On successive refinement of diversity for fading ISI channels

S. Dusad and S. N. Diggavi

***Abstract*— Rate and diversity impose a fundamental tradeoff in communications. This tradeoff was investigated for Intersymbol Interference (ISI) channels in [4]. A different point of view was explored in [1] where high-rate codes were designed so that they have a high-diversity code embedded within them. Such diversity embedded codes were investigated for flat fading channels and in this paper we explore its application to ISI channels. In particular, we investigate the rate tuples achievable for diversity embedded codes for scalar ISI channels through particular coding strategies. The main result of this paper is that the diversity multiplexing tradeoff for fading ISI channels is indeed successively refinable. This implies that for fading single input single output (SISO) ISI channels one can embed a high diversity code within a high rate code without any performance loss (asymptotically). This is related to a deterministic structural observation about the asymptotic behavior of frequency response of channel with respect to fading strength of time domain taps.**

I. Introduction

There exists a fundamental tradeoff between diversity (error probability) and multiplexing (rate). This tradeoff was characterized in the high SNR regime for flat fading channels with multiple transmit and multiple receive antennas (MIMO) [6]. This characterization was done in terms of multiplexing rate which captured the rate-growth (with SNR) and diversity order which represented reliability (at high SNR). This diversity multiplexing (D-M) tradeoff has been extended to several cases including fading ISI channels [4], [5]. The presence of ISI gives significant improvement of the diversity order. In fact, for the SISO case the improvement was equivalent to having multiple receive antennas equal to the number of ISI taps [4].

A different perspective for opportunistic communication was presented in [1], [2]. A strategy that combined high rate communications with high reliability (diversity) was investigated. Clearly, the overall code will still be governed by the rate-reliability tradeoff, but the idea was to ensure the high reliability (diversity) of at least part of the total information. These are called diversity-embedded codes [1], [2]. In [2] it was shown that when we have one degree of freedom (one transmit many receive or one receive many transmit antennas) the D-M tradeoff was successively refinable. That is, the high priority scheme (with higher diversity order) can attain the optimal diversity-multiplexing (D-M) performance as if the low priority stream was absent. However, the low priority scheme (with lower diversity order) attains the same D-M performance as that of the aggregate rate of the two streams. When there is more than one degree of freedom (for example, parallel fading channels) such a successive refinement property does not hold [3].

In this paper we investigate the diversity embedded codes for an ISI channel with single transmit and receive antenna. Since the Fourier basis is the eigenbasis for linear time invariant channels we can decompose the transmission into a set of parallel channels. Since it is known that the D-M tradeoff for parallel fading channels is not successively refinable [3], it is tempting to expect the same for fading ISI channels. However, the main result of this paper is that for SISO fading ISI channels the D-M tradeoff is indeed successively refinable. The correlations of the fading across the parallel channels seem to cause the difference in the behavior. The structural observations in lemma 3 give insight into these correlations. This will be made more precise in the paper.

The paper is organized as follows. In Section 2 we formulate the problem statement and present the notation. Section 3 gives a variation of the proof in [4] of the D-M tradeoff for ISI channels which makes a connection to the diversity embedded codes. We explore the role of correlation in successive refinement through a specific example. Section 4 presents the statement and the proof for the successive refinability of the D-M tradeoff for ISI channels. We conclude the paper with a brief discussion followed by the details of the proofs in the appendix.

II. Problem Statement

Consider communication over a quasi static fading channel with Inter-symbol Interference (ISI)

$$y[n] = h_0 x[n] + h_1 x[n-1] + \ldots + h_\nu x[n-\nu] + z[n] \quad (1)$$

The $\nu + 1$ i.i.d. fading coefficients are $h_i \sim \mathcal{CN}(0,1)$ and fixed for the duration of the block length $(N+\nu)$. The additive noise $z[n]$ is i.i.d. circularly symmetric Gaussian with unit variance. As is standard in these problems, we assume perfect channel knowledge only at the receiver.

The coding scheme is limited to one quasi-static transmission block of size $N+\nu$. Consider a sequence of coding schemes with transmission rate as a function of SNR given by $R(SNR)$ and an average error probability of decoding $P_e(SNR)$. Analogous to [6] we define the multiplexing rate r and the diversity order d as follows,

$$d = \lim_{SNR\to\infty} -\frac{\log P_e(SNR)}{\log(SNR)}, \; r = \lim_{SNR\to\infty} \frac{R(SNR)}{\log(SNR)}. \quad (2)$$

With these definitions, the D-M tradeoff for ISI channels was established in [4].

EPFL, Lausanne, Switzerland, S. Dusad was supported in part by SNSF Grant # 200021-105640/1. S. N. Diggavi is part of the SNSF supported NCCR-MICS center on wireless sensor networks. Email: {suhas.diggavi,sanket.dusad}@epfl.ch.

Theorem 1: [4] The diversity multiplexing tradeoff for the system model in (1) is bounded by,

$$(\nu+1)\left(1-\frac{N+\nu}{N}r\right) \leq d_{isi}(r) \leq (\nu+1)(1-r) \quad (3)$$

In this paper we explore the performance of diversity embedded codes over ISI channels. For clarity we focus on two streams but the procedure can be generalized to more than two levels.

Let $\mathcal{H}$ denote the message set from the first information stream and $\mathcal{L}$ denote that from the second information stream. The rates for the two message sets as a function of SNR are, respectively, $R_H(SNR)$ and $R_L(SNR)$. The decoder jointly decodes the two message sets and we can define two error probabilities, $P_e^H(SNR)$ and $P_e^L(SNR)$, which denote the average error probabilities for message sets $\mathcal{H}$ and $\mathcal{L}$ respectively. We want to characterize the tuple (r_H, d_H, r_L, d_L) of rates and diversities for the ISI channel that are achievable, where analogous to (2),

$$d_H = \lim_{SNR\to\infty} -\frac{\log P_e^H(SNR)}{\log(SNR)},\; r_H = \lim_{SNR\to\infty} \frac{R_H(SNR)}{\log(SNR)}$$
$$d_L = \lim_{SNR\to\infty} -\frac{\log P_e^L(SNR)}{\log(SNR)},\; r_L = \lim_{SNR\to\infty} \frac{R_L(SNR)}{\log(SNR)}$$

Also, we assume that $d_H \geq d_L$. Note that for the joint codebook $\{\mathcal{H}, \mathcal{L}\}$ the total multiplexing rate is $r_H + r_L$ and the diversity $d = min(d_H, d_L) = d_L$. We use the special symbol $\doteq$ to denote exponential equality *i.e.*, we write $f(SNR) \doteq SNR^b$ to denote

$$\lim_{SNR\to\infty} \frac{\log f(SNR)}{\log(SNR)} = b$$

and $\dot{\leq}$ and $\dot{\geq}$ are defined similarly.

From an information-theoretic point of view [2] focused on the case when there is one degree of freedom (*i.e.,* $\min(M_t, M_r) = 1$). In that case if we consider $d_H \geq d_L$ without loss of generality, the following result was established in [2].

Theorem 2: When $\min(M_t, M_r) = 1$, then the diversity-multiplexing trade-off curve is successively refinable, i.e., for any multiplexing rates r_H and r_L such that $r_H + r_L \leq 1$, the diversity orders $d_H \geq d_L$,

$$d_H = d^{opt}(r_H), \quad (4)$$
$$d_L = d^{opt}(r_H + r_L) \quad (5)$$

are achievable, where $d^{opt}(r)$ is the optimal diversity order given in [6].

■

Since the overall code has to still be governed by the rate-diversity trade-off given in [6], it is clear that the trivial outer bound to the problem is that $d_H \leq d^{opt}(r_H)$ and $d_L \leq d^{opt}(r_H + r_L)$. Hence Theorem 2 shows that the best possible performance can be achieved. This means that for $\min(M_t, M_r) = 1$, we can design ideal *opportunistic* codes. This analysis was done for flat fading channels and we will show a similar theorem for ISI fading channels.

III. ISI TRADEOFF

In this section we give an alternative interpretation of the D-M tradeoff for ISI channels for the particular case of two taps *i.e.*, $\nu = 1$. This exercise will help us to see the difference between the fading ISI channel and the i.i.d. parallel fading channel models.

Rewriting the equation (1) for the case of two taps, we have,

$$y[n] = h_0 x[n] + h_1 x[n-1] + z[n] \quad (6)$$

Assume a scheme in which one data symbol is sent in every $(\nu+1)$ transmissions from a QAM constellation of size SNR^r. With this strategy there is no interference between successive transmitted symbols and the receiver performs matched filtering to recover the symbol from the $\nu+1$ copies of the received signal. This gives us the matched filter upper bound to the diversity or the lower bound to the error probability,

$$d_{isi}(r) \leq 2(1-r)$$

where r is the multiplexing rate and d is the diversity order.

For the lower bound to the diversity consider a transmission strategy in which we assume that after transmission over a block length N, in the last $\nu = 1$ instants zero symbol is transmitted in order to avoid interblock interference. Therefore, the received vector over the block of length $N+\nu$ can be written as,

$$\begin{bmatrix} y[0] \\ y[1] \\ \vdots \\ y[N] \end{bmatrix} = \underbrace{\begin{bmatrix} h_0 & 0 & \dots & \dots & h_1 \\ h_1 & h_0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & \dots & h_1 & h_0 & 0 \\ 0 & \dots & 0 & h_1 & h_0 \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \\ 0 \end{bmatrix} + \mathbf{z} \quad (7)$$

where $\mathbf{z} = \begin{bmatrix} z[0] & z[1] & \dots & z[N] \end{bmatrix}^T$. Note now that the channel matrix $\mathbf{H}$ is a circulant matrix. Proceeding as in [4], look at the circulant matrix $\mathbf{H} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^*$ in the frequency domain where $\mathbf{Q}$ and $\mathbf{Q}^*$ are truncated DFT matrices and $\mathbf{\Lambda}$ is a diagonal matrix with the diagonal elements given by $\Lambda_l = h_0 + h_1 e^{-\frac{l2\pi j}{N+1}}$ for $l = 0, \dots, N$. If $(N+1)$ is divisible by two,we can view this as $\frac{(N+1)}{2}$ sets of 2 parallel independent channels. Communicating over each such set of 2 parallel channels at a rate $2r$, we get the effective multiplexing rate, $\tilde{r} = r\frac{N}{N+1}$. The diversity for this multiplexing rate is $2 - 2r = 2(1 - \frac{N+1}{N}\tilde{r})$ as given in [6].

Note that in the above argument we do not utilize the fact that correlations exist across these sets of independent channels anywhere. In this case correlations did not matter since it achieves[1] *the matched filter upper bound.*

In order to illustrate the impact of correlation between the frequency domain coefficients, we will specifically consider

[1] asymptotically in N.

the case for $N = 3$ and $\nu = 1$. Using (7) and the Fourier decomposition we see that we have four parallel channels,

$$\tilde{y}_l = \Lambda_l \tilde{x}_l + \tilde{z}_l \qquad l = 0, \ldots, 3$$

In the spirit of the above parallel channel argument we can view these as two sets of parallel channels $\{\tilde{y}_0, \tilde{y}_2\}$ and $\{\tilde{y}_1, \tilde{y}_3\}$ each consisting of two sub-channels. Given this view since the D-M tradeoff for parallel channels are not successively refinable we would expect that such a characterization also hold for the ISI D-M tradeoff. However, since $\Lambda_0 = h_0 + h_1$, $\Lambda_2 = h_0 - h_1$, $\Lambda_1 = h_0 - jh_1$, $\Lambda_3 = h_0 + jh_1$, we see that the fading across the two sets of parallel channels are correlated. In particular, if $|h_0|^2 \dot{>} SNR^{-(1-r)}$ then asymptotically $|\Lambda_l|^2 \dot{\leq} SNR^{-(1-r)}$ for at most one $l \in \{0, \ldots, 3\}$. Therefore, it is possible to code across these sets of parallel channels to get better performance instead of treating them independently.

This example gives the intuition to use the following method to prove the diversity multiplexing tradeoff for the ISI channel. Define a set $\mathcal{A}$ of events such that

$$\mathcal{A} = \left\{ \mathbf{h} : |h_0|^2 \dot{\leq} \frac{1}{SNR^{1-r}} \text{ and } |h_1|^2 \dot{\leq} \frac{1}{SNR^{1-r}} \right\} \quad (8)$$

For high SNR it follows that the probability of the set $\mathcal{A}$ occurring is $P(\mathcal{A}) = SNR^{-2(1-r)}$ and $P(\mathcal{A}^c) = 1 - SNR^{-2(1-r)}$. At each time instant independently transmit one symbol from a constellation with $d_{min}^2 \dot{\geq} SNR^{(1-r)}$ for N time instants and pad it with ν zero symbols. For detection, given that $\mathbf{h} \in \mathcal{A}^c$, we proceed as in the proof of lemma 3 with the detection of the $N + \nu$ length transmitted sequence in the frequency domain. Clearly, the error probability of the scheme with this decoder, denoted by $P_e^D(SNR)$, is an upper bound to the error probability *i.e.*, $P_e(SNR) \leq P_e^D(SNR)$. Therefore, we can write,

$$\begin{aligned}
&P_e(SNR) = P(\mathcal{A}) P_e(SNR \mid \mathbf{h} \in \mathcal{A}) + \\
&\quad P(\mathcal{A}^c) P_e(SNR \mid \mathbf{h} \in \mathcal{A}^c) \\
&\dot{\leq} P(\mathcal{A}) + \left(1 - SNR^{-2(1-r)}\right) P_e(SNR \mid \mathbf{h} \in \mathcal{A}^c) \\
&\dot{\leq} SNR^{-2(1-r)} + \left(1 - SNR^{-2(1-r)}\right) P_e^D(SNR \mid \mathbf{h} \in \mathcal{A}^c) \\
&\doteq SNR^{-2(1-r)}.
\end{aligned}$$

The last equality is true due to lemma 4 (given in Section 4) which states that if $\mathbf{h} \in \mathcal{A}^c$ the probability of error decays exponentially in SNR. This lemma in turn is based on a structural observation made in lemma 3 (also see Section 3) that at most ν coefficients in frequency domain will be smaller than $\min_l |h_l|^2$ and here given that $\mathbf{h} \in \mathcal{A}^c$ at most ν coefficients will be smaller than $SNR^{-(1-r)}$. This method of analysis turns out to be more useful for us and takes into account the fact that the sets of parallel channels are correlated.

This example was specifically for $N = 3$ and 2 taps. A similar analysis carries over for the case of general N and ν. As summarized in lemma 2 in Section 4, for a finite N with $(\nu+1)$ taps either all the taps will be of order less than $SNR^{-(1-r)}$ or at most ν taps will be of order less than $SNR^{-(1-r)}$.

IV. Successive Refinement of the ISI D-M tradeoff

In this section we will formally prove the successive refinement of the D-M tradeoff for ISI channels. The intuition of the effect of fading in the frequency domain is captured by the following result which is proved in the appendix.

Lemma 3: For a $(\nu+1)$ tap ISI channel we have the taps in the frequency domain are given by,

$$\Lambda_k = \sum_{m=0}^{\nu} h_m e^{-\frac{2\pi j}{(N+\nu)} km} \qquad k = \{0, \ldots, (N+\nu-1)\}$$

Define the sets $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{A}$ as,

$$\begin{aligned}
&\mathcal{F} = \{i : |\Lambda_i|^2 \dot{\leq} SNR^{-(1-r)}\}, && (9) \\
&\mathcal{G} = \{i : |\Lambda_i|^2 \doteq \max_{l \in \{0,1,\ldots,\nu\}} |h_l|^2\}, \\
&\mathcal{A} = \left\{ \mathbf{h} : |h_m|^2 \dot{\leq} \frac{1}{SNR^{1-r}} \quad \forall m \in \{0, 1, \ldots, \nu\} \right\} && (10)
\end{aligned}$$

With these definitions we have:

(a)) Given that $\mathbf{h} \in \mathcal{A}^c$, $|\mathcal{F}| \leq \nu$, i.e., at most ν taps in the frequency domain are (asymptotically) of magnitude less than $SNR^{-(1-r)}$.

(b)) $|\mathcal{G}^c| \leq \nu$ i.e., at least N taps of the $N + \nu$ taps in the frequency domain are (asymptotically) of magnitude $\max(|h_0|^2, |h_1|^2, \ldots, |h_\nu|^2)$,

$$|\{k : |\Lambda_k|^2 \dot{<} \max(|h_0|^2, |h_1|^2, \ldots, |h_\nu|^2)\}| \leq \nu.$$

∎

Note that (b) along with $\mathbf{h} \in \mathcal{A}^c$ implies (a) and therefore is the stronger claim. Here is an intuition of why such a result will hold. Consider the polynomial

$$\Lambda(z) = \sum_{m=0}^{\nu} h_m z^m,$$

which evaluates to the Fourier transform for $z = e^{-\frac{2\pi j}{(N+\nu)} k}$. Hence, if we evaluate the polynomial at $z = e^{-\frac{2\pi j}{(N+\nu)} k}$, for $k = \{0, \ldots, (N+\nu-1)\}$, at most ν values can be zero and at least N values are bounded away from zero. Therefore, if SNR is large enough, it is clear that at least N values would be "larger" than $SNR^{-(1-r)}$. The details of the proof are in the appendix.

Now consider transmission using uncoded QAM such that the minimum distance between any two points in the constellation d_{min} is such that $d_{min}^2 \dot{\geq} SNR^{(1-r)}$. Defining $\mathcal{F} = \{i : |\Lambda_i|^2 \dot{\leq} SNR^{-(1-r)}\}$ we have from the lemma above that $|\mathcal{F}| \leq \nu$. Ignore these ν channels and examine the remaining N channels in $\mathcal{F}^c$. We can show that the distance between codewords in these channels is still asymptotically larger than $SNR^{(1-r)}$. Since the pairwise error probability is a Q function, we can show that the error probability decays exponentially in SNR. This is summarized in the following lemma, the proof of which is in the appendix.

Lemma 4: Assume that the minimum distance d_{min} between any two points in the constellation ($\mathcal{X}$) from which the signal is transmitted is $d_{min}^2 \dot{\geq} SNR^{(1-r)}$. Assume uncoded transmission such that at each time instant one symbol is independently transmitted from the constellation for N time instants followed by a padding with ν zero symbols. For a finite period of communication (finite N) given that $\mathbf{h} \in \mathcal{A}^c$ (see (7)), the error probability P_e decays exponentially in SNR. ∎

The part (a) of lemma 3 and lemma 4 can be combined together to give an alternative proof of the diversity multiplexing tradeoff of the ISI channel. But to prove the successive refinement of the D-M tradeoff of the ISI channel we need the stronger result in the part (b) of lemma 3.

We will prove a lemma analogous to lemma 4 for the case of superposition coding. This will be useful in our proof to show the successive refinement of the D-M tradeoff for ISI channels. Since we are padding every N symbols with ν zeros, to communicate at an effective rate of r using uncoded QAM transmission, we need to send symbols from a QAM constellation of size $SNR^{\tilde{r}}$ where $\tilde{r} = \frac{r(N+\nu)}{N}$. Let $\mathcal{X}_H$ be QAM constellation instant of size $SNR^{\tilde{r}_H}$ and power constraint SNR. Similarly let $\mathcal{X}_L$ be a QAM constellation of size $SNR^{\tilde{r}_L}$ and power constraint $SNR^{1-\beta}$, where $\beta > \tilde{r}_H$. As before, define

$$\mathcal{A}_H = \left\{ \mathbf{h} : |h_m|^2 \dot{\leq} \frac{1}{SNR^{1-\tilde{r}_H}} \quad \forall m \in \{0, 1, \dots, \nu\} \right\} \tag{11}$$

Lemma 5: Using the $\mathcal{X}_H$ and $\mathcal{X}_L$ for signaling, assume uncoded superposition transmission such that at each time instant symbols are independently chosen and superposed from each constellation ($\mathcal{X}_H$, $\mathcal{X}_L$) for N time instants followed by a padding with ν zero symbols. For a finite period of communication (finite N) given that $\mathbf{h} \in \mathcal{A}_H^c$ (see (11)), the error probability of detecting the set of symbols sent from the higher constellation ($\mathcal{X}_H$) denoted by $P_e^H(SNR)$ decays exponentially in SNR. ∎

In this lemma we critically use the fact that *all* except at most ν taps in the frequency domain, are asymptotically of *equal* magnitude ($\max_{l \in \{0,1,\dots,\nu\}} |h_l|^2$).

Using these lemmas we will prove the following theorem on the successive refinement.

Theorem 6: Consider a ν tap point to point SISO ISI channel. The diversity multiplexing tradeoff for this channel is successively refinable, *i.e.*, for any multiplexing gains r_H and r_L such that $r_H + r_L \leq \frac{N}{N+\nu}$ the achievable diversity orders given by $d_H(r_H)$ and $d_L(r_L)$ are bounded as,

$$(\nu+1)\left(1 - \frac{N+\nu}{N} r_H\right) \leq d_H(r_H) \leq (\nu+1)\left(1 - r_H\right), \tag{12}$$

$$(\nu+1)\left(1 - \frac{N+\nu}{N} (r_H + r_L)\right) \leq d_L(r_L) \leq (\nu+1)\left(1 - (r_H + r_L)\right) \tag{13}$$

where N is finite and does not grow with SNR. ∎

Proof: To show the successive refinement we use superposition coding and assume two streams with uncoded QAM codebooks for each stream, as in [2]. Assume that given a total power constraint P we allocate powers P_H and P_L to the high and low priority streams respectively. We design the power allocation such that at high signal to noise ratio, we have $SNR_H \doteq SNR$ and $SNR_L \doteq SNR^{1-\beta}$ for $\beta \in [0, 1]$. Let $\mathcal{X}_H$ be QAM constellation instant of size $SNR^{\tilde{r}_H}$ with minimum distance $(d_{min}^H)^2 = SNR^{1-\tilde{r}_H}$. Similarly let $\mathcal{X}_L$ be a QAM constellation of size $SNR^{\tilde{r}_L}$ with minimum distance $(d_{min}^L)^2 = SNR^{1-\beta-\tilde{r}_L}$, where $\beta > \tilde{r}_H$. The symbol transmitted at the k^{th} instant is the superposition of a symbol from $\mathcal{X}_H$, $\mathcal{X}_L$ given by,

$$x[k] = x_H[k] + x_L[k] \qquad \text{where } x_H[k] \in \mathcal{X}_H,\ x_L[k] \in \mathcal{X}_L$$

It can be shown [2] that even with the above superposition coding, if $\beta > \tilde{r}_H$ the order of magnitude of the effective minimum distance between two points in the constellation $\mathcal{X}_H$ is preserved.

The upper bound in both (12) and (13) is trivial and follows from the matched filter bound. We will investigate the lower bound in (12). At each time instant superpose symbols from the higher and lower layers for N time instants and pad them with ν zero symbols at the end. We consider this particular transmission scheme and for detection, given that $\mathbf{h} \in \mathcal{A}_H^c$, (where $\mathcal{A}_H$ is as defined in equation (11)), we proceed as in lemma 3. Therefore, we can write,

$$\begin{aligned} P_e^H(SNR) &= P(\mathcal{A}_H) P_e(SNR \mid \mathbf{h} \in \mathcal{A}_H) + \\ &\quad P(\mathcal{A}_H^c) P_e(SNR \mid \mathbf{h} \in \mathcal{A}_H^c) \qquad (14) \\ &\dot{\leq} P(\mathcal{A}_H) + \left(1 - SNR^{-(\nu+1)(1-\tilde{r}_H)}\right) P_e(SNR \mid \mathbf{h} \in \mathcal{A}_H^c) \\ &\dot{\leq} SNR^{-(\nu+1)(1-\tilde{r}_H)} + \\ &\quad \left(1 - SNR^{-(\nu+1)(1-\tilde{r}_H)}\right) P_e^D(SNR \mid \mathbf{h} \in \mathcal{A}_H^c) \qquad (15) \end{aligned}$$

for communication at an effective rate of $r_H = \frac{N}{N+\nu} \tilde{r}_H$.

For decoding the higher layer we treat the signal on the lower layer as noise. Given that $\mathbf{h} \in \mathcal{A}_H^c$ and choosing $\beta > \tilde{r}_H$ we conclude from lemma 5 that the second term in (15) decays exponentially in SNR. Therefore,

$$P_e^H(SNR) \dot{\leq} \frac{1}{SNR^{(\nu+1)(1-\tilde{r}_H)}} \tag{16}$$

Or equivalently,

$$(\nu+1)\left(1 - \frac{N+\nu}{N} r_H\right) \leq d_H(r_H) \tag{17}$$

Once we have decoded the upper layer we subtract its contribution from the lower layer. Proceeding as above, define

$$\mathcal{A}_L = \left\{ \mathbf{h} : |h_m|^2 \dot{\leq} \frac{1}{SNR^{1-\tilde{r}_L-\beta}} \quad \forall m \in \{0, 1, \dots, \nu\} \right\} \tag{18}$$

$$
\begin{bmatrix} \mathbf{y}[0] \\ \mathbf{y}[1] \\ \vdots \\ \mathbf{y}[N] \\ \dots \\ \mathbf{y}[N+\nu] \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{h}_0 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{h}_\nu & \dots & \mathbf{h}_2 & \mathbf{h}_1 \\ \mathbf{h}_1 & \mathbf{h}_0 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{h}_\nu & \dots & \mathbf{h}_2 \\ \vdots & \vdots & & & \dots & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{h}_\nu & \mathbf{h}_{\nu-1} & \dots & \mathbf{h}_1 & \mathbf{h}_0 \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \\ \mathbf{0}_{\nu\times 1} \end{bmatrix}}_{\begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix}} + \underbrace{\begin{bmatrix} \mathbf{z}[0] \\ \mathbf{z}[1] \\ \vdots \\ \mathbf{z}[N] \\ \vdots \\ \mathbf{z}[N+\nu] \end{bmatrix}}_{\mathbf{z}} \tag{22}
$$

For high SNR it follows that,

$$
\begin{aligned} P(\mathcal{A}_L) &= SNR^{-(\nu+1)(1-\tilde{r}_L-\beta)}, \\ P(\mathcal{A}_L^c) &= 1 - SNR^{-(\nu+1)(1-\tilde{r}_L-\beta)}. \end{aligned}
$$

For the lower layer we have that $(d_{min}^L)^2 = \frac{SNR^{1-\beta}}{SNR^{\tilde{r}_L}} = SNR^{1-\beta-\tilde{r}_L}$. Using lemma 4, taking β arbitrarily close to $\tilde{r}_H$ we can conclude that,

$$
\begin{aligned} (\nu+1)\left(1 - \frac{N+\nu}{N}(r_H + r_L)\right) &\leq d_L(r_L) \\ &\leq (\nu+1)\left(1 - (r_H + r_L)\right) \end{aligned} \tag{19}
$$

Comparing this with Theorem 1 we can see that the diversity multiplexing tradeoff for the ISI channel is successively refinable since $d_H(r_H) = d_{isi}(r_H)$ and $d_L(r_L) = d_{isi}(r_H + r_L)$. ■

The intuition that was used in deriving the successive refinement of the SISO tradeoff for ISI channels was that given that $h \in \mathcal{A}$ at most ν taps in the frequency domain are zero and the remaining are "good" and of the same magnitude. This intuition can also be carried over to show the successive refinability of the SIMO channel with M_r receive antennas and one transmit antenna. In this case, the received vector at the n^{th} instant is given by,

$$
\mathbf{y}[n] = \mathbf{h}_0 x[n] + \mathbf{h}_1 x[n-1] + \ldots + \mathbf{h}_\nu x[n-\nu] + \mathbf{z}[n] \tag{20}
$$

where $\mathbf{y}, \mathbf{h}_i, \mathbf{z} \in \mathbb{C}^{M_r \times 1}$. Assume that the $\nu + 1$ fading coefficients are $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{M_r})$ and fixed for the duration of the block length $(N + \nu)$ and $\mathbf{h}_i$ is independent of $\mathbf{h}_j$. Let $h_i^{(p)}$ represent the i^{th} tap coefficient between the transmitter and the p^{th} receive antenna. We will denote $\mathbf{C} = circ\{c_1, c_2, \ldots, c_T\}$ to be the $T \times T$ circulant matrix given by

$$
\mathbf{C} = \begin{bmatrix} c_1 & c_2 & c_3 & \dots & c_{T-1} & c_T \\ c_T & c_1 & c_2 & \dots & c_{T-2} & c_{T-1} \\ \vdots & & \vdots & \ddots & & \vdots \\ c_2 & c_3 & c_4 & \dots & c_T & c_1 \end{bmatrix} \tag{21}
$$

Consider a transmission scheme in which we transmit uncoded symbols from a QAM constellation of size SNR^r for N time instants and pad them with ν zero symbols at the end. Consider a transmission scheme in which one data symbol is sent at every instant from a QAM constellation of size SNR^r. Therefore, the received vector over the block of length $N+\nu$ can be written as in equation (22) at the top of the page, where $\mathbf{H} \in \mathbb{C}^{(N+\nu)M_r \times (N+\nu)}$, $\mathbf{y} \in \mathbb{C}^{(N+\nu)M_r \times 1}$, $\mathbf{z} \in \mathbb{C}^{(N+\nu)M_r \times 1}$ and $\mathbf{x} \in \mathbb{C}^{N \times 1}$. By reordering the rows we can write the received vector in terms of circulant matrices as,

$$
\begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \\ \vdots \\ \mathbf{y}^{(M_r)} \end{bmatrix} = \begin{bmatrix} \mathbf{H}^{(1)} \\ \mathbf{H}^{(2)} \\ \vdots \\ \mathbf{H}^{(M_r)} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{0}_{\nu\times 1} \end{bmatrix} + \begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \vdots \\ \mathbf{z}^{(M_r)} \end{bmatrix} \tag{23}
$$

where $\mathbf{H}^{(1)}, \ldots, \mathbf{H}^{(M_r)} \in \mathbb{C}^{(N+\nu)\times(N+\nu)}$ are circulant matrices given by,

$$
\mathbf{H}^{(p)} = circ\{h_0^{(p)}, 0, \ldots, 0, h_\nu^{(p)}, \ldots, h_2^{(p)}, h_1^{(p)}\}
$$

for $p \in \{1, 2, \ldots, M_r\}$ and,

$$
\mathbf{y}^{(p)} = \begin{bmatrix} y^{(p)}[0] \\ y^{(p)}[1] \\ \vdots \\ y^{(p)}[N+\nu] \end{bmatrix}
$$

where $y^{(p)}[n]$ represents the symbol received at the p^{th} receive antenna in the n^{th} time instant.

Since the $\mathbf{H}^{(p)}$ are circulant matrices we can write them using the frequency domain notation as $\mathbf{H}^{(p)} = \mathbf{Q}\mathbf{\Lambda}^{(p)}\mathbf{Q}^*$ where $\mathbf{Q}, \mathbf{Q}^* \in \mathbb{C}^{(N+\nu)\times(N+\nu)}$ are truncated DFT matrices as defined earlier and $\mathbf{\Lambda}^{(p)}$ are diagonal matrices with the elements given by,

$$
\mathbf{\Lambda}^{(p)} = diag\left\{ \Lambda_k^{(p)} : \Lambda_k^{(p)} = \sum_{m=0}^{\nu} h_m^{(p)} e^{-\frac{2\pi j}{(N+\nu)}km} \right\}
$$

for $k = \{0, \ldots, (N+\nu-1)\}$. Therefore the equation (23) can be rewritten as,

$$\begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \\ \vdots \\ \mathbf{y}^{(M_r)} \end{bmatrix} = \begin{bmatrix} \mathbf{H}^{(1)} \\ \mathbf{H}^{(2)} \\ \vdots \\ \mathbf{H}^{(M_r)} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{0}_{\nu\times 1} \end{bmatrix} + \begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \vdots \\ \mathbf{z}^{(M_r)} \end{bmatrix} \quad (24)$$

$$= \begin{bmatrix} \mathbf{Q}\mathbf{\Lambda}^{(1)}\mathbf{Q}^* \\ \mathbf{Q}\mathbf{\Lambda}^{(2)}\mathbf{Q}^* \\ \vdots \\ \mathbf{Q}\mathbf{\Lambda}^{(M_r)}\mathbf{Q}^* \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{0}_{\nu\times 1} \end{bmatrix} + \begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \vdots \\ \mathbf{z}^{(M_r)} \end{bmatrix} \quad (25)$$

$$= \begin{bmatrix} \mathbf{Q}\mathbf{\Lambda}^{(1)}\tilde{\mathbf{Q}}^*\mathbf{x} \\ \mathbf{Q}\mathbf{\Lambda}^{(2)}\tilde{\mathbf{Q}}^*\mathbf{x} \\ \vdots \\ \mathbf{Q}\mathbf{\Lambda}^{(M_r)}\tilde{\mathbf{Q}}^*\mathbf{x} \end{bmatrix} + \begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \vdots \\ \mathbf{z}^{(M_r)} \end{bmatrix} \quad (26)$$

where $\tilde{\mathbf{Q}}^*$ is a $(N+\nu)\times N$ matrix which is obtained by deleting the last ν columns of the matrix $\mathbf{Q}^*$ (similar as in the proof of lemma 4).

Lemma 7: For a $(\nu+1)$ tap, M_r receive antennas ISI channel we have the taps in the frequency domain are given by,

$$\Lambda_k^{(p)} = \sum_{m=0}^{\nu} h_m^{(p)} e^{-\frac{2\pi j}{(N+\nu)}km}$$

for $k = \{0, \ldots, (N+\nu-1)\}$ and $p \in \{0, \ldots, M_r\}$. Define the sets $\mathcal{F}^{(p)}$, $\mathcal{G}^{(p)}$ and $\mathcal{M}$ as,

$$\mathcal{F}^{(p)} = \{k : |\Lambda_k^{(p)}|^2 \dot{\leq} SNR^{-(1-r)}\}, \quad (27)$$

$$\mathcal{G}^{(p)} = \{k : |\Lambda_k^{(p)}|^2 \doteq \max_{l\in\{0,1,\ldots,\nu\}} |h_l^{(p)}|^2\} \quad (28)$$

$$\mathcal{M} = \{\mathbf{h} : |h_i^{(p)}|^2 \dot{\leq} \frac{1}{SNR^{1-r}} \quad \forall i \in \{0, \ldots, \nu\},\ \forall p \in \{1, \ldots, M_r\}\} \quad (29)$$

With $\overline{\mathcal{G}^{(p)}}$ representing the complement of the set $\mathcal{G}^{(p)}$, we have

$$|\overline{\mathcal{G}^{(p)}}| \leq \nu \quad \forall p$$

and given that $\mathbf{h} \in \mathcal{M}^c$ this means that,

$$\exists p \in \{1, 2, \ldots, M_r\}\ s.\ t.\ |\mathcal{F}^{(p)}| \leq \nu \quad (30)$$

■

Proof: From lemma 3 for each p it is clear that $|\overline{\mathcal{G}^{(p)}}| \leq \nu$. Since $\mathbf{h} \in \mathcal{M}^c$ there exists at least one (i, p) pair such that $|h_i^{(p)}|^2 \dot{\leq} \frac{1}{SNR^{1-r}}$. Then from lemma 3 it follows that for this particular p, $|\mathcal{F}^{(p)}| \leq \nu$ and $|\overline{\mathcal{G}^{(p)}}| \leq \nu$. ■

As before, define

$$\mathcal{M}_H = \{\mathbf{h} : |h_i^{(p)}|^2 \dot{\leq} \frac{1}{SNR^{1-r_H}} \quad \forall i \in \{0, 1, \ldots, \nu\},\ \forall p \in \{1, 2, \ldots, M_r\}\} \quad (31)$$

Since all the tap coefficients are i.i.d we have that $P(\mathcal{M}) = SNR^{-M_r(\nu+1)(1-r)}$ and $P(\mathcal{M}_H) = SNR^{-M_r(\nu+1)(1-r_H)}$.

Lemma 8: Using the $\mathcal{X}_H$ and $\mathcal{X}_L$ defined earlier for signaling, assume uncoded superposition transmission such that at each time instant one symbol is independently transmitted from each constellation ($\mathcal{X}_H$, $\mathcal{X}_L$) for N time instants followed by a padding with ν zero symbols. For a finite period of communication (finite N) given that $\mathbf{h} \in \mathcal{M}_H^c$ (see (31)), the error probability of detecting the set of symbols sent from the higher constellation ($\mathcal{X}_H$) denoted by $P_e^H(SNR)$ decays exponentially in SNR. ■

We will just give an outline of the proof as the details are similar to the proof of lemma 5 given in the Appendix. From lemma 7 there exists at least one set of $(N+\nu)$ coefficients in the frequency domain through which $\hat{\mathbf{Q}}\mathbf{x}$ passes such that at most ν taps of the available $(N+\nu)$ taps in this set are of magnitude smaller than $SNR^{-(1-r_H)}$. Then from lemma 5 it directly follows that the error probability decays exponentially in SNR.

Theorem 9: Consider a ν tap point to point SIMO ISI channel with M_r receive antennas. The diversity multiplexing tradeoff for this channel is successively refinable, *i.e.*, for any multiplexing gains r_H and r_L such that $r_H + r_L \leq \frac{N}{N+\nu}$ the achievable diversity orders given by $d_H(r_H)$ and $d_L(r_L)$ are bounded as,

$$M_r(\nu+1)\left(1 - \frac{N+\nu}{N}r_H\right) \leq d_H(r_H) \leq M_r(\nu+1)(1-r_H), \quad (32)$$

$$M_r(\nu+1)\left(1 - \frac{N+\nu}{N}(r_H+r_L)\right) \leq d_L(r_L) \leq M_r(\nu+1)(1-(r_H+r_L)) \quad (33)$$

where N is finite and does not grow with SNR. ■

Proof: As in theorem 6 use superposition coding and assume two streams with uncoded QAM codebooks $\mathcal{X}_H$ and $\mathcal{X}_L$ for the higher and lower priority streams respectively. Choose $SNR_H \doteq SNR$ and $SNR_L \doteq SNR^{1-\beta}$ for $\beta \in [0,1]$. Also let $|\mathcal{X}_H| = SNR^{\tilde{r}_H}$ and $|\mathcal{X}_L = SNR^{\tilde{r}_L}$. As in theorem 6 the minimum distances are $(d_{min}^H)^2 = SNR^{1-\tilde{r}_H}$, $(d_{min}^L)^2 = SNR^{1-\beta-\tilde{r}_L}$, and if $\beta > \tilde{r}_H$. the order of magnitude of the effective minimum distance between two points in the constellation $\mathcal{X}_H$ is preserved.

The symbol transmitted at the k^{th} instant is the superposition of a symbol from $\mathcal{X}_H$, $\mathcal{X}_L$ given by,

$$x[k] = x_H[k] + x_L[k] \qquad \text{where } x_H[k] \in \mathcal{X}_H,\ x_L[k] \in \mathcal{X}_L$$

The upper bound in both (32) and (33) is trivial and follows from the matched filter bound. We will investigate the lower bound in (32). At each time instant superpose symbols from the higher and lower layers for N time instants and pad them with ν zero symbols at the end. We consider this particular transmission scheme and for detection, given that $\mathbf{h} \in \mathcal{M}_H^c$, (where $\mathcal{M}_H$ is as defined in equation (31)), we proceed as in lemma 3. Therefore, we can write,

$$\begin{aligned}P_e^H(SNR) &= P(\mathcal{M}_H)P_e(SNR \mid \mathbf{h} \in \mathcal{M}_H) \\ &\quad + P(\mathcal{M}_H^c)P_e(SNR \mid \mathbf{h} \in \mathcal{M}_H^c) \\ &\dot{\leq} SNR^{-M_r(\nu+1)(1-\tilde{r}_H)} + \qquad (34)\\ &\quad \left(1 - SNR^{-M_r(\nu+1)(1-\tilde{r}_H)}\right) P_e^D(SNR \mid \mathbf{h} \in \mathcal{M}_H^c). \qquad (35)\end{aligned}$$

for communication at an effective rate of $r_H = \frac{N}{N+\nu}\tilde{r}_H$.

From lemma 8, where we treat the signal on the lower layer as noise, we get that the second term in (35) decays exponentially in SNR. Therefore,

$$P_e^H(SNR) \dot{\leq} \frac{1}{SNR^{M_r(\nu+1)(1-\tilde{r}_H)}} \qquad (36)$$

Or equivalently,

$$M_r(\nu+1)\left(1 - \frac{N+\nu}{N}r_H\right) \leq d_H(r_H) \qquad (37)$$

Once we have decoded the upper layer we subtract its contribution from the lower layer. Proceeding as above, define

$$\begin{aligned}\mathcal{M}_L = \{\mathbf{h} : |h_i^{(p)}|^2 \dot{\leq} \frac{1}{SNR^{1-\tilde{r}_L-\beta}} \\ \forall i \in \{0,1,\ldots,\nu\},\ \forall p \in \{1,2,\ldots,M_r\}\}\end{aligned} \qquad (38)$$

For high SNR it follows that $P(\mathcal{M}_L) = SNR^{-M_r(\nu+1)(1-\tilde{r}_L-\beta)}$. Using lemma 8, taking β arbitrarily close to $\tilde{r}_H$ we can conclude that,

$$\begin{aligned}M_r(\nu+1)\left(1 - \frac{N+\nu}{N}(r_H + r_L)\right) &\leq d_L(r_L) \\ &\leq M_r(\nu+1)\left(1-(r_H+r_L)\right)\end{aligned} \qquad (39)$$

Comparing this with Theorem 1 we can see that the diversity multiplexing tradeoff for the ISI channel is successively refinable. ■

V. Discussion

In this paper we presented the successive refinement of the diversity multiplexing tradeoff for the SISO ISI fading channel. Moreover we showed that superposition of two uncoded QAM constellations was sufficient to achieve this successive refinement. Although parallel channels are not successively refinable, a set of correlated parallel channels might be refinable. The same result holds for multiple receive and single transmit antenna. It would be interesting to investigate whether a similar result would be true for ISI channels with a single receive and multiple transmit antennas.

VI. Appendix

A. Proof of lemma 3

Proof: The tap coefficients in the frequency domain are given by,

$$\Lambda_k = \sum_{m=0}^{\nu} h_m e^{-\frac{2\pi j}{(N+\nu)}km} \qquad k = \{0,\ldots,(N+\nu-1)\}$$

Defining $\theta = e^{-\frac{2\pi j}{(N+\nu)}}$ the above equation can be rewritten as,

$$\begin{aligned}\Lambda_k &= \sum_{m=0}^{\nu} h_m \theta^{km} \qquad k = \{0,\ldots,(N+\nu-1)\} \qquad (40)\\ &= \begin{bmatrix} 1 & \theta^k & \ldots & \theta^{k\nu}\end{bmatrix}\begin{bmatrix} h_0 & h_1 & \ldots & h_\nu\end{bmatrix}^t\end{aligned}$$

Take any set of $(\nu+1)$ coefficients in the frequency domain and index this set by $\mathcal{K} = \{k_0,\ldots,k_\nu\}$. Define,

$$\breve{\mathbf{\Lambda}} = \begin{bmatrix}\Lambda_{k_0}\\ \Lambda_{k_1}\\ \ldots \\ \Lambda_{k_\nu}\end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \theta^{k_0} & \ldots & \theta^{k_0\nu}\\ 1 & \theta^{k_1} & \ldots & \theta^{k_1\nu}\\ \vdots & & \ldots & \vdots\\ 1 & \theta^{k_\nu} & \ldots & \theta^{k_\nu\nu}\end{bmatrix}}_{\mathbf{V}}\underbrace{\begin{bmatrix}h_0\\ h_1\\ \ldots\\ h_\nu\end{bmatrix}}_{\mathbf{h}} \qquad (41)$$

where $\mathbf{V} \in \mathbb{C}^{(\nu+1)\times(\nu+1)}$ is a full rank Vandermonde matrix. Therefore its inverse exists and we denote it by $\mathbf{V}^{-1} = \mathbf{A}$. Denoting the rows of $\mathbf{A}$ as,

$$\mathbf{A} = \begin{bmatrix}\mathbf{a}^{(0)}\\ \mathbf{a}^{(1)}\\ \ldots\\ \mathbf{a}^{(\nu)}\end{bmatrix} \qquad (42)$$

we conclude that,

$$\mathbf{a}^{(l)}\mathbf{V} = \mathbf{e}^{(l)} \quad \text{and} \quad \mathbf{a}^{(l)} \neq \mathbf{0} \qquad (43)$$

where $\mathbf{e}^{(l)} \in \mathbb{C}^{1\times(\nu+1)}$ is the unit row vector with 1 at the l^{th} position and zero otherwise. Note that the entries of $\{\mathbf{a}^{(l)}\}$ do not depend on SNR. Therefore,

$$\mathbf{a}^{(l)} = \mathbf{e}^{(l)}\mathbf{V}^{-1} \qquad (44)$$

From (41) we have,

$$\mathbf{h} = \mathbf{V}^{-1}\breve{\mathbf{\Lambda}}$$

Multiplying both sides by $\mathbf{e}^{(l)}$ and using (44) we get,

$$\mathbf{e}^{(l)}\mathbf{h} = h_l = \mathbf{e}^{(l)}\mathbf{V}^{-1}\breve{\mathbf{\Lambda}} \stackrel{(a)}{=} \mathbf{a}^{(l)}\breve{\mathbf{\Lambda}} = \sum_{i=0}^{\nu} a_i^{(l)}\Lambda_{k_i}$$

Using the Cauchy-Schwartz inequality[2], we get,

$$|h_l|^2 = |\sum_{i=0}^{\nu} a_i^{(l)}\Lambda_{k_i}|^2 \leq (\sum_{i=0}^{\nu}|a_i^{(l)}|^2)(\sum_{i=0}^{\nu}|\Lambda_{k_i}|^2)$$

Using the fact that N is finite or does not grow with SNR it follows that the $\{a_i^{(l)}\}$ do not depend on SNR. Therefore, the above inequality can be written as

$$|h_l|^2 \dot{\leq} |\Lambda_{k_0}|^2 + |\Lambda_{k_1}|^2 + \ldots + |\Lambda_{k_\nu}|^2 \qquad (45)$$

Note that the above inequality holds for all h_l, $l = 0,\ldots,\nu$. Therefore, we get that for any set of $(\nu+1)$ coefficients in the frequency domain indexed by $\{k_0,\ldots,k_\nu\}$,

$$\max_{l\in\{0,1,\ldots,\nu\}} |h_l|^2 \dot{\leq} |\Lambda_{k_0}|^2 + |\Lambda_{k_1}|^2 + \ldots + |\Lambda_{k_\nu}|^2 \qquad (46)$$

[2] $|\mathbf{u}^*\mathbf{v}| \leq \|u\|.\|v\|$.

From the Cauchy-Schwartz inequality note that,

$$
\begin{aligned}
|\Lambda_{k_0}|^2 + |\Lambda_{k_1}|^2 + \ldots + |\Lambda_{k_\nu}|^2 = |\sum_{m=0}^{\nu} h_m \theta^{k_0 m}|^2 + \ldots \\
+ |\sum_{m=0}^{\nu} h_m \theta^{k_\nu m}|^2 \\
\leq (\sum_{m=0}^{\nu} |h_m|^2) \left(\sum_{m=0}^{\nu} |\theta^{k_0 m}|^2 + \ldots + \sum_{m=0}^{\nu} |\theta^{k_\nu m}|^2 \right) \\
\doteq (|h_0|^2 + |h_1|^2 + \ldots + |h_\nu|^2) \\
\doteq \max_{l \in \{0,1,\ldots,\nu\}} |h_l|^2 \qquad (47)
\end{aligned}
$$

Combining equations (46) and (47) we get,

$$
|\Lambda_{k_0}|^2 + |\Lambda_{k_1}|^2 + \ldots + |\Lambda_{k_\nu}|^2 \doteq \max_{l \in \{0,1,\ldots,\nu\}} |h_l|^2. \qquad (48)
$$

- Given that $\mathbf{h} \in \mathcal{A}^c$ we know that there exists at least one l such that $|h_l|^2 \dot{\geq} SNR^{-(1-r)}$. Therefore, if more than ν taps in the frequency domain are of magnitude less than $SNR^{-(1-r)}$ choose our set $\mathcal{K}$ to be these sets of coefficients. From equation (48) we get a contradiction. Therefore $|\mathcal{F}| \leq \nu$ proving (a).
- We know from (47) that,

$$
|\Lambda_k|^2 \dot{\leq} \max_{l \in \{0,1,\ldots,\nu\}} |h_l|^2 \qquad \forall k
$$

Since, $\mathcal{G} = \{i : |\Lambda_i|^2 \doteq \max_{l \in \{0,1,\ldots,\nu\}} |h_l|^2\}$,

$$
|\Lambda_k|^2 \dot{<} \max_{l \in \{0,1,\ldots,\nu\}} |h_l|^2 \qquad \forall k \in \mathcal{G}^c.
$$

If $|\mathcal{G}^c| > \nu$ then there exists a set $\mathcal{K} = \mathcal{G}^c$ of size at least $\nu + 1$ such that,

$$
|\Lambda_k|^2 \dot{<} \max_{l \in \{0,1,\ldots,\nu\}} |h_l|^2 \qquad \forall k \in \mathcal{K}
$$

But this is a contradiction to equation (48) and therefore we have $|\mathcal{G}^c| \leq \nu$ proving (b).

■

B. Proof of lemma 4

Proof: Consider the case where we have $\nu + 1$ taps, *i.e.*,

$$
y[n] = \sum_{m=0}^{\nu} h_m x[n-m] + z[n]
$$

We receive a vector of length $(N+\nu)$ denoted by $\mathbf{y}$. Denoting the transmitted sequence of length N by $\mathbf{x} \in \mathcal{X}^N$, the ν zero symbols padded at the end by $\mathbf{0}_{\nu \times 1}$ and the circulant channel matrix as $\mathbf{H}$, we have

$$
\mathbf{y} = \mathbf{H} \left[\begin{array}{cc} \mathbf{x} & \mathbf{0}_{\nu \times 1} \end{array} \right]^t + \mathbf{z} \qquad (49)
$$

Similar to analysis [4] we can write the circulant matrix $\mathbf{H} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^*$ where $\mathbf{Q} \in \mathbb{C}^{(N+\nu)\times(N+\nu)}$ and $\mathbf{Q}^* \in \mathbb{C}^{(N+\nu)\times(N+\nu)}$ are truncated DFT matrices and $\mathbf{\Lambda} \in \mathbb{C}^{(N+\nu)\times(N+\nu)}$ is a diagonal matrix with the diagonal elements given by,

$$
\Lambda_k = \sum_{m=0}^{\nu} h_m e^{-\frac{2\pi j}{(N+\nu)} km} \qquad k = \{0, \ldots, (N+\nu-1)\}
$$

and the entries of $\mathbf{Q}$ are given by,

$$
(\mathbf{Q})_{pq} = e^{-\frac{2\pi j}{(N+\nu)} pq} \text{ for } 0 \leq p \leq (N+\nu),\ 0 \leq q \leq (N+\nu)
$$

Note that $\mathbf{Q}$ is a Vandermonde matrix. Multiplying the received vector by $\mathbf{Q}^*$ we get,

$$
\tilde{\mathbf{y}} = \mathbf{Q}^* \mathbf{y} = \mathbf{\Lambda} \mathbf{Q}^* \left[\begin{array}{cc} \mathbf{x} & \mathbf{0}_{\nu \times 1} \end{array} \right]^{\mathbf{t}} + \mathbf{Q}\mathbf{z} = \mathbf{\Lambda} \tilde{\mathbf{Q}}^* \mathbf{x} + \tilde{\mathbf{z}}
$$

where $\tilde{\mathbf{Q}}^* \in \mathbb{C}^{(N+\nu)\times N}$ is a matrix obtained by deleting the last ν columns. Note that $\tilde{\mathbf{Q}}^*$ is also a Vandermonde matrix which implies it has rank N.

Note that from lemma 3 we know that at most ν taps of the available $(N+\nu)$ taps in the frequency domain can be of magnitude $|\Lambda_k|^2 \dot{\leq} SNR^{-(1-r)}$. Define a selection matrix $\mathbf{S} \in \mathbb{C}^{N \times (N+\nu)}$ such that,

$$
\mathbf{S}\mathbf{\Lambda}\tilde{\mathbf{Q}}^* = \hat{\mathbf{\Lambda}}\hat{\mathbf{Q}}
$$

where, $\hat{\mathbf{\Lambda}} \in \mathbb{C}^{N \times N}$, $\hat{\mathbf{\Lambda}} = diag\left(\{\Lambda_l : l \in \mathcal{F}^c\}\right)$. Similarly $\hat{\mathbf{Q}} \in \mathbb{C}^{N \times N}$ is the matrix $\tilde{\mathbf{Q}}$ with the ν rows corresponding to $\{\Lambda_l : l \in \mathcal{F}\}$ deleted. Note that $\hat{\mathbf{Q}}$ is still a full rank (rank N) Vandermonde matrix and denoting the singular values of $\hat{\mathbf{\Lambda}}\hat{\mathbf{Q}}$ by γ_k we have $\gamma_k \dot{>} SNR^{-(1-r)}$. Using this selection matrix we have,

$$
\hat{\mathbf{y}} = \mathbf{S}\tilde{\mathbf{y}} = \hat{\mathbf{\Lambda}}\hat{\mathbf{Q}}\mathbf{x} + \hat{\mathbf{z}} \qquad (50)
$$

Due to the fact that we are using uncoded QAM for transmission, the minimum norm distance between any two elements $\mathbf{x} \neq \mathbf{x}' \in \mathcal{X}^N$ is lower bounded by,

$$
\|\mathbf{x} - \mathbf{x}'\|^2 \dot{\geq} SNR^{(1-r)}.
$$

From the fact that $\hat{\mathbf{Q}}$ is full rank its smallest singular value is nonzero and independent of SNR. Defining $\hat{\mathbf{x}} = \hat{\mathbf{Q}}\mathbf{x}$ we can conclude that,

$$
\|\hat{\mathbf{x}} - \hat{\mathbf{x}}'\|^2 \doteq \|\mathbf{x} - \mathbf{x}'\|^2 \dot{\geq} SNR^{(1-r)} \qquad (51)
$$

As $\hat{\mathbf{\Lambda}}$ is a diagonal matrix,

$$
\begin{aligned}
\|\hat{\mathbf{\Lambda}}(\hat{\mathbf{x}} - \hat{\mathbf{x}}')\|^2 &= \sum_{l=0}^{N-1} |\Lambda_l (\hat{\mathbf{x}} - \hat{\mathbf{x}}')_l|^2 = \sum_{l=0}^{N-1} |\Lambda_l|^2 |(\hat{\mathbf{x}} - \hat{\mathbf{x}}')_l|^2 \\
&\doteq SNR^{-(1-r)+\epsilon} \sum_{l=0}^{N-1} |(\hat{\mathbf{x}} - \hat{\mathbf{x}}')_l|^2 \qquad (52) \\
&= SNR^{-(1-r)+\epsilon} \|(\hat{\mathbf{x}} - \hat{\mathbf{x}}')\|^2 \\
&\dot{\geq} SNR^{-(1-r)+\epsilon} SNR^{(1-r)} = SNR^{\epsilon}
\end{aligned}
$$

where (52) is true from lemma 3 for some $\epsilon > 0$. Since $Q(x)$ is a decreasing function in x, using the above equation, we conclude that the pairwise error probability of detecting the sequence $\mathbf{x}'$ given that $\mathbf{x}$ was transmitted is upper bounded by,

$$P_e(\mathbf{x} \to \mathbf{x}') \leq Q\left(\|\hat{\mathbf{\Lambda}}(\hat{\mathbf{x}} - \hat{\mathbf{x}}')\|^2\right) \dot{\leq} Q\left(SNR^{\epsilon}\right)$$

Therefore, by the union bound we have,

$$P_e(SNR) \dot{\leq} SNR^r Q\left(SNR^{\epsilon}\right) \dot{\leq} SNR^r e^{-\frac{SNR^{2\epsilon}}{2}}$$

as $Q(x)$ decays exponentially in x for large x *i.e.*, $Q(x) \leq e^{-\frac{x^2}{2}}$. Hence it follows that given $\mathbf{h} \in \mathcal{A}^c$ the error probability decays exponentially in SNR. Note that we only use the weaker form of lemma 3 over here *i.e.* we need at least N tap coefficients to be large but we don't need them to be of the same magnitude. ■

C. Proof of lemma 5

Proof: For decoding the higher layer we treat the signal on the lower layer as noise. Proceed as in the previous lemma (equation 50) with the selection matrix $\mathbf{S}$ chosen such that $\hat{\mathbf{\Lambda}} = diag\left(\{\Lambda_l : l \in \mathcal{G}\}\right)$, where $|\mathcal{G}| \geq N$. We get,

$$\begin{aligned}
\hat{\mathbf{y}} = \mathbf{S}\tilde{\mathbf{y}} &= \hat{\mathbf{\Lambda}} \underbrace{\hat{\mathbf{Q}}\mathbf{x_H}}_{\hat{\mathbf{x}}_H} + \hat{\mathbf{\Lambda}} \underbrace{\hat{\mathbf{Q}}\mathbf{x_L}}_{\hat{\mathbf{x}}_L} + \hat{\mathbf{z}} \\
&= \hat{\mathbf{\Lambda}}\hat{\mathbf{x}}_H + \underbrace{\hat{\mathbf{\Lambda}}\hat{\mathbf{x}}_H + \hat{\mathbf{z}}}_{\tilde{\mathbf{z}}} \\
&= \hat{\mathbf{\Lambda}}\hat{\mathbf{x}}_H + \tilde{\mathbf{z}}
\end{aligned}$$

The decoding rule we use to decode $\mathbf{x}_H$ is given by,

$$\tilde{\mathbf{x}}_H = \underset{\mathbf{x}_H}{\operatorname{argmin}} \|\hat{\mathbf{y}} - \hat{\mathbf{\Lambda}}\hat{\mathbf{Q}}\mathbf{x}_H\|^2$$

Therefore, the pairwise error probability of detecting the sequence $\mathbf{x}_{\mathbf{H}}^{'}$ given that $\mathbf{x_H}$ was transmitted is given by,

$$\begin{aligned}
P_e^H(\mathbf{x}_H \to \mathbf{x}_H^{'}) &= \sum_{\mathbf{x}_L \in \mathcal{X}_L^N} Pr(\mathbf{x}_L) P_e(\mathbf{x_H} \to \mathbf{x}_H^{'} | \Lambda, \mathbf{x}_L) \\
&= \sum_{\mathbf{x}_L \in \mathcal{X}_L^N} Pr(\mathbf{x}_L) Pr\left(\|\hat{\mathbf{y}} - \hat{\mathbf{\Lambda}}\hat{\mathbf{x}}_H\|^2 > \|\hat{\mathbf{y}} - \hat{\mathbf{\Lambda}}\hat{\mathbf{x}}_H^{'}\|^2\right) \\
&= \sum_{\mathbf{x}_L \in \mathcal{X}_L^N} Pr(\mathbf{x}_L) Q\left(\|\hat{\mathbf{\Lambda}}(\hat{\mathbf{x}}_H - \hat{\mathbf{x}}_H^{'})\| + \right. \\
&\qquad \left. 2Re\frac{< \hat{\mathbf{\Lambda}}(\hat{\mathbf{x}}_H - \hat{\mathbf{x}}_H^{'}), \hat{\mathbf{\Lambda}}\hat{\mathbf{x}}_L >}{\|\hat{\mathbf{\Lambda}}(\hat{\mathbf{x}}_H - \hat{\mathbf{x}}_H^{'})\|}\right)
\end{aligned} \tag{53}$$

Note that $Q(x)$ is a decreasing function in x. Therefore the equation (53) is upper bounded by,

$$\begin{aligned}
&P_e^H(\mathbf{x}_H \to \mathbf{x}_H^{'}) \leq \\
&\qquad \sum_{\mathbf{x}_L \in \mathcal{X}_L^N} Pr(\mathbf{x}_L) Q\left(\underbrace{\|\hat{\mathbf{\Lambda}}(\hat{\mathbf{x}}_H - \hat{\mathbf{x}}_H^{'})\| - 2\|\hat{\mathbf{\Lambda}}\hat{\mathbf{x}}_L\|}_{\Omega}\right)
\end{aligned} \tag{54}$$

Define Γ_{min} and Γ_{max} as,

$$\Gamma_{min} = \min_{i \in \mathcal{G}} |\Lambda_i|^2, \qquad \Gamma_{max} = \max_{i \in \mathcal{G}} |\Lambda_i|^2$$

Therefore, from lemma 3, we get

$$\Gamma_{min} \doteq \Gamma_{max} \doteq \max_{l \in \{0,1,\ldots,\nu\}} |h_l|^2 \doteq SNR^{-(1-\tilde{r}_H)+2\epsilon}$$

where the last equality follows for some $\epsilon > 0$ from lemma 3 as $\mathbf{h} \in \mathcal{A}_H^c$. Since $\|\hat{\mathbf{x}}_L\|^2 \dot{\leq} SNR^{1-\beta}$ and from equation (51) in the previous lemma, we can lower bound Ω as,

$$\begin{aligned}
\Omega &\geq \Gamma_{min}^{\frac{1}{2}} \|(\hat{\mathbf{x}}_H - \hat{\mathbf{x}}_H^{'})\| - 2\Gamma_{max}^{\frac{1}{2}} \|\hat{\mathbf{x}}_L\| \\
&\doteq SNR^{\frac{-(1-\tilde{r}_H)+2\epsilon}{2}} \left(\|\hat{\mathbf{x}}_H - \hat{\mathbf{x}}_H^{'}\| - \|\hat{\mathbf{x}}_L\|\right) \\
&\doteq SNR^{-\frac{(1-\tilde{r}_H)}{2}+\epsilon} \left(SNR^{\frac{1-\tilde{r}_H}{2}} - SNR^{\frac{1-\beta}{2}}\right) \\
&\doteq SNR^{\epsilon}
\end{aligned}$$

where the last step is valid as $\beta > \tilde{r}_H$ Therefore,

$$P_e^H(\mathbf{x}_H \to \mathbf{x}_H^{'}) \dot{\leq} Q(SNR^{\epsilon})$$

which decays exponentially in SNR. By the union bound as in the previous lemma we conclude that $P_e^H(SNR)$ decays exponentially in SNR.

■

References

[1] S. N. Diggavi, N. Al-Dhahir, and A. R. Calderbank. Diversity embedding in multiple antenna communications, Network Information Theory, pages 285-302. AMS volume 66, Series on Discrete Mathematics and Theoretical Computer Science. Appeared as a part of DIMACS workshop on Network Information Theory, March 2003.

[2] S. N. Diggavi and D. N. C. Tse. Fundamental limits of Diversity-Embedded Codes over Fading Channels IEEE International Symposium on Information Theory (ISIT), Adelaide, 2005.

[3] S. N. Diggavi and D. N. C. Tse. On opportunistic codes and broadcast codes with degraded message sets IEEE Information Theory Workshop (ITW), Uruguay, 2006.

[4] L. Grokop and D. N. C. Tse. Diversity Multiplexing Tradeoff in ISI channels *IEEE International Symposium on Information Theory (ISIT)*, pp 96, Chicago, 2004.

[5] A. Medles and D. T. M. Slock. Optimal diversity vs multiplexing tradeoff for frequency selective MIMO channels IEEE International Symposium on Information Theory (ISIT), Adelaide, 2005.

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