{\leq} \sum_{i=1}^{m} \left\{ I(X(i); Y_1(i)|Y_1^{i-1}, W_c) \right\} + mR_c\mathcal{P}_e^{(m)} + 1 \quad (13)$$

where (b) follows due to the independence of W_c and W_p. By defining $U(i) = \{Y_1^{i-1}, W_c\}$ and by the usual time-sharing arguments the proof is complete. Now, for Theorem 3.2, the steps are quite similar. Notice that in (13), we also have $I(W_p; Y_1(i)|Y_1^{i-1}, W_c) \leq I(W_p; Y_1(i)|Y_N^{i-1}, W_c)$ due to the Markov chain relationship. Therefore, using $U(i) = \{Y_N^{i-1}, W_c\}$ instead, we will be able to prove Theorem 3.2.

Appendix II
Proof of Theorem 3.3

For simplicity of notation, we do the proof for $K = 2$, but the general case follows identically. For $n \geq 2$,

$$mR_c \overset{(a)}{\leq} \sum_{i=1}^{m} \left\{ I(W_c, Y_{1,1}^{i-1}; Y_{n,1}(i)) + I(W_c, Y_{1,1}^{m}, Y_{1,2}^{i-1}; Y_{n,2}(i)) \right\}$$

The relationship (a) follows because of Fano's inequality and $Y_{n,1}^{i-1} \leftrightarrow Y_{1,1}^{i-1} \leftrightarrow X_1(i) \leftrightarrow Y_{n,1}(i)$ and $(Y_{n,1}^m, Y_{n,2}^{i-1}) \leftrightarrow (Y_{1,1}^m, Y_{1,2}^{i-1}) \leftrightarrow X_1(i) \leftrightarrow Y_{n,2}(i)$. For brevity we have left out the additive term $mR_c\mathcal{P}_e^{(m)} + 1$ Now we also have,

$$mR_p \overset{(b)}{\leq} \sum_{i=1}^{m} \left\{ I(X_1(i); Y_{1,1}(i)|Y_{1,1}^{i-1}, W_c) + I(X_2(i); Y_{1,2}(i)|Y_{1,1}^{m}, Y_{1,2}^{i-1}, W_c) \right\}$$

where (b) follows due to Fano's inequality and due to the independence of W_c and W_p. Again for brevity we leave out the additive term $mR_p\mathcal{P}_e^{(m)} + 1$. By defining $U_1(i) = \{W_c, Y_{1,1}^{i-1}\}, U_2(i) = \{W_c, Y_{1,1}^m, Y_{1,2}^{i-1}\}$ and using standard time-sharing arguments, the single-letter form of the theorem is proved. For arbitrary K we would use the same argument with $U_k(i) = (W_c, Y_{1,k}^{i-1}, \{Y_{1,s}^m\}_{s=1}^{k-1}),\ k = 1,\ldots,K$.

Appendix III
Proof of Theorem 3.4

We will prove the result for $\mathcal{C}_p(\mathbf{P}_H, \mathbf{P}_L)$, from which the Theorem follows. We start with the single-letter characterization of Theorem 3.3 for the parallel discrete problem. Basically, we apply the scalar entropy power inequality for each of the sub-channels to arrive at the desired result. Given the set of N channels $\mathbf{h} \overset{def}{=} \{\mathbf{h}_n = (h_{n,1},\ldots,h_{n,K}), n = 1,\ldots,N\}$, we define the error probability $\mathcal{P}_{e,H}^{(m)}(\mathbf{h})$ and $\mathcal{P}_{e,L}^{(m)}(\mathbf{h})$ as the maximum mean errors for messages m_H and m_L respectively when codes of length m are used. Given a power allocation $\mathbf{P} = \{P_1, P_2\}$ per sub-channel, from (4) we see that

$$R_L \leq I(X_1; Y_{1,1}|U_1) + I(X_2; Y_{1,2}|U_2) + \underbrace{R_L\mathcal{P}_{e,L}^{(m)}(\mathbf{h}) + \frac{1}{m}}_{\epsilon_L} \quad (14)$$

$$\overset{(a)}{=} \log\left(\frac{\alpha_{L,1}P_1|h_{1,1}|^2}{\sigma^2} + 1\right) + \log\left(\frac{\alpha_{L,2}P_2|h_{1,2}|^2}{\sigma^2} + 1\right) + \epsilon_L$$

where (a) follows because $h(Z_{1,k}) \leq h(Y_{1,k}|U_1) \leq h(Y_{1,k})$ and hence there exists such $\alpha_{L,k} \in [0,1]$ for $k = 1,2$. Now, using the first constraint on the common rate, we see that,

$$R_H \leq I(U_1; Y_{n,1}) + I(U_2; Y_{n,2}) + \underbrace{R_H\mathcal{P}_{e,H}^{(m)}(\mathbf{h}) + \frac{1}{m}}_{\epsilon_H} \quad (15)$$

$$\overset{(b)}{\leq} \log\left(\frac{\frac{P_1|h_{n,1}|^2}{\sigma^2} + 1}{\frac{\alpha_{L,1}P_1|h_{n,1}|^2}{\sigma^2} + 1}\right) + \log\left(\frac{\frac{P_2|h_{n,2}|^2}{\sigma^2} + 1}{\frac{\alpha_{L,2}P_2|h_{n,2}|^2}{\sigma^2} + 1}\right) + \epsilon_H$$

where (b) follows from the scalar entropy power inequality by using $Y_{n,k} = Y_{1,k} + \tilde{Z}_{n,k},\ k = 1,2$, where $\tilde{Z}_{n,k} \sim \mathbb{C}\mathcal{N}(0, \frac{\sigma^2}{|h_{n,k}|^2} - \frac{\sigma^2}{|h_{1,k}|^2})$. By defining $P_{L,k} = \alpha_{L,k}P_k$ and $P_{H,k} = (1 - \alpha_{L,k})P_k$, we get the desired result. In the same manner we can prove this for general K and general $n \in \{1,\ldots,N\}$.

Acknowledgements: The research of D. Tse is supported by the National Science Foundation through grant CCR-01-18784.

References

[1] T M. Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18:2–14, January 1972.

[2] S. N. Diggavi, N. Al-Dhahir, and A. R. Calderbank, *Diversity embedding in multiple antenna communications*, in DIMACS workshop on Network Information Theory, March 2003. Also part of AMS edited volume "Network Information Theory", edited by P. Gupta, G. Kramer and A J. van Wijngaarden, AMS volume 66, Series on Discrete Mathematics and Theoretical Computer Science, pp 285–302, 2004.

[3] S. N Diggavi and D. Tse, *Fundamental Limits of Diversity-Embedded Codes over Fading Channels*, IEEE ISIT, Adelaide, Australia, pp 510–514, September, 2005.

[4] A A. El-Gamal. Capacity of the product and sum of two unmatched broadcast channels. *Problemy Peridachi Informatsii*, 17(1):3–23, January-March 1980 (English translation *Problems in Information Transmission*, 16(1):1–16, January-March 1980).

[5] J. Korner and K. Marton. General broadcast channels with degraded message sets. *IEEE Trans. IT*, 23(1):60–64, January 1977.

[6] S Shamai (Shitz). A broadcast strategy for the Gaussian slowly fading channel. In *IEEE ISIT*, page 150, June 1997.

[7] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[8] L Zheng and D N C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. *IEEE Trans. IT*, 49(5):1073–1096, May 2003.