Diversity–Multiplexing Tradeoff in ISI Channels

Leonard H. Grokop, *Member, IEEE*, and David N. C. Tse, *Senior Member, IEEE*

***Abstract*—The optimal diversity–multiplexing tradeoff curve for the intersymbol interference (ISI) channel is computed and various equalizers are analyzed using this performance metric. Maximum-likelihood signal decoding (MLSD) and decision feedback equalization (DFE) equalizers achieve the optimal tradeoff without coding, but zero forcing (ZF) and minimum mean-square-error (MMSE) equalizers do not. However if each transmission block is ended with a period of silence lasting the coherence time of the channel, both ZF and MMSE equalizers become diversity-multiplexing optimal. This suggests that the bulk of the performance gain obtained by replacing linear decoders with computationally intensive ones such as orthogonal frequency-division multiplexing (OFDM) or Viterbi, can be realized in much simpler fashion—with a small modification to the transmit scheme.**

***Index Terms*—Decision feedback equalization (DFE), diversity–multiplexing tradeoff, intersymbol interference (ISI), maximum-likelihood sequence estimation , minimum mean-square-error (MMSE) equalization, zero-forcing (ZF) equalization.**

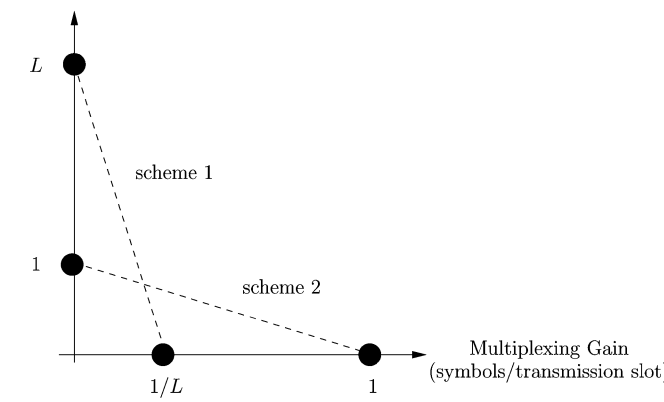

Fig. 1. Two diversity–multiplexing tradeoff curves corresponding to suboptimal communication schemes for the ISI channel (infinite block length). The y-axis plots the error probability exponent. The x-axis plots the data rate in symbols per transmission slot.

I. Introduction

TRADITIONALLY, intersymbol interference (ISI) on wireless channels has been viewed as a hindrance to communication from a complexity perspective. From a diversity perspective though, appropriate communication schemes can exploit the ISI by averaging the fluctuations in the channel gains across the different signal paths, leading to dramatic improvements in system performance. In order to compare such schemes one needs a performance metric. When the coherence time of the channel is short compared to the length of the codewords used to communicate across it (fast fading), the relevant performance metric is ergodic capacity. When this is not the case (slow fading), the error probability cannot be made arbitrarily small. One can still perform coding over the N-length blocks but the appropriate performance measure is now some form of optimal tradeoff between error probability and data rate. It is the slow-fading channel model that we consider in this paper.

We use the following example to illustrate the point. Consider two communication schemes operating over an ISI channel with L taps. Assume quaternary phase-shift keying (QPSK) is used and the receiver knows the channel. In the first scheme, one data symbol is sent every L transmission slots. There is no ISI. From the L replicas received, decode the one with the strongest channel coefficient. In the second scheme, data symbols are sent at the rate of one per transmission slot. A zero-forcing (ZF) equalizer is used at the receivier to remove the ISI.

Let us crudely analyze the tradeoff between error probability and data rate for each scheme. Although it is desirable to consider expressions for each of these quantities for a fixed SNR, for analytical simplicity we consider only how they scale with the SNR in the high-SNR regime. In this qualitative analysis we will imagine increasing the SNR in two different ways, first by handing the benefits to data rate and then by handing the benefits to error probability. This analysis is performed for each of the communication schemes.

Suppose we increase the SNR while keeping the data rate fixed (i.e., the constellation size fixed). Scheme one's error probability will go to zero like SNR^{-L} due to the diversity gain resulting from always selecting the best of L channels. However, scheme two's error probability will only decrease to zero like SNR^{-1}, as the system is equivalent to one with L receive antennas nulling $L-1$ interferers [1]. Now suppose instead we increase the data rate so that the constellation size grows with the SNR. Then, the error probability remains fixed, but the data rate of the first scheme increases only at $1/L$ times the rate of the second.

Characterizing these schemes on the basis of their minimum error probability or maximum data rate makes it difficult to compare their performance and determine which scheme is better. Rather we ask: When the data rate is made to scale with SNR at a certain rate, what error performance does each scheme achieve? One can envision a tradeoff curve between the scaling of the two quantities. The two scenarios analyzed above constitute endpoints of the tradeoff curve specific to the signaling schemes used. See Fig. 1. Note the analysis did not characterize the complete tradeoff curve, *only the endpoints*.

Manuscript received August 19, 2007; revised August 20, 2008. Current version published December 24, 2008. This work was supported by Qualcomm Inc. with matching funds from the California MICRO program, and by the National Science Foundation under Grant CCR-01-18784.

The authors are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: lgrokop@eecs.berkeley.edu; dtse@eecs.berkeley.edu).

Communicated by P. Viswanath, Associated Editor for Communications.
Digital Object Identifier 10.1109/TIT.2008.2008120

Some questions that come to mind are: what is the best tradeoff any scheme can achieve? What is the impact on the tradeoff curve if the decoder uses a suboptimal equalizer? To answer these questions we adopt the framework developed by Zheng and Tse [1].

In Section II, we specify the system model and define diversity gain and multiplexing gain. Our results are presented in Section III. These consist of a characterization of the optimal diversity–multiplexing tradeoff curve, and the diversity–multiplexing tradeoffs achieved by a variety of equalizers commonly used in practice. Proofs of these results are given in Section IV.

II. Problem Formulation and System Model

A. Setup

We consider a wireless point-to-point multipath channel with Rayleigh fading. The L fading coefficients are denoted by the vector $\boldsymbol{h} = [h_0, \ldots, h_{L-1}]^T$ and are independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$. The channel input/output model is expressed as

$$y[n] = \sum_{l=0}^{L-1} h_l x[n-l] + w[n] \tag{1}$$

for $n = 0, 1, \cdots$. Transmission begins at $n = 0$ before which time the input to the channel is zero. The additive noise $w[n]$ is i.i.d. circularly symmetric Gaussian, $w[n] \sim \mathcal{CN}(0, N_0)$. Throughout this paper we will assume, without loss of generality, that $N_0 = 1$ so that the transmit power is SNR. The channel realization is constant for all time n (slow-fading) and known to the receiver only (receiver channel side information (CSI)).

B. Tradeoff Curve

The following definitions are used to formulate the tradeoff between diversity gain and multiplexing gain. These are taken from [2]. By a *scheme* we refer to a family of codes $\mathcal{C}(\mathsf{SNR})$ of block length N, one at each SNR level. The data rate scales like $R(\mathsf{SNR}) \approx r \log_2 \mathsf{SNR}$ (bits per symbol), the error probability scales like $P_e(\mathsf{SNR}) \approx \mathsf{SNR}^{-d}$. More formally we have the following.

Definition II.1: A scheme $\mathcal{C}(\mathsf{SNR})$ achieves spatial multiplexing gain r and diversity gain d if the data rate and the average error probability satisfy respectively

$$\lim_{\mathsf{SNR}\to\infty} \frac{R(\mathsf{SNR})}{\log \mathsf{SNR}} = r \quad \text{and}$$
$$\lim_{\mathsf{SNR}\to\infty} \frac{\log P_e(\mathsf{SNR})}{\log \mathsf{SNR}} = -d.$$

For each r define $d^*(r)$ to be the supremum of the diversity gain achieved over all schemes. Also define $d^*_{\max} \triangleq d^*(0)$ which is the maximal diversity gain in the channel.

The curve $d^*(r)$ is termed the *optimal diversity–multiplexing tradeoff curve* and represents the maximum diversity gain any scheme can achieve with a simultaneous spatial multiplexing gain of r.

In this paper, we will use the notation $\doteq$ to denote asymptotic equality in the large SNR limit, that is, $X \doteq Y$ is equivalent to

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log X}{\log \mathsf{SNR}} = \lim_{\mathsf{SNR}\to\infty} \frac{\log Y}{\log \mathsf{SNR}}.$$

As an example of Definition II.1 consider the ISI channel in (1) with $L = 1$. To achieve a multiplexing gain of r, at each time slot we transmit one symbol from a quadrature amplitude modulation (QAM) constellation containing SNR^r points. In this way, the data rate is set to $R = r \log \mathsf{SNR}$. The distance between constellation points at the transmitter is $\doteq \mathsf{SNR}^{(1-r)/2}$ but at the receiver this distance is affected by the channel realization and becomes $\doteq |h_0|\mathsf{SNR}^{(1-r)/2}$. So an error will occur if the noise takes on a value $\doteq |h_0|\mathsf{SNR}^{(1-r)/2}$. As the noise has unit variance the probability of this occurring is $\doteq 0$ if $|h_0|\mathsf{SNR}^{(1-r)/2} \dot{>} 1$ and $\doteq 1$ if $|h_0|\mathsf{SNR}^{(1-r)/2} \dot{\leq} 1$. Thus, the error probability (P_e) is asymptotically equal to the probability of $\{|h_0|^2 \dot{\leq} \mathsf{SNR}^{r-1}\}$ occurring. Now $|h_0|^2$ has an exponential distribution so this event occurs with probability $\doteq \mathsf{SNR}^{r-1}$. Hence $P_e \doteq \mathsf{SNR}^{r-1}$ and the diversity–multiplexing tradeoff for this scheme is $d(r) = 1 - r$. This turns out to be the optimal diversity–multiplexing tradeoff for $L = 1$ and we write $d^*(r) = 1 - r$. Returning to Fig. 1 we see this tradeoff curve corresponds to the dashed line joining the points $(0,1)$ and $(1,0)$.

C. Encoding Schemes

For each equalizer we analyze two different transmission schemes.

Trailing zeros transmission scheme.
For the first $N - L + 1$ time slots of each N-length block the transmitter sends one symbol from a QAM constellation containing M' points, where

$$M' = \mathsf{SNR}^{r'}$$

with

$$r' = \frac{r}{1 - \frac{L-1}{N}}. \tag{2}$$

For simplicity, we will assume M' is a perfect square so that the QAM constellation is well formed. This is minor point since for $r > 0$ and large SNR, M' is very large. During the last $L - 1$ time slots the transmitter sends nothing—in other words, the zero symbol is sent at times $\{N-L+1, N-L+2, \ldots, N-1\}$, $\{2N - L + 1, 2N - L + 2, \ldots, 2N - 1\}$, etc. ... Thus, the transmission rate is

$$\begin{aligned}\left(1 - \frac{L-1}{N}\right)\log_2 M' &= \left(1 - \frac{L-1}{N}\right) r' \log_2 \mathsf{SNR} \\ &= r \log_2 \mathsf{SNR} \\ &= R.\end{aligned}$$

Decoding for this scheme is performed at the end of each N-length block based only on the associated receive vector. The delay is thus different for each symbol, ranging from L to N. Without loss of generality, we concentrate on the first block only. The input/output model is neatly represented in matrix form as

$$\boldsymbol{y} = \boldsymbol{H}x + \boldsymbol{w} \tag{3}$$

where $\boldsymbol{y} \in \mathbb{C}^N, \boldsymbol{x} \in \mathbb{C}^{N-L+1}, \boldsymbol{w} \in \mathbb{C}^N$, and $\boldsymbol{H} \in \mathbb{C}^{N\times(N-L+1)}$ like so

$$\boldsymbol{H} = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ \vdots & h_0 & \ddots & \vdots \\ h_{L-1} & \vdots & \ddots & \vdots \\ 0 & h_{L-1} & \ddots & h_0 \\ \vdots & 0 & \ddots & \vdots \\ \vdots & \vdots & \cdots & h_{L-1} \end{bmatrix}. \tag{4}$$

No trailing zeros transmission
In this scheme symbols are transmitted continuously from a QAM constellation of $M = \mathsf{SNR}^r$ points and decoded with infinite delay based on the entire sequence of observations $y[0], y[1], \cdots$.

The reason for drawing a distinction between these two schemes will become evident later on.

D. Decoding Schemes

In this subsection, we briefly review the various equalizers that will later be analyzed. See [3, pp. 468–477]. Denote the constellation used by $\mathcal{X}$.

1) Maximum-Likelihood Signal Decoding (MLSD) Equalizer: **Trailing zeros**
For decoding symbol $x[k]$ of a trailing zeros transmission the MLSD equalizer selects the $\hat{x}[k] \in \mathcal{X}$ with the highest likelihood of transmission given the received vector

$$\hat{x}[k] = \arg \max_{x[k]\in\mathcal{X}} \Pr(x[k]|\boldsymbol{y}). \tag{5}$$

No trailing zeros
In the case of no trailing zeros transmission, the MLSD equalizer implements the Viterbi algorithm [4]. For decoding symbol $x[k] \in \mathcal{X}$, this algorithm outputs

$$\hat{x}[k] = \arg \max_{x[k]\in\mathcal{X}} \Pr(x[k]|y[0], y[1], \cdots). \tag{6}$$

2) ZF Equalizer: **Trailing zeros**
For trailing zeros transmission ZF equalization removes the ISI by multiplying the received vector by the channel inverse. It then selects the nearest constellation point. The ZF filter is

$$\boldsymbol{G} = \boldsymbol{H}(\boldsymbol{H}^*\boldsymbol{H})^{-1}. \tag{7}$$

By $\boldsymbol{g}_k$ we denote the kth column of $\boldsymbol{G}$, that is

$$\boldsymbol{G} = [\boldsymbol{g}_0, \ldots, \boldsymbol{g}_{N-L+1}]. \tag{8}$$

The *filtered estimate* is

$$\hat{y}[k] = \boldsymbol{g}_k^*\boldsymbol{y} \tag{9}$$

and the decoder outputs

$$\hat{x}[k] = \arg \min_{x[k]\in\mathcal{X}} ||x[k] - \hat{y}[k]||. \tag{10}$$

No trailing zeros
In the case of no trailing zeros transmission the ZF equalizer passes the input sequence through an infine impulse response (IIR) filter before making a decision. Denote the transfer function of the channel

$$H(z) \triangleq \sum_{l=0}^{L-1} h_l z^{-l}.$$

The transfer function of the ZF filter is

$$G(z) = \frac{1}{H(z)}. \tag{11}$$

Denote the filter coefficients g_m such that

$$G(z) \triangleq \sum_{m=-\infty}^{\infty} g_m z^{-m}.$$

Then the decoder output is given by (10) with

$$\hat{y}[k] = \sum_{m=-\infty}^{\infty} g_m y[k-m]. \tag{12}$$

3) Minimum Mean-Square-Error (MMSE) Equalizer: **Trailing zeros**
For trailing zeros transmission the MMSE filter is

$$\boldsymbol{G} = \mathsf{SNR} \cdot \boldsymbol{H}(\boldsymbol{I} + \mathsf{SNR} \cdot \boldsymbol{H}^*\boldsymbol{H})^{-1}. \tag{13}$$

The output is then given by (10) with the filtered estimate of (9).
No trailing zeros
In the case of no trailing zeros transmission the MMSE filter is (see [3, pp. 472])

$$G(z) = \frac{\mathsf{SNR} \cdot H^*(z^*)}{1 + \mathsf{SNR} \cdot |H(z)|^2}. \tag{14}$$

The output is then given by (10) with the filtered estimate of (12).

4) DFE-ZF Equalizer: This is the minimum-phase ZF decision feedback equalizer (DFE). We only analyze it in the no trailing zeros case. Factorize $H(z)$ as

$$H(z) = A_0 \cdot H_{\min}(z) H_{\min}^*(z^*)$$

where $H_{\min}(z)$ and $H_{\max}(z)$ are the minimum and maximum phase components of $H(z)$, respectively, normalized so that the coefficient of z^0 in each is 1, and A_0 is the normalization factor. The precursor equalizer is then

$$G^{\rm pre}(z) = \frac{H_{\max}^*(z^*)}{A_0 H_{\max}(z)}$$

and the causal post-cursor equalizer

$$G^{\rm post}(z) = H_{\min}(z) H_{\max}^*(z) - 1.$$

The filtered estimate is then

$$\hat{y}[k] = \sum_{m=-\infty}^{\infty} g_m^{\rm pre} y[k-m] - \sum_{m=1}^{\infty} g_m^{\rm post} \hat{x}[k-m]. \tag{15}$$

where $g_m^{\rm pre}$ and $g_m^{\rm post}$ are the filter taps corresponding to $G^{\rm pre}(z)$ and $G^{\rm post}(z)$, respectively. The output is given by (10).

III. Results

In point form the main results are as follows.

- The optimal diversity–multiplexing tradeoff curve for the ISI channel.
- The optimal tradeoff can be achieved by sending independent QAM symbols with $L-1$ zeros attached to the end of each transmitted block and using a Viterbi decoder.
- The same transmission scheme achieves the optimal tradeoff curve using only ZF or MMSE equalization.
- If we do not attach $L-1$ trailing zeros, the optimal tradeoff curve is still achieved by the Viterbi equalizer, but no longer achieved by ZF or MMSE equalization.
- The optimal tradeoff can be achieved using QAM and decision-feedback equalization (minimum-phase ZF), with or without the $L-1$ trailing zeros.

A. Optimal Tradeoff Curve

We first review the concept of outage probability as it forms the basis for converse results.

1) Outage Probability: Roughly speaking, the outage probability represents the probability that the channel fades below the level required for reliable communication. An outage is defined as the event that the mutual information does not support a target data rate

$$\{h : I(x[k]; y[k]|\boldsymbol{h} = h) < R\}. \tag{16}$$

The mutual information is a function of the input distribution $P(x[k])$, and the channel realization. Without loss of optimality, the input distribution can be taken to be Gaussian with variance SNR in which case

$$P_{\text{out}}(R) = \inf_{\alpha(\omega)\geq 0:\frac{1}{2\pi}\int_{-\pi}^{\pi}\alpha(\omega)d\omega=1} \Pr\left(\frac{1}{2\pi}\int_0^{2\pi}\log(1+\mathsf{SNR}\alpha(\omega)|H(\omega)|^2)d\omega < R\right) \tag{17}$$

where

$$H(\omega) = \sum_{l=0}^{L-1} h_l e^{j\omega l}. \tag{18}$$

Intuitively, we expect the outage probability to lower-bound the error probability as when the channel is "in outage" errors will occur frequently. This intuition can be made rigorous in the following asymptotic sense.

Lemma III.1: The error probability satisfies

$$P_e \dot{\geq} P_{\text{out}}. \tag{19}$$

For a proof of this result see [2].

2) Optimal Tradeoff: To characterize the optimal tradeoff curve we use the outage probability to lower-bound the error probability, and illustrate (in the next section) a communication scheme that achieves the same SNR exponent.

In fact, we need not compute the outage probability precisely, but only bound it from below. An easy way of doing this is via the matched-filter bound. Here we assume the receiver has individual access to each of the L multipaths. See Fig. 2. As in the decoding process it is free to combine these L paths to recreate the true signal, this assumption generates an outer bound on performance.

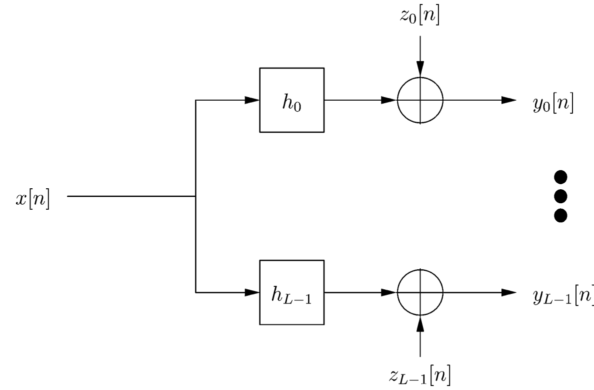

Fig. 2. Matched filter bound.

Theorem III.2: The outage probability P_{out} of the ISI channel satisfies

$$P_{\text{out}} \dot{\geq} \mathsf{SNR}^{-L(1-r)}.$$

With individual access to each path, full diversity can be attained without compromising rate. It is therefore not surprising that the matched filter bound yields the full tradeoff curve—one equal to L times the optimal tradeoff curve for a single-antenna system. What is surprising though is that the match-filter bound is tight, and can be achieved without coding.

Theorem III.3: The optimal diversity–multiplexing tradeoff curve for the ISI channel is

$$d^*(r) = L(1-r).$$

This curve is plotted as the benchmark in both Figs. 4 and 5. One would expect the ISI to cause a significant degradation in system performance, at least over some range of r. Theorem III.3 says that despite receiving the multipaths in a conglomeration, the error probability can be made to decay with SNR as quickly as if they were received in isolation, *at every multiplexing gain.*

B. Equalizers

1) MLSD:

Theorem III.4. (Trailing Zeros): The diversity–multiplexing tradeoff achieved with the trailing zeros transmission scheme and an ML decoder (given by (5)) is

$$d(r) = L(1-r') \tag{20}$$

where r' is given by (2).

See Fig. 4. In the limit $N \to \infty$, $r' \to r$ and this scheme achieves the outage bound and is hence optimal in terms of diversity–multiplexing tradeoff. For finite N, at every diversity gain, the multiplexing gain is reduced by a factor of $N/N-L+1$, which is precisely the ratio between the total block length and the transmission length. This result holds in the case of transmission with any $L-1$ trailing symbols, so long as they are known to the receiver (thus conveying no information).

For MLSD, trailing zeros do not improve large block length performance. That is, we have the following.

Theorem III.5. (No Trailing Zeros): The diversity–multiplexing tradeoff achieved using the no trailing zeros transmission scheme and an infinite delay MLSD equalizer (given by (6) is

$$d(r) = L(1-r).$$

See Fig. 5.

2) ZF:

Theorem III.6. (Trailing Zeros): The diversity–multiplexing tradeoff achieved using the trailing zeros transmission scheme and a ZF equalizer (given by (7), (9), and (10)) is

$$d(r) = L(1-r').$$

See Fig. 4. Surprisingly with $L-1$ trailing zeros, the ZF equalizer achieves the same tradeoff curve as the MLSD equalizer. For continuous transmission this is not the case.

Theorem III.7. (No Trailing Zeros): For the no trailing zeros transmission scheme with infinite delay ZF equalization at the receiver (given by (11), (12), and (10)) is

$$d(r) \leq 1 - r.$$

This inequality is strict for at least $L = 2$.

See Fig. 5. To understand why continuous transmission performs poorly consider the $L = 2$ case. The receiver observes (21). The root of the problem is the following: for a given channel realization there is only one direction in the infinite-dimensional receive space in which the received sequence $\{y[0], y[1], \cdots\}$ can be projected in order to cancel the interference, i.e., the received sequence is passed through the channel inverting filter $1/(h_0 + h_1 z^{-1})$. This filter has two possible implementations.

If $\underline{|h_0| > |h_1|}$

$$\frac{1}{h_0 + h_1 z^{-1}} = \frac{1}{h_0(1 + h_1/h_0 z^{-1})} = \frac{1}{h_0}\left(1 - \frac{h_1}{h_0}z^{-1} + \frac{h_1}{h_0}z^{-2} - \cdots\right)$$

and the interference-free direction is given by the sequence of filter taps

$$g_m = \frac{(-1)^m}{h_0}\left(\frac{h_1}{h_0}\right)^m, \quad \text{for } m = 0, 1, 2, \cdots$$

with $g_m = 0$ for $m = -1, -2, \cdots$, i.e., the equalizer retrieves $x[k]$ from $y[k]$ and nulls $x[k-1]$ using $\{y[k-1], y[k-2], \cdots\}$.

If $\underline{|h_0| < |h_1|}$

$$\frac{1}{h_0 + h_1 z^{-1}} = \frac{z}{h_1(1 + h_0/h_1 z)} = \frac{z}{h_1}\left(1 - \frac{h_0}{h_1}z + \frac{h_0}{h_1}z^2 - \cdots\right)$$

and the interference-free direction is given by the sequence of filter taps

$$g_m = \frac{(-1)^{m+1}}{h_0}\left(\frac{h_1}{h_0}\right)^m, \quad \text{for } m = -1, -2, \cdots$$

with $g_m = 0$ for $m = 0, 1, 2, \cdots$, i.e., the equalizer retrieves $x[k]$ from $y[k+1]$ and nulls $x[k+1]$ using $\{y[k+2], y[k+3], \cdots\}$. Note that in this case the equalizer incurs infinite delay. Thus, an outage occurs if $|h_0| \approx |h_1|$ as the noise will be amplified immensely. More precisely, the effective signal-to-noise ratio can be shown to be

$$\mathsf{SNR}_{\text{eff}} = ||h_0|^2 - |h_1|^2|\mathsf{SNR}$$

and thus an error occurs if $||h_0|^2 - |h_1|^2|$ is small—of the order SNR^{r-1} (see Lemma VII.6). This is an order-one event (see Lemma VII.4), so the error probability is proportional to SNR^{r-1}. The same reasoning applies for arbitrary L. An error occurs if the magnitude of one of the poles in (11) approaches unity. This is an order-one event and hence the maximal diversity gain of the infinite-length ZF equalizer is 1 and $d(r) \leq 1 - r$.

The insertion of $L-1$ zeros into the transmission stream at regular intervals changes the situation. The receiver now observes (22) (see the bottom of the page). The interference spans an $N-L$-dimensional subspace of the N-dimensional space in

$$\begin{bmatrix} \vdots \\ y[k] \\ y[k+1] \\ \vdots \\ \vdots \end{bmatrix} = \cdots + \begin{bmatrix} \vdots \\ h[0] \\ h[1] \\ \vdots \\ \vdots \end{bmatrix} x[k] + \begin{bmatrix} \vdots \\ \vdots \\ h[0] \\ h[1] \\ \vdots \end{bmatrix} x[k+1] + \cdots + \begin{bmatrix} \vdots \\ w[k] \\ w[k+1] \\ \vdots \\ \vdots \end{bmatrix}. \tag{21}$$

$$\begin{bmatrix} y[0] \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ y[N-1] \end{bmatrix} = \begin{bmatrix} h[0] \\ \vdots \\ h[L-1] \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} x[0] + \begin{bmatrix} \vdots \\ h[0] \\ \vdots \\ h[L-1] \\ \vdots \\ \vdots \end{bmatrix} x[1] + \cdots + \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ h[0] \\ \vdots \\ h[L-1] \end{bmatrix} x[N-L+1] + \begin{bmatrix} w[0] \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ w[N-1] \end{bmatrix}. \tag{22}$$

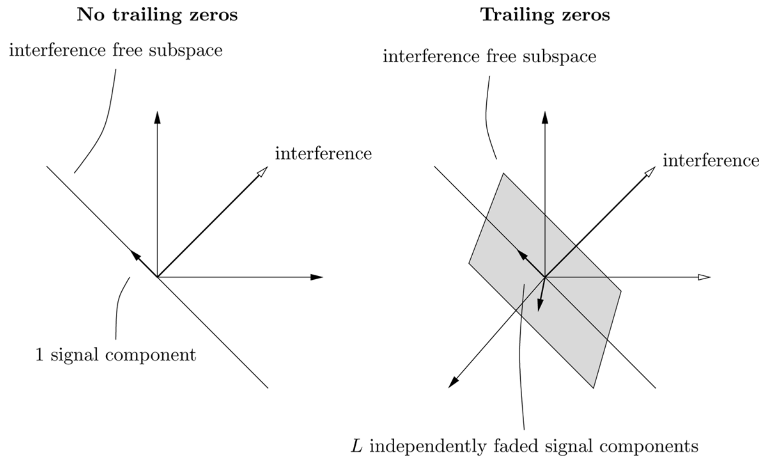

Fig. 3. Adding $L-1$ trailing zeros to the end of each transmitted block increases the number of interference free dimensions from one to L.

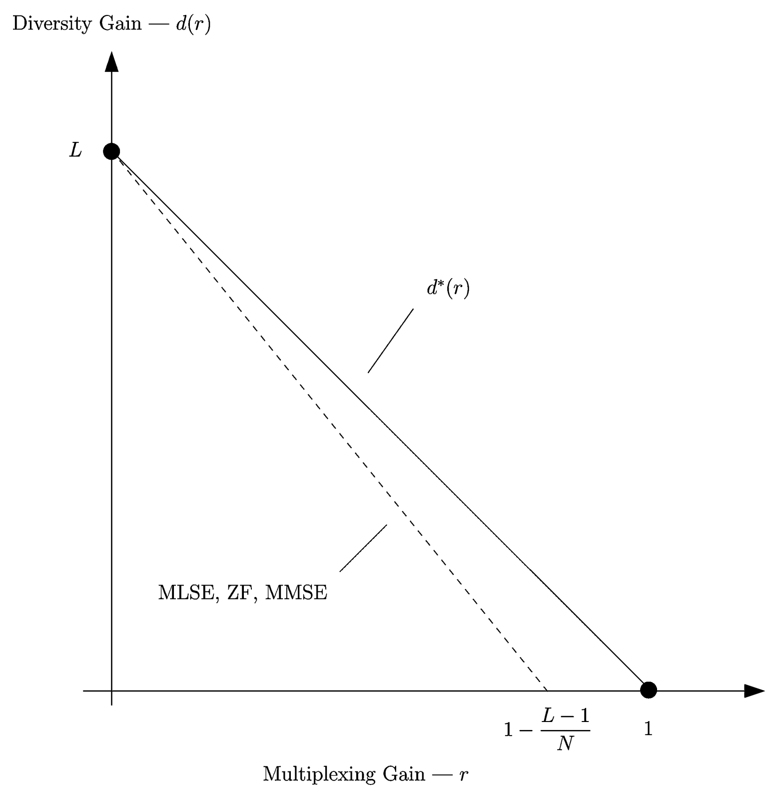

Fig. 4. Optimal tradeoff curve, and those achieved by various equalizers using QAM transmission with the last $L-1$ zeros of each N length block set to zero.

which $x[k]$ resides. Thus, the interference-free subspace spans L dimensions. The receiver can project the received vector $\boldsymbol{y}$ onto L orthogonal directions, yielding L independent observations. Match filter combining of these yields a filtered sequence with a maximal diversity gain of L, and (in the limit $N \to \infty$) we will have $d(r) = L(1-r)$. See Fig. 3.

3) MMSE:

Theorem III.8. (Trailing Zeros): The diversity–multiplexing tradeoff achieved using the trailing zeros transmission scheme and an MMSE equalizer (given by (9), (10), and (13)) is

$$d(r) = L(1-r').$$

See Fig. 4. Again it is somewhat surprising that with $L-1$ trailing zeros, this equalizer achieves the same tradeoff curve as the MLSD equalizer. Like the ZF equalizer, without these zeros, its performance is distinctly worse.

Theorem III.9. (No Trailing Zeros): For the no trailing zeros transmission scheme with infinite delay MMSE equalization at the receiver (given by (12), (10), and (14))

$$d(r) \le \min\left\{1-r, L\left(1-\frac{1}{\frac{L}{(L-1)(L+1)}}r\right), \min_{k\in\{1,\dots,L-2\}}(k+1/2)\left(1-\frac{1}{\frac{k+1/2}{k(k+1)}}r\right)\right\}.$$

This L-part curve is plotted in Fig. 5.

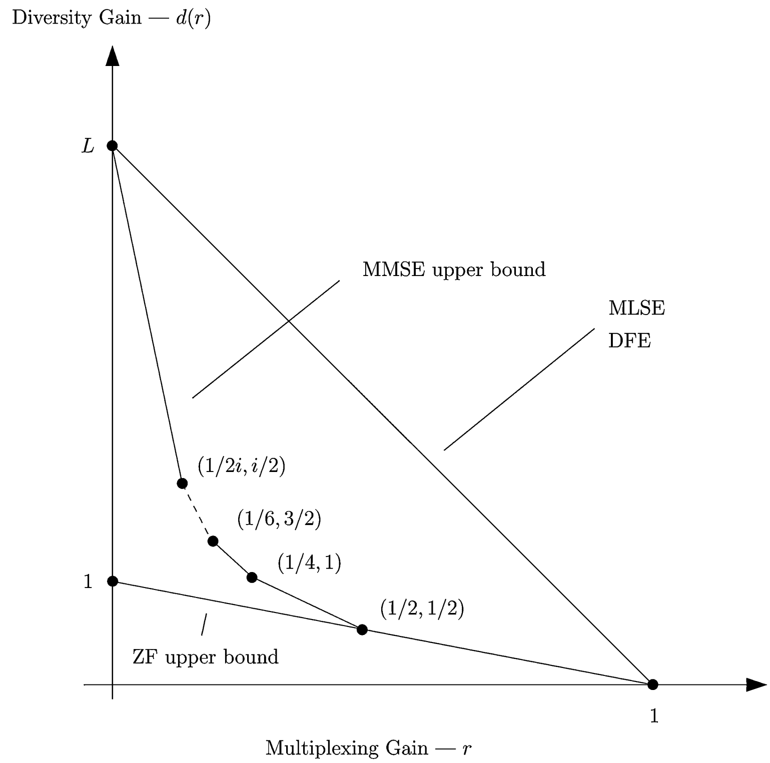

Fig. 5. Optimal tradeoff curve, and those achieved by various equalizers using continuous QAM transmission.

4) DFE-ZF:

Theorem III.10: For the no trailing zeros transmission scheme with minimum-phase infinite-length DFE-ZF equalization at the receiver (given in Section II-D4) the error probability satisfies

$$d(r) = L(1-r).$$

See Fig. 5. To understand why decision feedback leads to a vast improvement in the tradeoff curve consider again the $L=2$ case. Assume $x[k-1]$ has been correctly decoded. The filter has two possible implementations.

If $\underline{|h_0| > |h_1|}$

$$A_0 = h_0, \quad H_{\min} = 1 + h_1/h_0 z^{-1}, \quad \text{and} \quad H_{\max} = 1.$$

The precursor equalizer is simply $1/h_0$ and the postcursor equalizer is $h_1/h_0 z^{-1}$. Thus, the DFE obtains $x[k]$ from $y[k]$ by just dividing through by h_0 and subtracting off the residual $h_1/h_0 x[k-1]$ term.

If $\underline{|h_0| < |h_1|}$

$$A_0 = h_1 z^{-1}, \quad H_{\min} = 1, \quad \text{and} \quad H_{\max} = 1 + h_0/h_1 z.$$

The precursor equalizer is now

$$G^{\text{pre}}(z) = \frac{1 + h_0^*/h_1^* z^{-1}}{h_1 z^{-1}(1 + h_0/h_1 z)}.$$

We can express it as a product of an all-pass filter and a scaling constant filter

$$G^{\text{pre}}(z) = \frac{h_1^* + h_0^* z^{-1}}{h_0 + h_1 z^{-1}} \cdot \frac{1}{h_1^*}.$$

A little thought reveals that the all-pass filter merely swaps the channel coefficients and conjugates them so that at the input to the scaling constant filter the channel appears to be $H_{\text{all}}(z) = h_1^* + h_0^* z^{-1}$ rather than $H(z) = h_0 + h_1 z^{-1}$. As the magnitude response of the filter is flat, the noise variance remains the same.[1] The post-cursor equalizer is $h_0^*/h_1^* z^{-1}$. Thus, for $|h_0| < |h_1|$ the DFE performs an almost identical operation to the one it performs for $|h_0| > |h_1|$ (but incurring infinite delay). If we denote the output of the all-pass filter by $y_{\text{all}}[k] = h_1^* x[k] + h_0^* x[k-1] + w_{\text{all}}[k]$ then $x[k]$ is obtained from $y_{\text{all}}[k]$ by just dividing through by h_1^* and subtracting off the residual $h_0^*/h_1^* x[k-1]$ term.

In comparison to the ZF equalizer, which makes an error whenever $|h_0| \approx |h_1|$—an order-one event, the DFE-ZF only makes an error if both h_0 and h_1 are small—of the order SNR^{r-1}. This is an order-two event so the error probability is proportional to $\mathsf{SNR}^{2(r-1)}$. The same reasoning applies for arbitrary L. Assume $x[k-1], \ldots, x[k-L+1]$ were decoded correctly. If the channel is minimum phase then given

$$y[k] = h_0 x[k] + \cdots + h_{L-1} x[k-L+1]$$

$x[k]$ is obtained by diving through by h_0 and subtracting off the interference terms. If not, the channel is first converted to minimum phase using an appropriate all-pass filter and a similar decoding operation is then performed. For an error to occur, all L channel coefficients must fade to the order of SNR^{r-1}. This is an order-L event and consequently the error probability is proportional to $\mathsf{SNR}^{L(r-1)}$.

[1]More generally, the all-pass filter converts the maximum-phase channel to a minimum-phase one.

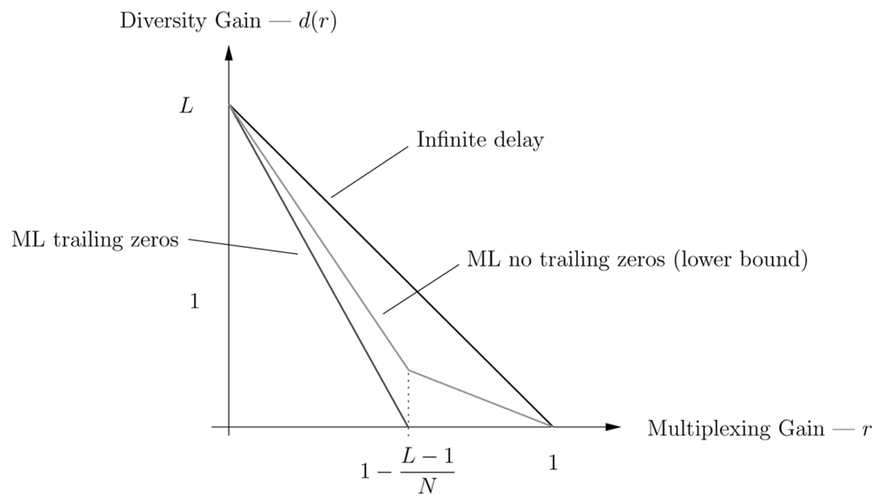

Fig. 6. Tradeoff curves for finite-delay MLSD equalization with and without trailing zeros, versus infinite delay MLSD equalization $(L = 2)$.

C. Optimal Tradeoff for Finite Delay

One of the conclusions of the previous subsection was that trailing zeros communication achieves the outage bound in the limit of infinite delay. It is natural to wonder whether trailing zeros communication is also optimal for finite delay. The answer is no. Before elaborating upon this we give a definition of decoding delay.

Definition III.11: A communication scheme for the ISI channel incurs a maximum decoding delay of N time slots if for all $n \geq 0$, $\hat{x}[n]$ can be expressed as a function of $y[0], \ldots, y[n+N-1]$.

From here on we will use the symbol N to denote decoding delay as defined above. This definition is consistent with the N defined for the trailing zeros scheme in Section II-C.

Returning to the question of optimality of the trailing zeros scheme for finite delay, consider the following successive cancellation strategy. Transmit symbols continuously starting at time $n = 0$, i.e., use the no trailing zeros transmission scheme. Decode $x[0]$ from $y[0]$ and remove the ISI by subtracting $h_1\hat{x}[0]$ from $y[1]$. Then decode $x[1]$ and subtract $h_1\hat{x}[1]$ from $y[2]$, etc., $\ldots$. This is a full rate, diversity 1 scheme. The tradeoff achieved is

$$d(r) = 1 - r. \tag{23}$$

This can be seen by induction. The probability of decoding $x[0]$ incorrectly is

$$\Pr(\hat{x}[0] \neq x[0]) \doteq \mathsf{SNR}^{-(1-r)}.$$

If we assume the probability of decoding $x[n-1]$ incorrectly is $\doteq \mathsf{SNR}^{-(1-r)}$ then the probability of decoding $x[n]$ incorrectly is

$$\begin{aligned}
&\Pr(\hat{x}[n] \neq x[n]) \\
&\quad\doteq \Pr(\hat{x}[n] \neq x[n] | \hat{x}[n-1] = x[n-1]) \\
&\qquad\times \Pr(\hat{x}[n-1] = x[n-1]) \\
&\qquad+ \Pr(\hat{x}[n] \neq x[n] | \hat{x}[n-1] \neq x[n-1]) \\
&\qquad\times \Pr(\hat{x}[n-1] \neq x[n-1]) \\
&\quad\doteq \mathsf{SNR}^{-(1-r)}
\end{aligned}$$

for $n \geq 1$. □

Comparing the tradeoff achieved by this scheme (see (23)) to that achieved by trailing zeros transmission (see (20)) reveals the trailing zeros tradeoff is suboptimal *at least* for $r \in (1 - \frac{L}{N+1}, 1]$. From this successive cancellation example we see that the performance loss of the trailing zeros scheme comes from wasting degrees of freedom, $L-1$ out of every N are not used. For $L = 2$ we are able to show something stronger.

Theorem III.12: For $L = 2$ no trailing zeros transmission and MLSD equalization achieves

$$d(r) \geq \max\left\{1 - r, 2\left(1 - \frac{r}{1 - \frac{1}{2N-1}}\right)\right\}. \tag{24}$$

This curve is plotted in Fig. 6. It illustrates that trailing zeros transmission is strictly suboptimal *for all* r (for at least $L = 2$). More generally, any scheme using a prefix of length $L - 1$ is strictly suboptimal in terms of diversity-multiplexing tradeoff, though for $N \gg L$ the loss is negligible. Orthogonal frequency-division multiplexing (OFDM) is an example of such a scheme and is hence suboptimal for finite N. The question of whether or not the outage bound can be achieved with a finite delay decoder remains open.

IV. Proofs

A. Optimal Tradeoff Curve

Proof of Theorem III.2: We use a matched filter bound. Suppose the receiver has individual access to the L multipaths via the following receiver vector:

$$\boldsymbol{y}^{\mathrm{MF}}[n] = \boldsymbol{h}x[n] + \boldsymbol{w}^{\mathrm{MF}}[n] \tag{25}$$

for all n. Here the additive white Gaussian noise (AWGN) vectors $\boldsymbol{z}[n] \sim \mathcal{CN}(0, \boldsymbol{I}/L)$ are independent and identically distributed (i.i.d.) across n. The receiver can recover the original $y[n]$ via

$$y[n] = \sum_{k=0}^{L-1} y_k^{\mathrm{MF}}[n-k].$$

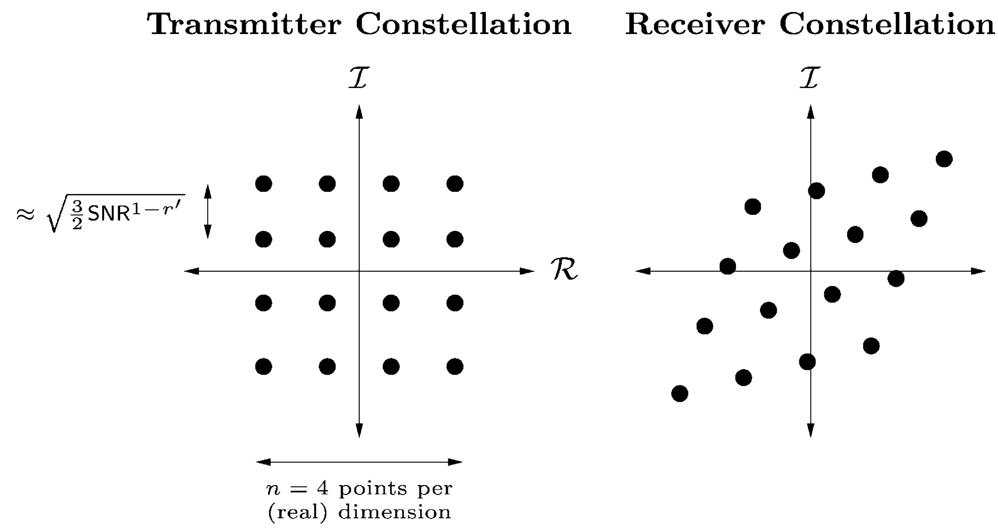

Fig. 7. Transmitter QAM constellation with $\mathsf{SNR}^{r'}$ codewords. Receiver constellation is a linearly transformed version of the $2N$-dimensional transmitter constellation. The example above is for $N = 1$ but can be abstracted to higher dimensions corresponding to $N > 1$.

which matches (1). We now use (16) to compute the outage probability for the system of (25)

$$\begin{aligned} P_{\text{out}}(R) &= \inf_{p_X} \Pr\{h : I(X;Y|\boldsymbol{h} = h) < R\}. \\ &= \Pr\{\log(1 + \|\boldsymbol{h}\|^2 \mathsf{SNR}) < R\}. \\ &\doteq \Pr\{\|\boldsymbol{h}\|^2 < \mathsf{SNR}^{r-1}\}. \\ &\doteq \mathsf{SNR}^{-L(1-r)} \end{aligned} \tag{26}$$

for $0 < r < 1$, where the second step follows from the fact that Gaussian inputs maximize the mutual information over AWGN channels and the last step follows by Lemma VII.2. □

B. MLSD Equalizer

Proof of Theorem III.4: The two main elements of this proof are the use of geometric arguments to construct an expression for the error probability in terms of the minimum singular value of the matrix $\boldsymbol{H}/\|\boldsymbol{h}\|$, and a desirable property this minimum singular value possesses.

In order to upper-bound the symbol error probability of maximum-likelihood sequence estimation we compute the block error probability. Recall the transmitter sends $N - L + 1$ symbols followed by $L - 1$ zeros. Decoding commences upon reception of the Nth symbol. We refer to the M^{N-L+1} possible transmit sequences as "codewords" (although coding is not performed over time) and denote them $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{M^{N-L+1}-1}$. Assume without loss of generality that $\boldsymbol{x}_0$ was transmitted. The receiver observes $\boldsymbol{y} = [y[0], \ldots, y[N-1]]^T$ given by (3).

Geometrically, we view the complex-valued N-length codewords as real-valued length-$2N$ vectors in $\mathbb{R}^{2N}$. The received vector is viewed as an element in $\mathbb{R}^{2N}$ in the same way. The transmitter constellation for a block of N QAM symbols is a regular $2N$-dimensional lattice containing M^{N-L+1} points with

$$n \triangleq \mathsf{SNR}^{r'/2}$$

per dimension. From (3) we see the receiver constellation is a linearly transformed $2N$-dimensional lattice. The linear operator that performs the transformation between these two lattices is the matrix $\boldsymbol{H}$ (see (4)). See Fig. 7.

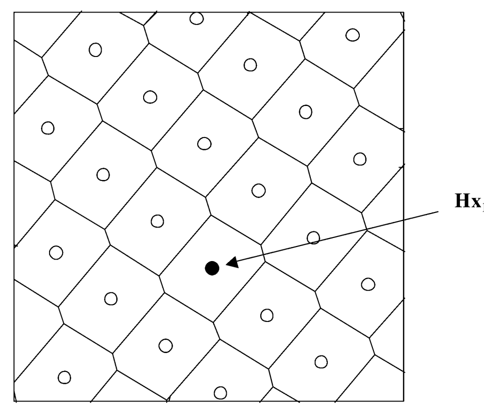

Fig. 8. Geometric view of ML decoding: an error is made if the received vector lands outside of the Voronoi cell generated by $\boldsymbol{H}x_1$.

As the noise $\boldsymbol{w}$ is AWG and circularly symmetric, we interpret the job of the ML decoder geometrically—it selects the point in the receiver constellation closest to the received vector $\boldsymbol{y}$. An error occurs if the closest point to the received vector is not the one that was transmitted. If we form a Voronoi tessellation of the receiver constellation using the points as generators, then the ML decoder makes an error whenever the received vector lands outside the Voronoi cell generated by the transmitted point. See Fig. 8. We use these geometric ideas to upper-bound the error probability as follows.

Let $\boldsymbol{x}_{i^*}$ denote the closest codeword to $\boldsymbol{x}_0$ at the receiver in $\mathbb{C}^N$ under the complex L^2 Euclidean norm, i.e.,

$$\boldsymbol{x}_{i^*} = \arg\min_i \|\boldsymbol{H}(\boldsymbol{x}_i - \boldsymbol{x}_0)\|.$$

Define $\boldsymbol{e}_0 = \boldsymbol{H}(\boldsymbol{x}_i - \boldsymbol{x}_0)/\|\boldsymbol{H}(\boldsymbol{x}_i - \boldsymbol{x}_0)\|$ and let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{N-1}$ be a set of orthogonal unit vectors in $\mathbb{C}^N$, orthogonal to $\boldsymbol{e}_0$ so that $\{\boldsymbol{e}_0, \ldots, \boldsymbol{e}_{N-1}\}$ constitutes a basis for $\mathbb{C}^N$. Denote the region enclosed by the hypersphere of radius $\|\boldsymbol{H}(\boldsymbol{x}_i - \boldsymbol{x}_0)\|/2$ centered at $\boldsymbol{H}x_0$ in the receiver space, by $\mathcal{S}$. As $\boldsymbol{H}x_{i^*}$ is the

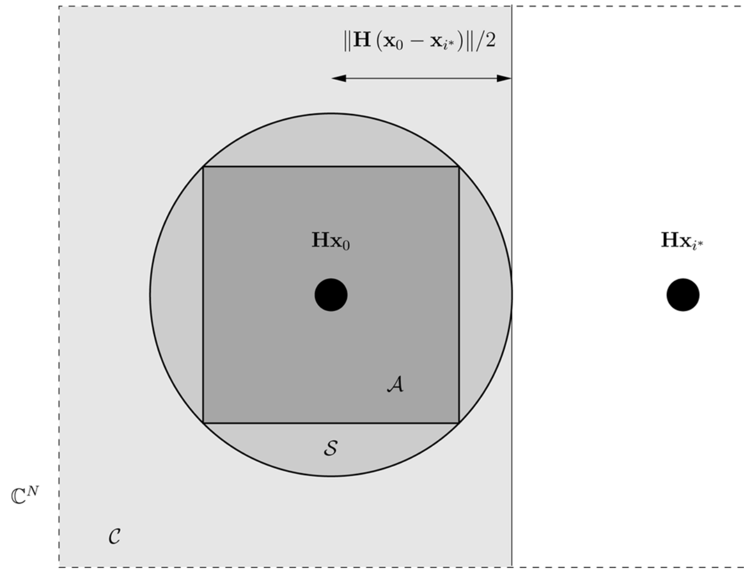

Fig. 9. The probability of error is upper-bounded by the probability the received vector $\boldsymbol{y}$ lands outside the hypercube $\mathcal{A}$, which is of the same order as the pairwise error probability—the probability $\boldsymbol{y}$ lands outside $\mathcal{C}$.

closest codeword to $\boldsymbol{H}x_0$, the error event is a subset of the event that the received vector falls outside $\mathcal{S}$. In other words, an error does not occur if $y \in \mathcal{S}$. Circumscribe a hypercube of width $\|\boldsymbol{H}(\boldsymbol{x}_{i^*} - \boldsymbol{x}_0)\|/\sqrt{2}$ inside S. Denote it $\mathcal{A}$. See Fig. 9. Then, as $\mathcal{A} \subset \mathcal{S}$ an error *does not* occur if $y \in \mathcal{A}$. Denote the real and imaginary parts of X, $\mathcal{R}\{X\}$, and $\mathcal{I}\{X\}$, respectively. Thus, an error does not occur if

$$\mathcal{R}\{(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}_0)^*\boldsymbol{e}_k\} \leq \frac{\|\boldsymbol{H}(\boldsymbol{x}_{i^*} - \boldsymbol{x}_0)\|}{2\sqrt{2}} \quad \text{and}$$
$$\mathcal{I}\{(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}_0)^*\boldsymbol{e}_k\} \leq \frac{\|\boldsymbol{H}(\boldsymbol{x}_{i^*} - \boldsymbol{x}_0)\|}{2\sqrt{2}}$$

for all $k \in \{1, \ldots, N\}$. Now the error probability conditioned on a channel realization of $\boldsymbol{h}$ can be bounded as per (27) (shown at the bottom of the page), where the third step follows from the circular symmetry and i.i.d. assumptions on $\boldsymbol{w}$, the fourth from the simple inequality $mx \geq 1 - (1-x)^m$ for all x and m an integer, and the fifth from the fact $\boldsymbol{w} \sim \mathcal{CN}(0,1)$. The notation $\mathcal{L}\{X\} < Y$ for $\mathcal{L} \in \{\mathcal{R}, \mathcal{I}\}$ indicates both the events $\mathcal{R}\{X\} < Y$ and $\mathcal{I}\{X\} < Y$ occur.

The last line contains an unexpected result. Recall the pairwise error probability between two codewords

$$\Pr(\boldsymbol{x}_0 \to \boldsymbol{x}_i|\boldsymbol{h}) = Q(\sqrt{\|\boldsymbol{H}(\boldsymbol{x}_i - \boldsymbol{x}_0)\|^2/2}).$$

So $\Pr(e|\boldsymbol{h})$ is roughly equal to the pairwise error probability of confusing $\boldsymbol{x}_0$ with its *channel dependent* nearest neighbor, $\boldsymbol{x}_{i^*}$. More specifically, and importantly, the two error probabilities are of the same asymptotic order in SNR. What is also evident is that for each channel realization, the high-SNR asymptotics of the error probability are completely determined by the properties of a single codeword—the nearest neighbor. When computing the average error probability the union bound is redundant.

We scale the codewords to take on integer values. The spacing between neighboring symbols in the QAM constellation is

$$\sqrt{\frac{3\mathsf{SNR}/2}{\mathsf{SNR}^{r'} - 1/4}},$$

which for large SNR is $\sqrt{3/2}\mathsf{SNR}^{(1-r')/2}$. If we set

$$\overline{\boldsymbol{x}}_k = \boldsymbol{x}_k\sqrt{2/3}\mathsf{SNR}^{-(1-r')/2}$$

$$\begin{aligned}
\Pr(e|\boldsymbol{h}) &\leq 1 - \Pr(e^c|\boldsymbol{h}) \\
&= 1 - \Pr(\mathcal{L}\{(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}_0)^*\boldsymbol{e}_k\} \leq \|\boldsymbol{H}(\boldsymbol{x}_{i^*} - \boldsymbol{x}_0)\|/2\sqrt{2}; \mathcal{L} \in \{\mathcal{R}, \mathcal{I}\}, \forall k|\boldsymbol{h}) \\
&= 1 - \Pr(\mathcal{L}\{\boldsymbol{w}^*\boldsymbol{e}_k\} \leq \|\boldsymbol{H}(\boldsymbol{x}_{i^*} - \boldsymbol{x}_0)\|/2\sqrt{2}; \mathcal{L} \in \{\mathcal{R}, \mathcal{I}\}, \forall k|\boldsymbol{h}) \\
&= 1 - (1 - \Pr(\mathcal{R}\{\boldsymbol{w}^*\boldsymbol{H}(\boldsymbol{x}_{i^*} - \boldsymbol{x}_0)\} \geq \|\boldsymbol{H}(\boldsymbol{x}_{i^*} - \boldsymbol{x}_0)\|^2/2\sqrt{2}|\boldsymbol{h}))^{2N} \\
&\leq 2N.\Pr(\mathcal{R}\{\boldsymbol{w}^*\boldsymbol{H}(\boldsymbol{x}_{i^*} - \boldsymbol{x}_0)\} \geq \|\boldsymbol{H}(\boldsymbol{x}_{i^*} - \boldsymbol{x}_0)\|^2/2\sqrt{2}|\boldsymbol{h}) \\
&= 2N.Q\left(\sqrt{\|\boldsymbol{H}(\boldsymbol{x}_{i^*} - \boldsymbol{x}_0)\|^2/2\sqrt{2}}\right) \\
&\leq 2N.Q\left(\sqrt{\min_i \|\boldsymbol{H}(\boldsymbol{x}_i - \boldsymbol{x}_0)\|^2/2\sqrt{2}}\right)
\end{aligned} \qquad (27)$$

the constellation spacing for $\bar{\boldsymbol{x}}$ is 1, i.e., for n even

$$\bar{\boldsymbol{x}}_k \in \left\{-\frac{n-1}{2},\ldots,-\frac{3}{2},-\frac{1}{2},\frac{1}{2},\frac{3}{2},\ldots,\frac{n-1}{2}\right\} \times j\left\{-\frac{n-1}{2},\ldots,-\frac{3}{2},-\frac{1}{2},\frac{1}{2},\frac{3}{2},\ldots,\frac{n-1}{2}\right\}.$$

Denote the normalized vector discriminating between $\boldsymbol{x}_0$ and $\boldsymbol{x}_i$

$$\boldsymbol{a} = \bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_0.$$

Each element $a[k]$ of $\boldsymbol{a}$ lies in the set $\{-n+1,\ldots,n-1\} \times j\{-n+1,\ldots,n-1\}$, but as an error does not occur if we confuse the transmitted codeword with itself, $a[0],\ldots,a[N-L+1]$ cannot all be zero. Define the set

$$\mathcal{A}_n \triangleq \{\boldsymbol{a} \in (\{-n+1,\ldots,n-1\} \times j\{-n+1,\ldots,n-1\})^{N-L+1} \text{ at least one } a[k] \neq 0\}.$$

Then

$$\begin{aligned}
&\Pr(e|\boldsymbol{h}) \\
&\le 2N.Q\left(\sqrt{3\mathsf{SNR}^{1-r'} \min_i \|\boldsymbol{H}(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_0)\|^2/8\sqrt{2}}\right) \\
&= 2N.Q\left(\sqrt{3\mathsf{SNR}^{1-r'} \min_{\boldsymbol{a}\in\mathcal{A}_n} \|\boldsymbol{H}\boldsymbol{a}\|^2/8\sqrt{2}}\right) \\
&= 2N.Q\left(\sqrt{3\mathsf{SNR}^{1-r'}\|\boldsymbol{h}\|^2 \min_{\boldsymbol{a}\in\mathcal{A}_n} \|\bar{\boldsymbol{H}}\boldsymbol{a}\|^2/8\sqrt{2}}\right) \\
&\le 2N.Q\left(\sqrt{3\mathsf{SNR}^{1-r'}\|\boldsymbol{h}\|^2 \inf_{\boldsymbol{h}\in\mathbb{C}^L} \inf_{\boldsymbol{a}\in\mathcal{T}} \|\bar{\boldsymbol{H}}\boldsymbol{a}\|^2/8\sqrt{2}}\right) \\
&= 2N.Q\left(\sqrt{3\mathsf{SNR}^{1-r'}\|\boldsymbol{h}\|^2 \inf_{\boldsymbol{h}\in\mathbb{C}^L} \lambda^2_{\min}(\bar{\boldsymbol{H}})/8\sqrt{2}}\right) \\
&\le 2N.Q\left(\sqrt{3\mathsf{SNR}^{1-r'}\|\boldsymbol{h}\|^2\bar{\lambda}/8\sqrt{2}}\right)
\end{aligned}$$

where $\lambda_{\min}(\bar{\boldsymbol{H}})$ is the minimum singular value of $\bar{\boldsymbol{H}}$

$$\bar{\lambda} \triangleq \inf_{\boldsymbol{h}\in\mathbb{C}^L} \lambda^2_{\min}(\bar{\boldsymbol{H}}) \tag{28}$$

$$\bar{\boldsymbol{H}} \triangleq \frac{\boldsymbol{H}}{\|\boldsymbol{h}\|} \tag{29}$$

and the set

$$\mathcal{T} \triangleq \{\boldsymbol{a} \in \mathbb{C}^{N-L+1} : \|\boldsymbol{a}\| \ge 1\}. \tag{30}$$

The average error probability is obtained by taking the expectation over the channel

$$\begin{aligned}
P_e &\le E\left[2N.Q\left(\sqrt{3\mathsf{SNR}^{1-r'}\|\boldsymbol{h}\|^2\bar{\lambda}/8\sqrt{2}}\right)\right] \\
&\le 2N.E\left[e^{-3\mathsf{SNR}^{1-r'}\|\boldsymbol{h}\|^2\bar{\lambda}/8\sqrt{2}}\right] \\
&= 2N\left(\frac{32\sqrt{2}}{3\bar{\lambda}}\right)^L \mathsf{SNR}^{-L(1-r')}.
\end{aligned}$$

As $\bar{\lambda}$ is not a function of SNR, we need only show that $\bar{\lambda} > 0$.

Lemma IV.1:

$$\bar{\lambda} > 0.$$

Proof:

$$\begin{aligned}
\bar{\lambda} &= \inf_{\boldsymbol{h}\in\mathbb{C}^L}\left(\inf_{\boldsymbol{a}\in\mathcal{T}} \|\bar{\boldsymbol{H}}\boldsymbol{a}\|^2\right) \\
&= \inf_{\boldsymbol{a}\in\mathcal{T}} \inf_{\boldsymbol{h}\in\mathbb{C}^L} \frac{\|\boldsymbol{H}\boldsymbol{a}\|^2}{\|\boldsymbol{h}\|^2} \\
&= \inf_{\boldsymbol{a}\in\mathcal{T}} \inf_{\boldsymbol{h}\in\mathbb{C}^L} \frac{\|\boldsymbol{A}\boldsymbol{h}\|^2}{\|\boldsymbol{h}\|^2} \\
&= \inf_{\boldsymbol{a}\in\mathcal{T}} \inf_{\boldsymbol{h}\in\mathcal{S}} \|\boldsymbol{A}\boldsymbol{h}\|^2 \\
&= \inf_{\boldsymbol{h}\in\mathcal{S}} \inf_{\boldsymbol{a}\in\mathcal{T}} \|\boldsymbol{H}\boldsymbol{a}\|^2 \\
&= \inf_{\boldsymbol{h}\in\mathcal{S}} \lambda^2_{\min}(\boldsymbol{H})
\end{aligned}$$

where the matrix

$$\boldsymbol{A} = \begin{bmatrix} a_0 & 0 & \cdots & 0 \\ a_1 & a_0 & \ddots & \vdots \\ \vdots & a_1 & \ddots & 0 \\ a_{L-1} & \vdots & \ddots & a_0 \\ \vdots & a_{L-1} & \ddots & a_1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{N-1} & a_{N-2} & \cdots & a_{N-L+1} \end{bmatrix} \tag{31}$$

and the set

$$\mathcal{S} \triangleq \{\boldsymbol{h} \in \mathbb{C}^L : \|\boldsymbol{h}\| = 1\}. \tag{32}$$

By inspection the columns of $\boldsymbol{H}$ are linearly independent so $\boldsymbol{H}$ is full rank for each $\boldsymbol{h} \in \mathcal{S}$ and hence $\lambda_{\min}(\boldsymbol{H}) > 0$ pointwise over $\mathcal{S}$. But $\mathcal{S}$ is compact and the singular values of $\boldsymbol{H}$ are continuous functions of its entries and therefore continuous functions of $\boldsymbol{h}$. Hence, $\lambda_{\min}(\boldsymbol{H})$ achieves its infimum over $\mathcal{S}$ and

$$\inf_{\boldsymbol{h}\in\mathcal{S}} \lambda^2_{\min}(\boldsymbol{H}) = \min_{\boldsymbol{h}\in\mathcal{S}} \lambda^2_{\min}(\boldsymbol{H}) > 0.$$

Thus, $\bar{\lambda} > 0$. ■

We conclude

$$P_e \ \dot{\le}\ \mathsf{SNR}^{-L(1-r')}.$$

By Lemma VII.7 this bound is tight and hence $d(r) = L(1-r')$. □

Proof of Theorem III.12: We bound the performance of MLSD by analyzing a suboptimal decoder that works as follows. Let $\boldsymbol{y} = [y[0],\ldots,y[N-1]]^T$, $\boldsymbol{w} = [w[0],\ldots,w[N-1]]^T$, and $\epsilon > 0$ be an arbitrarily small constant. Define the following events:

$$\begin{aligned}
E_0 &\triangleq \{|h_0|^2 > \mathsf{SNR}^{r-1+\epsilon}\} \\
E_1 &\triangleq \{|h_0|^2 < \mathsf{SNR}^{r-1+\epsilon}\} \bigcup \left\{|h_1|^2 > \mathsf{SNR}^{\frac{N}{N-1}r-1+2\epsilon}\right\} \\
E_2 &\triangleq \{|h_0|^2 < \mathsf{SNR}^{r-1+\epsilon}\} \bigcup \left\{|h_1|^2 < \mathsf{SNR}^{\frac{N}{N-1}r-1+2\epsilon}\right\}.
\end{aligned}$$

For $r \in [0, 1-1/N)$ decode $x[0]$ by declaring an error if E_2 occurs and otherwise selecting $\hat{x}[0]$ according to (9) and (10) with

$$\boldsymbol{g}_0^* = \begin{cases} \left[\frac{1}{h_0}, 0, \ldots, 0\right], & \text{if } E_0 \text{ occurs} \\ \left[0, \ldots, 0, \frac{1}{h_1}, -\frac{h_0}{h_1^2}, \ldots, (-1)^N \frac{h_0^{N-2}}{h_1^{N-1}}\right], & \text{if } E_1 \text{ occurs.} \end{cases}$$

To clarify $\boldsymbol{g}_0 \in \mathbb{C}^N$. Thus, if E_0 occurs the filtered estimate is

$$\boldsymbol{g}_0^* \boldsymbol{y} = x[0] + w[0]/h_0$$

and if E_1 occurs the filtered estimate is

$$\boldsymbol{g}_0^* \boldsymbol{y} = x[0] + \frac{h_0^{N-1}}{h_1^{N-1}} x[N-1] + \boldsymbol{g}_0^* \boldsymbol{w}.$$

For $r \in [1-1/N, 1]$ always decode $x[0]$ by selecting $\hat{x}[0]$ according to (9) and (10) with

$$\boldsymbol{g}^* = \left[\frac{1}{h_0}, 0, \ldots, 0\right].$$

Analysis for $r \in [0, 1-1/N)$. By Lemma VII.6, the error probability can be bounded as per (33), shown at the bottom of the page, in which we have used Lemma VII.2 in the sixth step. Taking $\epsilon \to 0$ yields the desired result for $r \in [0, 1-1/N)$. For $r \in [1-1/N, 1]$

$$\begin{aligned} P_e &\dot{\leq} \Pr(\mathsf{SNR}_{\text{eff}} < \mathsf{SNR}^r) \\ &= \Pr(|h_0|^2 \mathsf{SNR} < \mathsf{SNR}^r) \\ &\doteq \mathsf{SNR}^{-(1-r)}. \end{aligned}$$

To decode $x[k]$ for $k > 0$, the receiver uses a successive cancellation procedure to subtract off the interference from $x[k-1], x[k-2], \ldots, x[0]$. Using the union bound we have the error probability for decoding the kth symbol

$$P_e(k) \dot{\leq} \begin{cases} k \cdot \mathsf{SNR}^{-2\left(1-\frac{r}{1-\frac{1}{2N-1}}\right)}, & \text{for } r \in \left[0, 1-\frac{1}{N}\right) \\ k \cdot \mathsf{SNR}^{-(1-r)}, & \text{for } r \in \left[1-\frac{1}{N}, 1\right]. \end{cases}$$

Thus

$$d(r) \geq \max\left\{1-r, 2\left(1-\frac{r}{1-\frac{1}{2N-1}}\right)\right\}. \qquad \square$$

.

C. ZF Equalizer

Proof of Theorem III.6: Denoting the filtered estimates by $\hat{\boldsymbol{y}} = [\hat{y}[0], \ldots, \hat{y}[N-L+1]]^T$, we have

$$\begin{aligned} \hat{\boldsymbol{y}} &= \boldsymbol{G}^* \boldsymbol{y} \\ &= (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{H}x + (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{w} \\ &= \boldsymbol{x} + (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{w} \\ &= \boldsymbol{x} + \boldsymbol{z} \end{aligned}$$

where $\boldsymbol{z} \triangleq (\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{w}$ represents the filtered noise. The total noise variance

$$\begin{aligned} \mathbb{E}||\boldsymbol{z}||^2 &= \mathbb{E} \sum_{k=0}^{N-1} |z_k|^2 \\ &= \mathbb{E}\ \mathrm{tr}\ [\boldsymbol{z}z^*] \\ &= \mathrm{tr}\ [\mathbb{E}(\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{w}w^*\boldsymbol{H}(\boldsymbol{H}^*\boldsymbol{H})^{-1}] \\ &= \mathrm{tr}\ [(\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*(\boldsymbol{I})\boldsymbol{H}(\boldsymbol{H}^*\boldsymbol{H})^{-1}] \\ &= \mathrm{tr}\ [(\boldsymbol{H}^*\boldsymbol{H})^{-1}]. \end{aligned}$$

The effective SNR for decoding the kth symbol is

$$\begin{aligned} \mathsf{SNR}_{\text{eff}}(k) &= \frac{\mathsf{SNR}}{\mathbb{E}|z_k|^2} \\ &\geq \frac{\mathsf{SNR}}{\mathbb{E}||\boldsymbol{z}||^2} \\ &= \frac{\mathsf{SNR}}{\mathrm{tr}\ [(\boldsymbol{H}^*\boldsymbol{H})^{-1}]} \\ &= \frac{||\boldsymbol{h}||^2\mathsf{SNR}}{\mathrm{tr}\ [(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1}]} \\ &= \frac{||\boldsymbol{h}||^2\mathsf{SNR}}{\sum_{k=0}^{N-L} \lambda_k^{-2}(\bar{\boldsymbol{H}})} \\ &\geq \frac{1}{N}\lambda_{\min}^2(\bar{\boldsymbol{H}})||\boldsymbol{h}||^2\mathsf{SNR} \\ &\geq \frac{1}{N} \inf_{\boldsymbol{h}\in\mathbb{C}^L} \lambda_{\min}^2(\bar{\boldsymbol{H}})||\boldsymbol{h}||^2\mathsf{SNR} \\ &= \frac{\bar{\lambda}}{N}||\boldsymbol{h}||^2\mathsf{SNR} \end{aligned}$$

$$\begin{aligned} P_e &\doteq \Pr(\mathsf{SNR}_{\text{eff}} < \mathsf{SNR}^r) \\ &\dot{\leq} \Pr(\mathsf{SNR}_{\text{eff}} < \mathsf{SNR}^r, E_0) + \Pr(\mathsf{SNR}_{\text{eff}} < \mathsf{SNR}^r, E_1) + \Pr(E_2) \\ &\leq \Pr(\mathsf{SNR}_{\text{eff}} < \mathsf{SNR}^r | E_0) + \Pr(\mathsf{SNR}_{\text{eff}} < \mathsf{SNR}^r | E_1) + \Pr(E_2) \\ &= \Pr(|h_0|^2\mathsf{SNR} < \mathsf{SNR}^r | E_0) + \Pr\left(\frac{\mathsf{SNR}}{\frac{|h_0|^{2(N-1)}}{|h_1|^{2(N-1)}}\mathsf{SNR} + \frac{1}{|h_1|^2}\left(1 + \frac{|h_0|^2}{|h_1|^2} + \cdots + \frac{|h_0|^{2(N-2)}}{|h_1|^{2(N-2)}}\right)} < \mathsf{SNR}^r \,\middle|\, E_1\right) + \Pr(E_2) \\ &= 0 + \Pr\left(\frac{\mathsf{SNR}}{\mathsf{SNR}^{1-r-\epsilon(N-1)} + \mathsf{SNR}^{1-\frac{N}{N-1}r-2\epsilon}\left(1 + \cdots + \mathsf{SNR}^{-\frac{N-2}{N-1}r-(N-2)\epsilon}\right)} < \mathsf{SNR}^r\right) + \Pr(E_2) \\ &\doteq \Pr\left(\frac{\mathsf{SNR}}{\mathsf{SNR}^{1-r-\epsilon(N-1)}} < \mathsf{SNR}^r\right) + \mathsf{SNR}^{r-1+\epsilon}\mathsf{SNR}^{\frac{N}{N-1}r-1+2\epsilon} \\ &= 0 + \mathsf{SNR}^{\frac{2N-1}{N-1}r-2+3\epsilon} \\ &= \mathsf{SNR}^{-2\left(1-\frac{r}{1-\frac{1}{2N-1}}\right)+3\epsilon}. \end{aligned} \tag{33}$$

where $\bar{\boldsymbol{H}}$ is defined in (29) and $\bar{\lambda}$ is defined in (28). Thus, by Lemma VII.6

$$\begin{aligned} P_e &\doteq \Pr(\mathsf{SNR}_{\text{eff}}(k) < \mathsf{SNR}^{r'}) \\ &\le \Pr(\|\boldsymbol{h}\|^2 < N\bar{\lambda}^{-1}\mathsf{SNR}^{r'-1}) \\ &\doteq \bar{\lambda}^{-L}\mathsf{SNR}^{-L(1-r')}. \end{aligned}$$

In the last step, we have invoked Lemma VII.2. Lemma IV.1 tells us that $\bar{\lambda} > 0$ hence

$$P_e \dot{\le} \mathsf{SNR}^{-L(1-r')}.$$

Lemma VII.7 demonstrates this bound is tight. Hence, $d(r) = L(1-r')$ and the proof is complete. □

Proof of Theorem III.7: The effective signal-to-noise ratio (see [5, p. 620])

$$\mathsf{SNR}_{\text{eff}}^{\text{ZF}} = \mathsf{SNR} \cdot \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{d\omega}{|H(\omega)|^2}\right)^{-1}, \tag{34}$$

where $H(\omega)$ is defined in (8). As $|H(\omega)|^2$ is periodic

$$\int_0^{2\pi} \frac{d\omega}{|H(\omega)|^2} = \int_0^{2\pi} \frac{d\nu}{|H(\omega^* + \nu)|^2} \tag{35}$$

for any ω^*. Choose ω^* to be the frequency at which $|H(\omega)|^2$ is minimized

$$\omega^* \triangleq \arg \min_{\omega \in [0, 2\pi)} |H(\omega)|^2. \tag{36}$$

Then

$$\begin{aligned} &|H(\omega^* + \nu)|^2 \\ &= \left|\sum_{l=0}^{L-1} h_l e^{j(\omega^*+\nu)l}\right|^2 \\ &= \left|\sum_{l=0}^{L-1} h_l e^{j\omega^* l} + \sum_{l=1}^{L-1} h_l e^{j\omega^* l}(e^{j\nu l} - 1)\right|^2 \\ &\dot{\le}\ |H(\omega^*)|^2 + \left|\sum_{l=1}^{L-1} h_l e^{j\omega^* l}(e^{j\nu l} - 1)\right|^2 \\ &= |H(\omega^*)|^2 + \left|\sum_{l=1}^{L-1} h_l e^{j\omega^*(l-1)}\left(\frac{e^{j\nu l} - 1}{e^{j\nu} - 1}\right)\right|^2 |e^{j\nu} - 1|^2 \\ &\dot{\le}\ |H(\omega^*)|^2 + \sum_{l=1}^{L-1} |h_l|^2 \left|\frac{e^{j\nu l} - 1}{e^{j\nu} - 1}\right|^2 \nu^2 \\ &\dot{\le}\ |H(\omega^*)|^2 + \|\boldsymbol{h}\|^2\nu^2. \end{aligned}$$

We have used Lemma VII.1 in the third step, the fact that $|e^{jX} - 1|^2 \le X^2$ in the third last and the fact that $|e^{jXl} - 1|^2 \le l^2|e^{jX} - 1|^2$ in the second last. We can further bound this quantity by bounding $|H(\omega^*)|^2$. For any $\bar{\omega} \in [0, 2\pi)$

$$\begin{aligned} |H(\omega^*)|^2 &\le |H(\bar{\omega})|^2 \\ &= \left|\sum_{l=0}^{L-1} h_l e^{j\bar{\omega} l}\right|^2 \\ &= \left||h_0| + e^{j(\bar{\omega} - \angle h_0)} \sum_{l=1}^{L-1} e^{j\bar{\omega}(l-1)}\right|^2. \end{aligned}$$

If we choose $\bar{\omega}$ to satisfy

$$\theta_0 + \pi - \bar{\omega} = \angle\left(\sum_{l=1}^{L-1} h_l e^{j(\bar{\omega}-1)l}\right) \tag{37}$$

and define the random variable

$$\bar{h_0} \triangleq \sum_{l=1}^{L-1} h_l e^{j(\bar{\omega}-1)l}, \tag{38}$$

then

$$|H(\omega^*)|^2 \le ||h_0| - |\bar{h_0}||^2.$$

To see that such an $\bar{\omega}$ can always be found, observe that $\angle(\sum_{l=1}^{L-1} h_l e^{j(\bar{\omega}-1)l})$ is periodic in $\bar{\omega}$ with period 2π and thus must intersect $\pi - \bar{\omega} + \theta_0 \bmod 2\pi$ at least once in the range $[0, 2\pi)$. Combining these bounds we have

$$|H(\omega^* + \nu)|^2 \dot{\le} ||h_0| - |\bar{h_0}||^2 + \|\boldsymbol{h}\|^2\nu^2. \tag{39}$$

Substituting (35) into (34) and using the bound of (39) and Lemma VII.6

$$\begin{aligned} P_e &\doteq \Pr\left(\mathsf{SNR}_{\text{eff}}^{\text{ZF}} < \mathsf{SNR}^r\right) \\ &\dot{\ge} \Pr\left(\int_0^{2\pi} \frac{d\nu}{||h_0| - |\bar{h_0}||^2 + \|\boldsymbol{h}\|^2\nu^2} > \mathsf{SNR}^{1-r}\right) \\ &\ge \Pr\left(\frac{\epsilon}{||h_0| - |\bar{h_0}||^2 + \|\boldsymbol{h}\|^2\epsilon^2} > \mathsf{SNR}^{1-r}\right) \end{aligned}$$

or any $\epsilon < \pi$. So

$$\begin{aligned} P_e &\dot{\ge} \Pr(||h_0| - |\bar{h_0}||^2 + \epsilon^2\|\boldsymbol{h}\|^2 < \epsilon \cdot \mathsf{SNR}^{r-1}) \\ &\ge \Pr(||h_0| - |\bar{h_0}||^2 < \epsilon \cdot \mathsf{SNR}^{r-1}, \\ &\qquad \|\boldsymbol{h}\|^2 < \epsilon^{-1}\mathsf{SNR}^{r-1}). \end{aligned}$$

Choose $\epsilon = \mathsf{SNR}^{r-1}$. Then

$$\begin{aligned} P_e &\dot{\ge} \Pr\left(||h_0| - |\bar{h_0}||^2 < \mathsf{SNR}^{2(r-1)}, \|\boldsymbol{h}\|^2 < 1\right) \\ &\dot{\ge} \mathsf{SNR}^{-(1-r)} \end{aligned}$$

by Lemma VII.8. □

D. MMSE Equalizer

Proof of Theorem III.8: Ideally, we would like the ISI to have a Gaussian distribution, as in this case the performance of the MMSE equalizer is bounded by that of the ZF equalizer and the proof is complete. Unfortunately, the ISI symbols are selected from a QAM distribution, so the proof requires some work.

We denote the kth column of $\boldsymbol{H}$ by $\boldsymbol{c}_k$, i.e., $\boldsymbol{H} = [\boldsymbol{c}_0, \ldots, \boldsymbol{c}_{N-L}]$. The filtered estimate for the kth symbol

$$\begin{aligned} \hat{y}[k] &= \boldsymbol{g}_k^* \boldsymbol{y}_k \\ &= \boldsymbol{g}_k^* \left(\boldsymbol{c}_k x[k] + \boldsymbol{w} + \sum_{l \ne k} \boldsymbol{c}_i x[i]\right). \end{aligned}$$

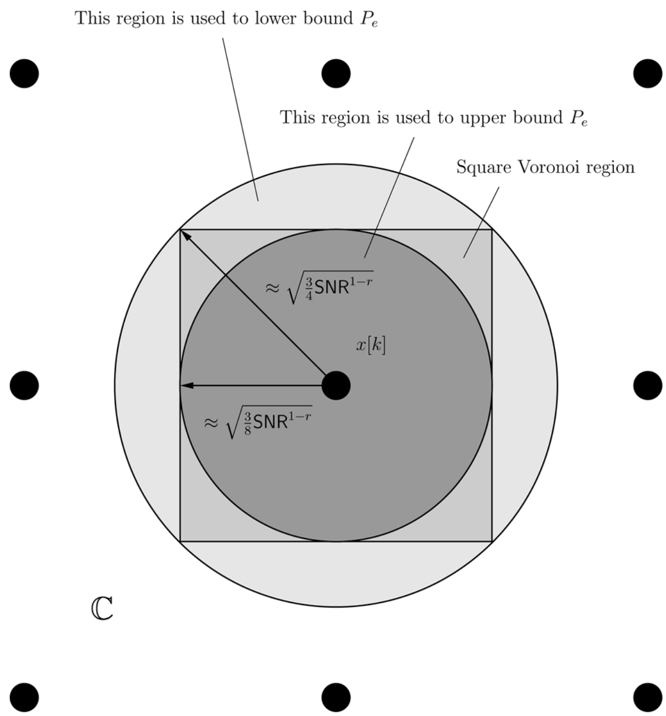

Fig. 10. An error occurs if the received point lies outside the square Voronoi region surrounding $x[k]$. The error probability can therefore be lower-bounded by the probability the received point lies outside the large disc (as it contains the Voronoi region) and upper-bounded by the probability the received point lies outside the small disc (as it is contained within the Voronoi region).

Let $\epsilon > 0$ be an arbitrarily small constant. A decoding error on the kth symbol occurs when the noise + interference is such that the received point $\hat{y}[k]$ lies outside the square Voronoi region belonging to the symbol $x[k]$. See Fig. 10. The probability of this event can be bounded from above by the probability that the received point lies outside the disc centered at $x[k]$ that is just small enough to be contained within the square Voronoi region. Thus, the error probability for the kth symbol can be upper-bounded by the probability that the magnitude of the noise + interference is greater than half the distance between the codewords. For large SNR, the distance between codewords is $\sqrt{3/2}|\boldsymbol{g}_k^*\boldsymbol{c}_k|\mathsf{SNR}^{(1-r)/2}$ so the error probability is bounded above as per (40), shown at the bottom of the page, in which the triangle inequality $|X+Y| < |X| + |Y|$ has been used in the second step. Note from the definition of $\boldsymbol{G}$ in (8) that

$$\begin{bmatrix} \boldsymbol{g}_0^*\boldsymbol{c}_0 & \boldsymbol{g}_0^*\boldsymbol{c}_1 & & \\ \boldsymbol{g}_1^*\boldsymbol{c}_0 & \boldsymbol{g}_1^*\boldsymbol{c}_1 & & \\ & & \ddots & \\ & & & \boldsymbol{g}_{N-L}^*\boldsymbol{c}_{N-L} \end{bmatrix} = \boldsymbol{G}^*\boldsymbol{H}.$$

Define the norm

$$\|\boldsymbol{X}\|_A \triangleq \sum_{i,j} |\boldsymbol{X}_{i,j}|.$$

Then

$$\begin{aligned} \sum_{i\neq k} |\boldsymbol{g}_k^*\boldsymbol{c}_i| &\leq \sum_k \sum_{i\neq k} |\boldsymbol{g}_k^*\boldsymbol{c}_i| \\ &\leq \sum_k \sum_{i\neq k} |\boldsymbol{g}_k^*\boldsymbol{c}_i| + \sum_i |\boldsymbol{g}_i^*\boldsymbol{c}_i - 1| \\ &= \|\boldsymbol{G}^*\boldsymbol{H} - \boldsymbol{I}\|_A. \end{aligned}$$

Let $\boldsymbol{Q}^*\boldsymbol{\Lambda}\boldsymbol{Q} = \bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}}$ be the singular value decomposition (SVD) of $\bar{\boldsymbol{H}} = \boldsymbol{H}/\|\boldsymbol{h}\|$. Then from (13)

$$\begin{aligned} \boldsymbol{G}^*\boldsymbol{H} &= \mathsf{SNR}(\boldsymbol{I} + \mathsf{SNR}\boldsymbol{H}^*\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{H} \\ &= (\boldsymbol{I} + \mathsf{SNR}^{-1}(\boldsymbol{H}^*\boldsymbol{H})^{-1})^{-1} \\ &= (\boldsymbol{I} + \|\boldsymbol{h}\|^{-2}\mathsf{SNR}^{-1}(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1})^{-1} \\ &= \boldsymbol{Q}^*(\boldsymbol{I} + \|\boldsymbol{h}\|^{-2}\mathsf{SNR}^{-1}\boldsymbol{\Lambda}^{-1})^{-1}\boldsymbol{Q}. \end{aligned}$$

With $\|\boldsymbol{h}\|^2 > \mathsf{SNR}^{r'-1+\epsilon}$, for sufficiently large SNR (more specifically $\mathsf{SNR} > \bar{\lambda}^{1/(r'+\epsilon)}$), we may perform a series expansion

$$\begin{aligned} \boldsymbol{G}^*\boldsymbol{H} &= \boldsymbol{Q}^*(\boldsymbol{I} - \|\boldsymbol{h}\|^{-2}\mathsf{SNR}^{-1}\boldsymbol{\Lambda}^{-1} \\ &\quad + \|\boldsymbol{h}\|^{-4}\mathsf{SNR}^{-2}\boldsymbol{\Lambda}^{-2} - \cdots)\boldsymbol{Q} \\ &= \boldsymbol{I} - \|\boldsymbol{h}\|^{-2}\mathsf{SNR}^{-1}(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1} \\ &\quad + \|\boldsymbol{h}\|^{-4}\mathsf{SNR}^{-2}(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-2} - \cdots. \end{aligned} \tag{41}$$

$$\begin{aligned} P_e &\leq \Pr\left(\left|\boldsymbol{g}_k^*\left(\boldsymbol{w} + \sum_{i\neq k}\boldsymbol{c}_i x[i]\right)\right| > \sqrt{\frac{3}{8}}|\boldsymbol{g}_k^*\boldsymbol{c}_k|\mathsf{SNR}^{(1-r')/2}\right) \\ &\leq \Pr\left(|\boldsymbol{g}_k^*\boldsymbol{w}| > |\boldsymbol{g}_k^*\boldsymbol{c}_k|\mathsf{SNR}^{(1-r')/2} - \sum_{i\neq k}|\boldsymbol{g}_k^*\boldsymbol{c}_i||x[i]|\right) \\ &\dot{\leq} \Pr\left(|\boldsymbol{g}_k^*\boldsymbol{w}| > |\boldsymbol{g}_k^*\boldsymbol{c}_k|\mathsf{SNR}^{(1-r')/2} - \sum_{i\neq k}|\boldsymbol{g}_k^*\boldsymbol{c}_i|\mathsf{SNR}^{1/2}\right) \\ &\leq \Pr\left(|\boldsymbol{g}_k^*\boldsymbol{w}| > |\boldsymbol{g}_k^*\boldsymbol{c}_k|\mathsf{SNR}^{(1-r')/2} - \sum_{i\neq k}|\boldsymbol{g}_k^*\boldsymbol{c}_i|\mathsf{SNR}^{1/2}\,\middle|\, \|\boldsymbol{h}\|^2 > \mathsf{SNR}^{r'-1+\epsilon}\right) \\ &\quad + \Pr(\|\boldsymbol{h}\|^2 < \mathsf{SNR}^{r'-1+\epsilon}) \end{aligned} \tag{40}$$

Thus

$$\begin{aligned}\sum_{i\neq k}|\boldsymbol{g}_k^*\boldsymbol{c}_i| \le \| &\|\boldsymbol{h}\|^{-2}\mathsf{SNR}^{-1}(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1} \\ &- \|\boldsymbol{h}\|^{-4}\mathsf{SNR}^{-2}(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-2}+\cdots\|_A \\ &\le \|\boldsymbol{h}\|^{-2}\mathsf{SNR}^{-1}\|(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1}\|_A \\ &+ \|\boldsymbol{h}\|^{-4}\mathsf{SNR}^{-2}\|(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-2}\|_A+\cdots.\end{aligned}$$

This step follows from the triangle inequality. We wish to show that each of the terms $\|(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1}\|_A$, $\|(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-2}\|_A$, $\cdots$ is $\dot{\le}$ SNR^0 so that they can be ignored. Let $\|\cdot\|_F$ denote the Frobenius norm, i.e.,

$$\|X\|_F \triangleq \sqrt{\sum_{i,j}|A_{i,j}|^2}$$

and $\boldsymbol{e}_k$ denote the vector in $\mathbb{C}^{N-L+1}$ with a 1 in its kth entry and 0's everywhere else. Then

$$\begin{aligned}\left\|(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1}\right\|_A^2 &= \left(\sum_{i,j}|[(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1}]_{i,j}|\right)^2 \\ &\le N(N-L+1)\sum_{i,j}|[(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1}]_{i,j}|^2 \\ &\doteq \left\|(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1}\right\|_F^2 \\ &= \sum_{k=0}^{N-L}\|(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1}\boldsymbol{e}_k\|^2 \\ &\le \sum_{k=0}^{N-L}\max_{\boldsymbol{a}\in\mathbb{C}^{N-L+1}:\|\boldsymbol{a}\|=1}\|(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1}\boldsymbol{a}\|^2 \\ &= (N-L+1)\lambda_{\max}^2((\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}})^{-1}) \\ &\doteq \lambda_{\min}^{-4}(\bar{\boldsymbol{H}}) \\ &\le \bar{\lambda}^{-2}\end{aligned}$$

where $\bar{\lambda}$ is defined in (28). Thus, when $\|\boldsymbol{h}\|^2 > \mathsf{SNR}^{r'-1+\epsilon}$

$$\begin{aligned}\sum_{i\neq k}|\boldsymbol{g}_k^*\boldsymbol{c}_i| &\le \|\boldsymbol{h}\|^{-2}\mathsf{SNR}^{-1}\bar{\lambda}^{-2}+\|\boldsymbol{h}\|^{-4}\mathsf{SNR}^{-2}\bar{\lambda}^{-4}+\cdots \\ &= \frac{\|\boldsymbol{h}\|^{-2}\mathsf{SNR}^{-1}\bar{\lambda}^{-2}}{1+\|\boldsymbol{h}\|^{-2}\mathsf{SNR}^{-1}\bar{\lambda}^{-4}} \\ &\doteq \frac{\mathsf{SNR}^{-r'-\epsilon}\bar{\lambda}^{-2}}{1+\mathsf{SNR}^{-r'-\epsilon}\bar{\lambda}^{-4}} \\ &\doteq \mathsf{SNR}^{-r'-\epsilon}.\end{aligned}\tag{42}$$

The last step follows from Lemma IV.1 and the fact that $\bar{\lambda}$ is not a function of SNR. For $\|\boldsymbol{h}\|^2 > \mathsf{SNR}^{r'-1+\epsilon}$ it is also evident from (41) that

$$\begin{aligned}|\boldsymbol{g}_k^*\boldsymbol{c}_k| \ge 1 - \Bigg\|&\|\boldsymbol{h}\|^{-2}\mathsf{SNR}^{-1}\left(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}}\right)^{-1} \\ &-\|\boldsymbol{h}\|^{-4}\mathsf{SNR}^{-2}\left(\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}}\right)^{-2}+\cdots\Bigg\|_A \\ &\ge 1-\mathsf{SNR}^{-r'-\epsilon} \\ &\doteq 1.\end{aligned}\tag{43}$$

The second last step follows from the reasoning which lead to (42). We now have bounds for $\sum_{i\neq k}|\boldsymbol{g}_k^*\boldsymbol{c}_i|$ and $|\boldsymbol{g}_k^*\boldsymbol{c}_k|$. Now we focus on the statistics of the noise. Denote the filtered noise $\boldsymbol{z}=\boldsymbol{G}^*\boldsymbol{w}$. Then conditioned on $\|\boldsymbol{h}\|^2 > \mathsf{SNR}^{r'-1+\epsilon}$, $\boldsymbol{g}_k^*\boldsymbol{w}=z_k$ is a Gaussian random variable with zero mean and variance

$$\begin{aligned}\mathbb{E}|z_k|^2 &\le \mathbb{E}\|\boldsymbol{z}\|^2 \\ &= \mathbb{E}\mathrm{tr}[\boldsymbol{z}\boldsymbol{z}^*] \\ &= \mathrm{tr}\Big[\mathbb{E}\left(\boldsymbol{I}+\mathsf{SNR}\boldsymbol{H}^*\boldsymbol{H}\right)^{-1}\mathsf{SNR}\boldsymbol{H}^*\boldsymbol{w}w^* \\ &\qquad \boldsymbol{H}\mathsf{SNR}\left(\boldsymbol{I}+\mathsf{SNR}\boldsymbol{H}^*\boldsymbol{H}\right)^{-1}\Big] \\ &= \|\boldsymbol{h}\|^2\mathsf{SNR}^2\mathrm{tr}\Big[\left(\boldsymbol{I}+\|\boldsymbol{h}\|^2\mathsf{SNR}\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}}\right)^{-1} \\ &\qquad \bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}}\left(\boldsymbol{I}+\|\boldsymbol{h}\|^2\mathsf{SNR}\bar{\boldsymbol{H}}^*\bar{\boldsymbol{H}}\right)^{-1}\Big] \\ &= \sum_{k=0}^{N-L}\frac{\|\boldsymbol{h}\|^2\mathsf{SNR}^2\lambda_k^2(\bar{\boldsymbol{H}})}{\left(1+\|\boldsymbol{h}|^2\mathsf{SNR}\lambda_k^2(\bar{\boldsymbol{H}})\right)^2} \\ &\le \frac{N-L+1}{\|\boldsymbol{h}\|^2\lambda_{\min}^2(\bar{\boldsymbol{H}})} \\ &\le (N-L+1)\bar{\lambda}^{-1}\mathsf{SNR}^{1-r-\epsilon} \\ &\doteq \mathsf{SNR}^{1-r-\epsilon}.\end{aligned}\tag{44}$$

The last step follows from Lemma IV.1 and the fact that $\bar{\lambda}$ is not a function of SNR. Combining the bounds of (42), (43), and (44) and invoking Lemma VIII.2 we have

$$\begin{aligned}P_e \;&\dot{\le}\; \Pr\Big(|\boldsymbol{g}_k^*\boldsymbol{w}| > \mathsf{SNR}^{1/2-r'/2}-\mathsf{SNR}^{1/2-r'-\epsilon}. \\ &\qquad \Big|\|\boldsymbol{h}\|^2 > \mathsf{SNR}^{r'-1+\epsilon}\Big) \\ &\quad + \Pr\left(\|\boldsymbol{h}\|^2 < \mathsf{SNR}^{r'-1+\epsilon}\right) \\ &\doteq \Pr(|\boldsymbol{g}_k^*\boldsymbol{w}|^2 > \mathsf{SNR}^{1-r'}\,|\,\|\boldsymbol{h}\|^2 > \mathsf{SNR}^{r'-1+\epsilon}) \\ &\quad + \mathsf{SNR}^{-L(1-r'-\epsilon)} \\ &\doteq Q(\sqrt{\mathsf{SNR}^{\epsilon}})+\mathsf{SNR}^{-L(1-r'-\epsilon)} \\ &\le e^{-\mathsf{SNR}^{\epsilon}}+\mathsf{SNR}^{-L(1-r'-\epsilon)} \\ &\doteq \mathsf{SNR}^{-L(1-r'-\epsilon)}.\end{aligned}$$

Taking $\epsilon\to 0$ yields the desired result. □

Proof of Theorem III.7: This proof is long and detailed so we start with an overview.

Discussion: For the ZF equalizer we argued that the error probability was asymptotically bounded below by the outage probability for the scalar AWGN channel linking $x[k]$ and $\hat{y}[k]$ (see Lemma VII.6). We then saw that the error probability was entirely determined by the tail behavior of the effective signal-to-noise ratio at the output of this channel, namely

$$P_e \,\dot{\ge}\, \Pr(\mathsf{SNR}_{\text{eff}} < \mathsf{SNR}^r).\tag{45}$$

It is tempting to appeal to (45) for the MMSE equalizer. The flaw in this approach is that the channel linking $x[k]$ and $\hat{y}[k]$ contains residual ISI which does not have a Gaussian distribution. Consequently, the effective noise (noise + interference) is not Gaussian and we cannot use the technique we used for the ZF equalizer. Instead, we lower-bound the error probability by removing the residual ISI term and analyzing the resulting expression. It seems counterintuitive to think this procedure yields a tight bound. However, it can be shown that the resulting bound is asymptotically equal to the one arrived at when (45) is used with the residual ISI term left in there (but treated as a Gaussian with the same variance as its true distribution). The simple explanation for this is that when an error occurs at high SNR, the additive noise term always either dominates the residual ISI term, or asymptotically matches it. Thus, removing the residual ISI term does not have an impact on the tradeoff curve. The moral of the story is the MMSE equalizer's error probability, like the ZF's, is governed by (45) and the non-Gaussian nature of the ISI is a side issue.

So we now look at (45) in more detail. The effective SNR at the output of the infinite-length MMSE equalizer is (see [5, p. 625])

$$\mathsf{SNR}_{\text{eff}} = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{d\omega}{1+\mathsf{SNR}|H(\omega)|^2}\right)^{-1} - 1.$$

Ideally, we would like to evaluate this integral explicitly but this requires the factorization of a polynomial of degree $2(L-1)$, which cannot be performed for arbitrary L. For $L=2$ we have

$$\mathsf{SNR}_{\text{eff}} = \sqrt{\begin{array}{c}1+2(|h_0|^2+|h_1|^2)\mathsf{SNR}\\ +(|h_0|^2-|h_1|^2)^2\mathsf{SNR}^2\end{array}} - 1. \quad (46)$$

If we treat the residual ISI as Gaussian with the same variance the bound in (45) will be tight (Lemma VII.6) and P_e will asymptotically equal

$$\Pr(|h_1|^2 < \mathsf{SNR}^{2r-1}, ||h_0|^2-|h_1|^2| < \mathsf{SNR}^{r-1}). \quad (47)$$

This is the probability that $|h_1|^2$ is of order SNR^{2r-1} and $|h_0|^2$ is within SNR^{r-1} of $|h_1|^2$. As $|h_0|^2$ and $|h_1|^2$ are i.i.d. exponential random variables with mean 1, the probability of such an event is (see Lemma VII.4 and Corollary VII.3)

$$\mathsf{SNR}^{-(1-2r)^+}\mathsf{SNR}^{r-1}$$

where $X^+ = X$ if $X>0$ and $X^+=0$ if $X<0$. Thus

$$d(r) = (1-2r)^+ + 1 - r. \quad (48)$$

This two-part curve matches the bound of Theorem III.9. The conclusion we draw is the following: if the residual ISI is treated as Gaussian then at least for $L=2$ the bound of Theorem III.9 is tight.

Unfortunately, this simple analysis does not shed much light on the mechanism that causes an error to occur—a topic that we will now begin to discuss. For this discussion, we will concentrate on the $L=2$ case and assume the residual ISI is Gaussian so that inequality (45) is tight. Let the random variable ω^* denote the frequency at which the channel attains its minimum value (see (36)). From (46) we have

$$P_e \doteq \Pr\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{d\omega}{1+\mathsf{SNR}|H(\omega)|^2} > \mathsf{SNR}^{-r}\right).$$

Thus, in order for an error to occur the channel $|H(\omega)|^2$ must fade sufficiently deeply and over a sufficiently large bandwidth so as to cause the integral in the above equation to blow up to a value as great as SNR^{-r}. We claim the following is a typical error event:

$$E_1(r) \triangleq \begin{cases}\{||h_0|-|h_1||^2 \doteq \mathsf{SNR}^{2(r-1)}, |h_1|^2 \doteq 1\}, & \text{if } r>1/2\\ \{||h_0|-|h_1||^2 \doteq \mathsf{SNR}^{-1}, |h_1|^2 \doteq \mathsf{SNR}^{2r-1}\}, & \text{if } r<1/2.\end{cases}$$

We will interpret this event shortly but first we validate the above claim by verifying that $E_1(r)$ causes an error, and that $\Pr(E_1(r)) \doteq \mathsf{SNR}^{-d^*(r)}$.

For $r > 1/2$, the two conditions imply we have both $|h_1|^2 \doteq 1$ and $|h_0|^2 \doteq 1$. Thus

$$\begin{aligned}||h_0|^2-|h_1|^2| &= ||h_0|-|h_1||||h_0|+|h_1||\\ &\doteq ||h_0|-|h_1||\\ &\doteq \mathsf{SNR}^{r-1}.\end{aligned}$$

So for this range of r event $E_1(r)$ implies both conditions in (47) are satisfied and hence an error occurs. Also

$$\begin{aligned}&\Pr(E_1(r))\\ &= \Pr\left(||h_0|-|h_1||^2 \doteq \mathsf{SNR}^{2(r-1)}\right)\\ &= \Pr\left(||h_0|^2-|h_1|^2|^2 \doteq \mathsf{SNR}^{2(r-1)}(|h_0|+|h_1|)^2\right)\\ &= \Pr(||h_0|^2-|h_1|^2| \doteq \mathsf{SNR}^{r-1})\\ &= \mathsf{SNR}^{r-1}\end{aligned}$$

by Lemma VII.4, which has an exponent matching (48).

For $r < 1/2$, the two conditions imply we have both $|h_1|^2 \doteq \mathsf{SNR}^{2r-1}$ and $|h_0|^2 \doteq \mathsf{SNR}^{2r-1}$. Thus

$$\begin{aligned}||h_0|^2-|h_1|^2| &\doteq ||h_0|-|h_1||\mathsf{SNR}^{2r-1}\\ &\doteq \mathsf{SNR}^{-1/2}\mathsf{SNR}^{r-1/2}\\ &\doteq \mathsf{SNR}^{r-1}.\end{aligned}$$

So also for this range of r event $E_1(r)$ implies both conditions in (47) are satisfied and hence an error occurs. Additionally it is straightforward to show that

$$\Pr(E_1(r)) \doteq \mathsf{SNR}^{3r-2}$$

which matches the exponent in (48), implying $E_1(r)$ is typical. □

Now to interpret event E_1. From the above discussion we see we can rewrite it as

$$E_1(r) = \begin{cases}\{||h_0|-|h_1||^2 \doteq \mathsf{SNR}^{2(r-1)}, |h_0||h_1| \doteq 1\}, & \text{if } r>1/2\\ \{||h_0|-|h_1||^2 \doteq \mathsf{SNR}^{-1}, |h_0||h_1| \doteq \mathsf{SNR}^{2r-1}\}, & \text{if } r<1/2\end{cases}$$

A little computation reveals

$$||h_0|-|h_1||^2 = |H(\omega^*)|^2 \quad \text{and}$$
$$|h_0||h_1| = \left.\frac{d^2|H(\omega)|^2}{d\omega^2}\right|_{\omega=\omega^*}.$$

Thus, for $r > 1/2$ the typical error event is caused by the channel minimum $|H(\omega^*)|^2$ fading to the level $\mathsf{SNR}^{2(r-1)}$. For $r < 1/2$ it is caused both by the channel minimum $|H(\omega^*)|^2$ fading to the level SNR^{-1} *and* the second derivative of the

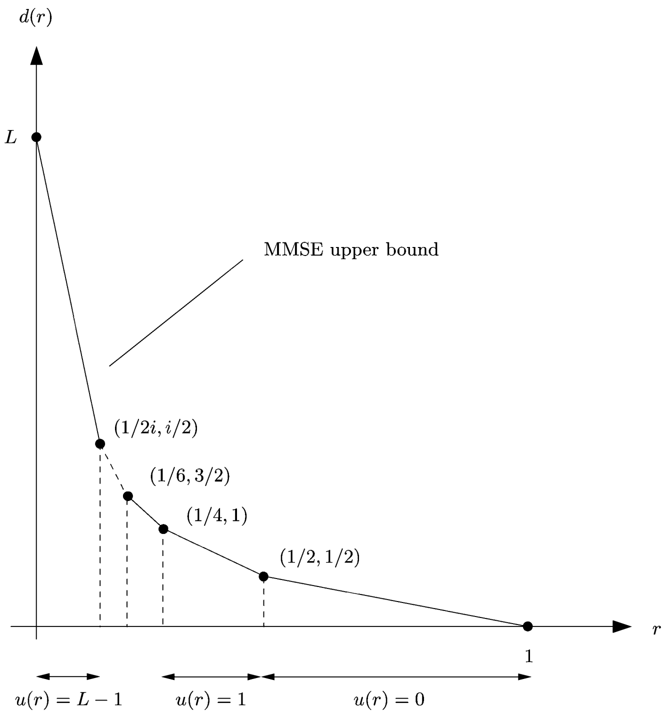

Fig. 11. Illustration of the function $u(r) : (0,1) \rightarrow \{0,\ldots,L-1\}$.

channel at the minimum, reducing to SNR^{2r-1}. The insight gained from this discussion forms the basis of the Proof of Theorem III.9, which we now commence.

Proof of Theorem III.9: We lower-bound the error probability by removing the residual ISI, as alluded to earlier. From (1) and (12) the filtered estimate can be written in the form

$$\hat{y}[k] = A_x[0]x[k] + \sum_{i\neq k} A_x[k-i]x[i] + z[k]$$

where

$$A_x[m] \triangleq \sum_{l=0}^{L-1} h_l g_{m-l}, \tag{49}$$

and

$$z[k] \triangleq \sum_{m=-\infty}^{\infty} g_m w[k-m].$$

Let

$$\boldsymbol{x}_{\neq k} \triangleq \{\ldots, x[k-2], x[k-1], x[k+1], x[k+2], \ldots\}$$

denote the ISI sequence for the kth symbol. A decoding error on the kth symbol occurs when the noise + interference is such that the received point $\hat{y}[k]$ lies outside the square Voronoi region belonging to the symbol $x[k]$. See Fig. 10. The probability of this event can be bounded below by the probability that the received point lies outside the disc centered at $x[k]$ that is just large enough to contain the square Voronoi region. Thus, the error probability for the kth symbol can be lower-bounded by the probability that the noise + interference exceeds $1/\sqrt{2}$ times the distance between the codewords. For large SNR, the distance between codewords is $\sqrt{3/2}|A_x[0]|\mathsf{SNR}^{(1-r)/2}$ so

$$P_e \doteq \Pr\left(\left|z[k] + \sum_{i\neq k} A_x[k-i]x[i]\right| > |A_x[0]|\mathsf{SNR}^{(1-r)/2}\right)$$

$$= \sum_{\boldsymbol{x}} \Pr\left(\left|z[k] + \sum_{i\neq k} A_x[k-i]x[i]\right| > |A_x[0]|\mathsf{SNR}^{(1-r)/2} | \boldsymbol{x}_{\neq k} = \boldsymbol{x}\right) \Pr(\boldsymbol{x}_{\neq k} = \boldsymbol{x}).$$

For a given ISI sequence $\boldsymbol{x}_{\neq k}$, $z[k] + \sum_{i\neq k} A_x[k-i]x[i]$ is a circularly symmetric complex Gaussian random variable centered around its mean $\sum_{i\neq k} A_x[k-i]x[i]$. A little thought reveals that removing the residual ISI term provides a lower bound to the above expression. This observation is justified in Lemma VII.9. Using it we write

$$\begin{aligned} P_e &\geq \sum_{\boldsymbol{x}} \Pr(|z[k]| > |A_x[0]|\mathsf{SNR}^{(1-r)/2}) \Pr(\boldsymbol{x}_{\neq k} = \boldsymbol{x}) \\ &= \Pr(|z[k]|^2 > |A_x[0]|^2 \mathsf{SNR}^{1-r}) \\ &= \int_{\boldsymbol{h}\in\mathbb{C}^L} \Pr(|z[k]|^2 > |A_x[0]|^2\mathsf{SNR}^{1-r}|\boldsymbol{h}) p(\boldsymbol{h}) d\boldsymbol{h} \\ &= \int_{\boldsymbol{h}\in\mathbb{C}^L} \exp\left(-\frac{|A_x[0]|^2\mathsf{SNR}^{1-r}}{\mathbb{E}|z[k]|^2}\right) p(\boldsymbol{h}) d\boldsymbol{h}. \end{aligned}$$

For notational convenience we now define a simple function $u : (0,1) \rightarrow \{0,\ldots,L-1\}$ that takes $r \in (0,1)$ as input and returns the corresponding multiplexing region as an integer in $\{0,\ldots,L-1\}$, by

$$u(r) \triangleq \begin{cases} 0, & \text{if } \frac{1}{2} \leq r < 1 \\ m, & \text{if } \frac{1}{2(m+1)} \leq r < \frac{1}{2m} \text{ for } m \in \{1,\ldots,L-2\} \\ L-1, & \text{if } 0 < r < \frac{1}{2(L-1)}. \end{cases}$$

See Fig. 11 for an illustration of $u(r)$.

Throughout this section, we will use the notation

$$\theta_k \triangleq \angle h_k.$$

We will define new random variables $\tilde{h}_0(r),\ldots,\tilde{h}_{L-1}(r)$, which for $u(r) < L-1$ are functions solely of $h_{u(r)+1},\ldots,h_{L-1}$ and θ_0, and for $u(r) = L-1$ are functions solely of θ_0 and θ_{L-1}. We refer to them as the *typical channel coefficients* and in future references drop the stated dependence on r for notational simplicity.

$u(r) < L-1$

Let $\tilde{\omega}(r)$ be the largest solution of

$$\theta_0 + \pi = \angle\left((-1)^{u(r)-L-1} \sum_{i=u(r)+1}^{L-1} \binom{i-1}{u(r)} h_i e^{j\tilde{\omega}(r)i}\right) \tag{50}$$

in the range $[0, 2\pi)$. To see that (50) always has at least one solution rewrite it as

$$0 = \sum_{i=u(r)+1}^{L-1} \binom{i-1}{u(r)} |h_i| \sin(i\tilde{\omega}(r) + \theta_i - \theta_0)).$$

Integrating the right-hand side with respect to (w.r.t.) $\tilde{\omega}(r)$ from $-\pi$ to π, reveals its average value equals zero. Thus, as it is a continuous function of $\tilde{\omega}(r)$ it must take on the value zero for at least one value of $\tilde{\omega}(r)$. In the remainder of the proof, we drop

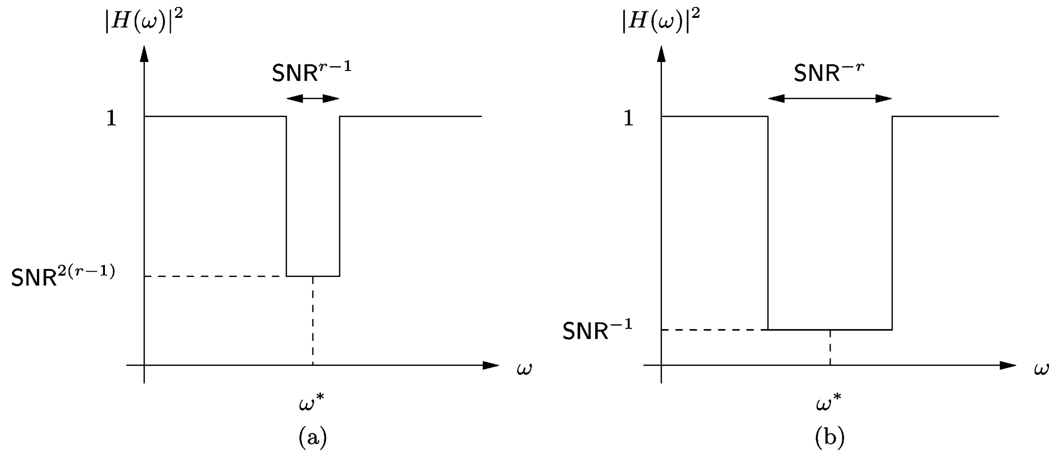

Fig. 12. (a) Typical channel realization for $r > 1/2$. (b) Typical channel realization for $r < 1/2$.

the stated dependence of r (for notational simplicity) and just write $\tilde{\omega}$.

Now define the $\tilde{h}_l$ in terms of $h_{u(r)+1}, \ldots, h_{L-1}$ and θ_0 by a somewhat recursive definition

$$\tilde{h}_l \triangleq \begin{cases} -\sum_{i=l+1}^{L-1} \binom{i}{l} \tilde{h}_i e^{j(i-l)\tilde{\omega}}, & \text{for } l = 0, \ldots, u(r) \\ h_i, & \text{for } l = u(r)+1, \ldots, L-1. \end{cases} \tag{51}$$

A nonrecursive definition is

$$\tilde{h}_l \triangleq \begin{cases} -\sum_{i=u(r)+1}^{L-1} \gamma_{i,l}^{u(r)} h_i e^{j\tilde{\omega} i}, & \text{for } l = 0, \ldots, u(r) \\ h_i, & \text{for } l = u(r)+1, \ldots, L-1 \end{cases}$$

for appropriate constants $\gamma_{i,l}^{u(r)}$, which if desired can be explicitly computed without too much trouble, using (51). □

$\underline{u(r) = L-1}$

For $u(r) = L-1, \tilde{\omega}$ is defined simply by

$$\tilde{\omega} = \frac{\theta_0 - \theta_{L-1}}{L-1}.$$

Note that $\tilde{\omega}$ takes on the same value for $u(r) = L-1$ and $u(r) = L-2$. Define $\tilde{h}_0, \ldots, \tilde{h}_{L-1}$ by

$$\tilde{h}_l \triangleq \begin{cases} -\sum_{i=l+1}^{L-1} \binom{i}{l} \tilde{h}_i e^{j(i-l)\tilde{\omega}}, & \text{for } l = 0, \ldots, L-2 \\ \mathsf{SNR}^{2(L-1)r-1}, & \text{for } l = L-1. \end{cases} \tag{52}$$

□

Next define random variables

$$\bar{h}_l \triangleq -\sum_{i=l+1}^{L-1} \binom{i}{l} h_i e^{j(i-l)\tilde{\omega}}$$

for $l = 0, \ldots, L-2$. Define $\bar{h}_{L-1}$ by $|\bar{h}_{L-1}| \triangleq 0$ and $\angle \bar{h}_{L-1} \triangleq \angle h_{L-1} + \pi$.

Continue Discussion

We now describe an error causing mechanism that generates the L-part bound of Theorem III.9. We believe this bound to be tight[2] which would imply that this mechanism is a typical cause of error. The following explanation will make clear the significance of the newly defined random variables.

For $r > 1/2$ an error occurs when the channel minimum fades to the level $\mathsf{SNR}^{2(r-1)}$, i.e., $|H(\omega^*)|^2 \doteq \mathsf{SNR}^{2(r-1)}$ and the curvature of the channel around this minimum is such that $|H(\omega)|^2 \doteq \mathsf{SNR}^{2(r-1)}$ in an interval of width SNR^{r-1} around ω^*. For $r < 1/2$ an error occurs when the channel minimum fades to the level $|H(\omega^*)|^2 \doteq \mathsf{SNR}^{-1}$ but the channel flattens around the minimum to compensate such that $|H(\omega)|^2 \doteq \mathsf{SNR}^{-1}$ in an interval of width SNR^{-r}. An illustration of these channel realizations is given in Fig. 12.

To see why the channel minimum saturates at the level SNR^{-1}, observe that at high-SNR, $\mathsf{SNR}_{\text{eff}}^{MMSE}$ differs from $\mathsf{SNR}_{\text{eff}}^{\text{ZF}}$ only by a "1+" that appears in the denominator (see (34) and (46)).

What is the probability that the channel takes one of the forms illustrated in Fig. 12? Each of these forms has two properties: (i) the minimum fades to a particular level and (ii) the channel around the minimum flattens to a particular level. Recall

$$|H(\omega^*)|^2 = \left| h_0 + h_1 e^{j\omega^*} + \cdots + h_{L-1} e^{j\omega^*(L-1)} \right|^2.$$

First consider $r > 1/2$. The first property of the typical channel for $r > 1/2$ is that the channel minimum takes on the value $\mathsf{SNR}^{2(r-1)}$. The typical way in which $|H(\omega^*)|^2 \doteq \mathsf{SNR}^{2(r-1)}$ occurs involves $L-2$ of the coefficients taking on order SNR^0 magnitudes, lets say $|h_l|^2 \doteq \mathsf{SNR}^0$ for $l = 1, \ldots, L-1$, and one of them, lets say h_0, pointing precisely in the right direction in the complex plane, and having precisely the right magnitude, to cancel out the sum $h_1 e^{j\omega^*} + \cdots + h_{L-1} e^{j(L-1)\omega^*}$. This sum corresponds to $-\bar{h}_0$. One may be tempted to conclude that this requires two events to occur:

1) $\angle h_0 \approx \angle h_1 e^{j\omega^*} + \cdots + h_{L-1} e^{j(L-1)\omega^*}$, and
2) $|h_0| \approx |h_1 e^{j\omega^*} + \cdots + h_{L-1} e^{j(L-1)\omega^*}|$,

both of which are improbable. But these events are only both improbable if ω^* is some arbitrarily selected frequency. The fact it is the minimizing frequency means that only the second event

[2]This was already proved for the $L = 2$ case under the Gaussian ISI assumption.

is improbable—ω^* will automatically take on a value such that the first event occurs naturally.

More precisely, the second event is

$$||h_0| - |h_1 e^{j\omega^*} + \cdots + h_{L-1} e^{j(L-1)\omega^*}|^2|^2 \doteq \mathsf{SNR}^{2(r-1)}.$$

The probability this occurs depends on the magnitude of $h_1 e^{j\omega^*} + \cdots + h_{L-1} e^{j(L-1)\omega^*}$. For $r > 1/2$ this quantity has order SNR^0 magnitude (we will explain why this is so in a moment) and the probability of the second event is SNR^{r-1}. This is then also the probability of $|H(\omega^*)|^2 \doteq \mathsf{SNR}^{2(r-1)}$ occurring.

The second property of the typical channel for $r > 1/2$ is that the width of the fade around the minimum is SNR^{r-1}. This occurs naturally. To see why, notice that the natural curvature of the channel in a small neighborhood surrounding its minimum is quadratic, i.e.,

$$|H(\omega^* + \epsilon)|^2 = |H(\omega^*)|^2 + K\epsilon^2$$

for some constant K. Let $0 < k_0 < k_1$ be arbitrary constants. Suppose $|H(\omega^*)|^2 = k_0 \mathsf{SNR}^{2(r-1)}$. Then the value of $\epsilon(r)$ required such that $|H(\omega^* + \epsilon(r))|^2 = k_1 \mathsf{SNR}^{2(r-1)}$ is

$$\begin{aligned}\epsilon(r) &= \sqrt{\frac{k_1 - k_0}{K}} \mathsf{SNR}^{r-1} \\ &\doteq \mathsf{SNR}^{r-1}.\end{aligned}$$

We have assumed here that K is independent of $|H(\omega^*)|^2$. □

We conclude that for $r > 1/2$ the typical error event consists merely of the channel minimum fading to the level $\mathsf{SNR}^{2(r-1)}$ and this occurs with probability SNR^{r-1}. Returning to our previous statement that for $r > 1/2$ the magnitude of $h_1 e^{j\omega^*} + \cdots + h_{L-1} e^{j(L-1)\omega^*}$ has order SNR^0, we see that this is because the flatness requirement occurs naturally and does not place any constraints on the coefficients $h_1, \ldots, h_{L-1}$.

For $r < 1/2$ the channel must fade over an increased bandwidth in order to cause a decoding error. Consider a partial Taylor series expansion of $|H(\omega)|^2$ around ω^*

$$\begin{aligned}&|H(\omega^* + \epsilon)|^2 - |H(\omega^*)|^2 \\ &\quad = f_2 \epsilon^2 + f_4 \epsilon^4 + \cdots + f_{2(L-1)} \epsilon^{2(L-1)} + o\left(\epsilon^{2(L-1)}\right) \\ &\quad = f_2 \mathsf{SNR}^{-2r} + f_4 \mathsf{SNR}^{-4r} + \\ &\qquad \cdots + f_{2(L-1)} \mathsf{SNR}^{-2(L-1)r} + o\left(\mathsf{SNR}^{-2(L-1)r}\right).\end{aligned}$$

Only the even powered terms have been included—the odd powered terms can never dominate the expansion around the minimum in an asymptotic sense, as this would contradict the minimality of $|H(\omega^*)|$. An error requires

$$|H(\omega^* + \epsilon)|^2 - |H(\omega^*)|^2 \doteq \mathsf{SNR}^{-1}.$$

This means that for $1/4 < r < 1/2$ we must have $f_2 \doteq \mathsf{SNR}^{2r-1}$, for $1/6 < r < 1/4$ we must have both $f_2 \doteq \mathsf{SNR}^{2r-1}$ and $f_4 \doteq \mathsf{SNR}^{4r-1}$, for $1/8 < r < 1/6$ we must have $f_2 \doteq \mathsf{SNR}^{2r-1}$, $f_4 \doteq \mathsf{SNR}^{4r-1}$, and $f_6 \doteq \mathsf{SNR}^{6r-1}$, and so on. The term f_{2k} corresponds to the $2k$th derivative of the channel evaluated at ω^*. Thus, these derivatives must successively vanish as $r \downarrow 0$. The typical way in which this happens is by the channel coefficients (viewed as phasors in the complex plane) aligning themselves in particular configurations which we refer to as *channel modes*. The modes are given by the relationships between the typical channel coefficients $\tilde{h}_0, \ldots, \tilde{h}_{L-1}$. For $1/2 < r < 1$, the channel coefficients will align to cancel f_0, i.e., they will align so as to almost completely cancel $|H(\omega^*)|^2$. This is the zeroth channel mode. For $1/4 < r < 1/2$, the channel coefficients will align so as to almost completely cancel both the channel minimum f_0 and the second derivative f_2. The is the first channel mode. For $1/6 < r < 1/4$ they will align so as to almost completely cancel the channel minimum, the second derivative, and the fourth derivative. This is the second channel mode. And so on. As the channel mode increases there are successively fewer degrees of freedom to play with, namely, for a given r, θ_0 and the channel coefficients $h_{u(r)+1}, \ldots, h_{L-1}$ are left unconstrained. Thus, by the time we get to $1/2(L-1) < r < 1/2(L-2)$ there are tight constraints on all channel coefficients except h_{L-1} and the phase of h_0. In the range $0 < r < 1/2(L-1)$ something new happens—there are no more degrees of freedom left to cancel more derivatives because all channel coefficients are already constrained (except θ_0 and θ_{L-1}). Consequently, in order to satisfy $f_{2(L-1)} \doteq \mathsf{SNR}^{2(L-1)r-1}$, $f_{2L} \doteq \mathsf{SNR}^{2Lr-1}, \ldots$ all channel coefficients must fade in magnitude—mere alignment is no longer sufficient. It is this behavior which gives rise to the L-part curve in Fig. 5.

Example: $L = 3$

For $1/2 < r < 1$, the channel coefficients align in the complex plane so as to form channel mode 0. This is illustrated in Fig. 13(a). The constrained parameter is $|h_0|$. The minimum is reduced to SNR^{-1}. For $1/4 < r < 1/2$ the channel coefficients align into channel mode 1. See Fig. 13(b). The constrained parameters are now $\theta_1, |h_0|$ and $|h_1|$. The second derivative at the minimum is reduced to SNR^{2r-1}. This can be seen from the figure. Suppose ω^* is increased slightly to $\omega^* + \epsilon$. To first order, the $\tilde{h}_1 e^{j\omega^*}$ phasor will rotate counterclockwise slightly such that the tip of this arrow moves downwards by $2|h_2|\epsilon$. The $\tilde{h}_2 e^{j2\omega^*}$ phasor will rotate clockwise slightly such that the tip of this arrow moves upwards by $2|h_2|\epsilon$. The nett effect is that the sum of the three phasors will remain unchanged. For $0 < r < 1/4$, the channel coefficients align into channel mode 2. See Fig. 14(c). The constrained parameters are $\theta_1, |h_0|, |h_1|$, and $|h_2|$. In the figure, mode 2 seems similar to mode 1. The difference is that all channel coefficients fade. □

Let $\tilde{\boldsymbol{h}} = [\tilde{h}_0, \ldots, \tilde{h}_{L-1}] \in \mathbb{C}^L$. For a given realization of the channel coefficients $h_{u(r)+1}, \ldots, h_{L-1}$, the $u(r)$th channel mode is characterized by the $L-u(r)-1$-dimensional subspace of $\mathbb{C}^L$ defined by the system of $L-u(r)-1$ linear constraints on the variables $\tilde{h}_0, \cdots \tilde{h}_{u(r)}$ given in (51) for $u(r) < L-1$ and (52) for $u(r) = L-1$. An error occurs when $\boldsymbol{h}$ lands very close to this subspace.

Continue Proof

More precisely, define

$$\epsilon(r) \triangleq \begin{cases} \mathsf{SNR}^{-r}, & \text{for } r \leq 1/2 \\ \mathsf{SNR}^{1-r}, & \text{for } r \geq 1/2. \end{cases}$$

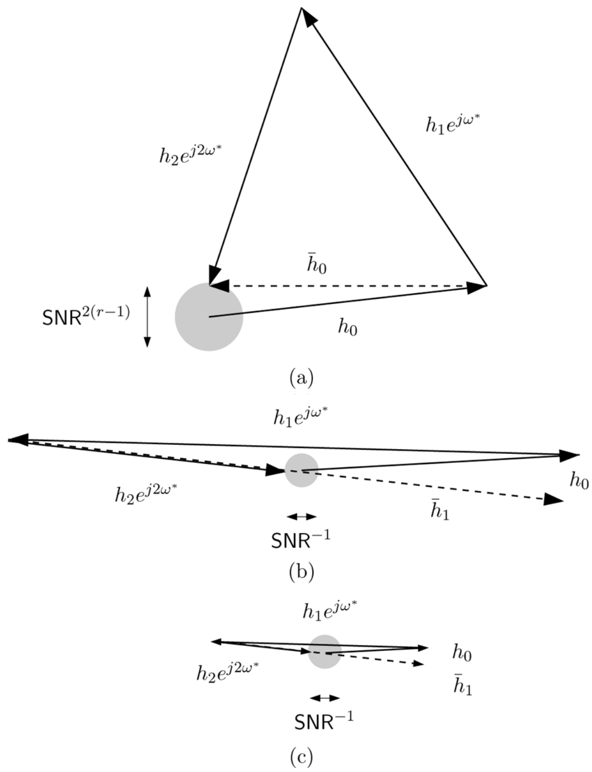

Fig. 13. When an error occurs at high-SNR, the channel coefficients align in particular configurations in the complex plane. In this example, $L = 3$: (a) corresponds to channel mode 0 which occurs for $1/2 < r < 1$; (b) corresponds to channel mode 1 which occurs for $1/4 < r < 1/2$; and (c) corresponds to channel mode 2 which occurs for $0 < r < 1/4$.

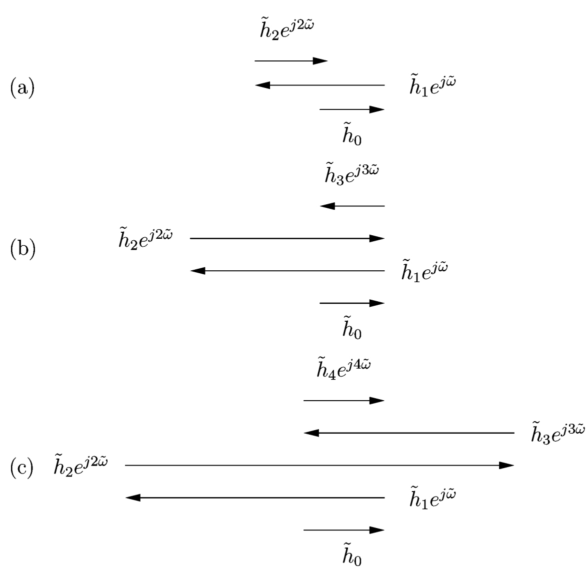

Fig. 14. Illustration of the $L-2$th channel mode for (a) $L = 3$, (b) $L = 4$, and (c) $L = 5$.

We will show that an error occurs whenever

$$\{|h_l - \tilde{h}_l|^2 < \epsilon(r)^{1-2l}\mathsf{SNR}^{r-1}\}$$

for $l = 0, \dots, u(r)$.

Example: $L = 3$

In Fig. 14, the $L-2$th channel modes are drawn for various values of L. Using the same arguments that were used in the preceding example, one may verify pictorially that small perturbations of ω around ω^* have no effect up to order $L-2$. □

We start by defining the event

$$\mathcal{O}_r \triangleq \bigcap_{l=0}^{u(r)} \{|h_l - \bar{h}_l|^2 < \epsilon(r)^{1-2l}\mathsf{SNR}^{r-1}\}$$

which deviates from the previous discussion only by the replacement of $\tilde{h}_l$ by $\bar{h}_l$. Also define

$$\mathcal{Q}_r \triangleq \{||h_0| - |h_0^*||^2 > \epsilon(r)\mathsf{SNR}^{r-1}/2\}$$

with

$$h_0^* \triangleq -\sum_{l=1}^{L-1} h_l e^{j\omega^* l}.$$

The set $\mathcal{O}_r$ is the error event that we analyze to generate the L-part bound of Theorem III.9. The set $\mathcal{Q}_r$ is only introduced to simplify the proof as

$$||h_0| - |h_0^*||^2 \le |H(\omega^*)|^2 \le ||h_0| - |\bar{h}_0||^2. \tag{53}$$

It does not represent anything fundamental. We can lower-bound the error probability by only integrating over those $\boldsymbol{h} \in \mathcal{O}_r \cap \mathcal{Q}_r$.

$$P_e \ge \int_{\boldsymbol{h}\in\mathcal{O}_r\cap\mathcal{Q}_r} \exp\left(-\frac{|A_x[0]|^2\mathsf{SNR}^{1-r}}{\mathbb{E}|z[k]|^2}\right) p(\boldsymbol{h})d\boldsymbol{h}.$$

We can express $|A_x[0]|^2$ in terms of frequency domain quantities using (14)

$$|A_x[0]|^2 = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\mathsf{SNR}|H(\omega)|^2}{1+\mathsf{SNR}|H(\omega)|^2} d\omega\right)^2$$

and thus $|A_x[0]|^2 \le 1$. $\mathbb{E}|z[k]|^2$ can be expressed in terms of the $H(\omega)$ as

$$\mathbb{E}|z[k]|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\mathsf{SNR}^2|H(\omega)|^2}{(1+\mathsf{SNR}|H(\omega)|^2)^2} d\omega.$$

For $\boldsymbol{h} \in \mathcal{O}_r \cap \mathcal{Q}_r$, we have $1 < 2\mathsf{SNR}|H(\omega^*)|^2 \le 2\mathsf{SNR}|H(\omega)|^2$ for all ω by (53) and thus

$$\begin{aligned}\mathbb{E}|z[k]|^2 &\ge \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\mathsf{SNR}^2|H(\omega)|^2}{(3\mathsf{SNR}|H(\omega)|^2)^2} d\omega \\ &\doteq \int_{-\pi}^{\pi} \frac{d\omega}{|H(\omega)|^2} \\ &= \int_{-\pi}^{\pi} \frac{d\nu}{|H(\tilde{\omega}+\nu)|^2}.\end{aligned}$$

The last step follows from the periodicity of $H(\omega)$. We now bound $|H(\omega)|$ in the vicinity of $|H(\tilde{\omega})|$.

$$\begin{aligned}|H(\tilde{\omega}+\nu)|^2 &= \left|\sum_{l=0}^{L-1} h_l e^{jl(\tilde{\omega}+\nu)}\right|^2 \\ &= \left|\sum_{l=0}^{L-1} h_l e^{jl(\tilde{\omega})} \left(\sum_{i=0}^{l} \binom{l}{i} (e^{j\nu}-1)^i\right)\right|^2 \\ &= \left|\sum_{l=0}^{L-1}\sum_{i=0}^{l} \binom{l}{i} h_l e^{jl(\tilde{\omega})}(e^{j\nu}-1)^i\right|^2\end{aligned}$$

$$
\begin{aligned}
&= \left| \sum_{i=0}^{L-1} \sum_{l=i}^{L-1} \binom{l}{i} h_l e^{jl(\tilde{\omega})} (e^{j\nu} - 1)^i \right|^2 \\
&\dot{\leq} \sum_{i=0}^{L-1} \left| \sum_{l=i}^{L-1} \binom{l}{i} h_l e^{jl(\tilde{\omega})} \right|^2 |e^{j\nu} - 1|^{2i} \\
&= \sum_{i=0}^{L-1} \left| \sum_{l=i}^{L-1} \binom{l}{i} h_l e^{j(l-i)(\tilde{\omega})} \right|^2 |e^{j\nu} - 1|^{2i} \\
&\leq \sum_{i=0}^{L-1} \left| h_i + \sum_{l=i+1}^{L-1} \binom{l}{i} h_l e^{j(l-i)(\tilde{\omega})} \right|^2 \nu^{2i} \\
&= \sum_{i=0}^{L-1} |h_i - \bar{h}_i|^2 \nu^{2i}.
\end{aligned}
$$

We have used Lemma VII.1 in the fourth step and the fact that $|e^{jX} - 1|^2 \leq X^2$ in the second last. Substituting this bound into the previous equation yields

$$
\begin{aligned}
\mathbb{E}|z[k]|^2 &\dot{\geq} \int_{-\pi}^{\pi} \frac{d\nu}{\sum_{i=0}^{L-1} |h_i - \bar{h}_i|^2 \nu^{2i}} \\
&\geq \frac{\epsilon(r)}{\sum_{i=0}^{L-1} |h_i - \bar{h}_i|^2 \epsilon(r)^{2i}} \\
&\geq \frac{\epsilon(r)}{\sum_{i=0}^{L-1} \epsilon(r) \mathsf{SNR}^{r-1}} \\
&\doteq \mathsf{SNR}^{1-r}.
\end{aligned}
$$

The second step follows from the fact that the function within the integral is monotone decreasing for $\nu > 0$ and monotone increasing for $\nu < 0$. This means

$$
\frac{|A_x[0]|^2 \mathsf{SNR}^{1-r}}{\mathbb{E}|z[k]|^2} \dot{\leq} 1.
$$

Thus

$$
\begin{aligned}
P_e &\dot{\geq} \int_{\boldsymbol{h} \in \mathcal{O}_r \cap \mathcal{Q}_r} p(\boldsymbol{h}) d\boldsymbol{h} \\
&= \Pr(\boldsymbol{h} \in \mathcal{O}_r \cap \mathcal{Q}_r) \\
&= \Pr\left(\bigcap_{l=0}^{u(r)} \{ |h_l - \bar{h}_l|^2 < \epsilon(r)^{(1-2l)} \mathsf{SNR}^{r-1} \}, \right. \\
&\qquad \left. ||h_0| - |h_0^*||^2 > \epsilon(r)^{(1-2l)} \mathsf{SNR}^{r-1}/2 \right) \\
&\dot{\geq} \Pr\left(\bigcap_{l=0}^{u(r)} \{ |h_l - \tilde{h}_l|^2 + |\tilde{h}_l - \bar{h}_l|^2 \right. \\
&\qquad < \epsilon(r)^{(1-2l)} \mathsf{SNR}^{r-1} \}, \\
&\qquad \left. ||h_0| - |h_0^*||^2 > \epsilon(r)^{(1-2l)} \mathsf{SNR}^{r-1}/2 \right).
\end{aligned}
$$

Next we show that for any $\alpha > 0$ and for each $l \in \{0, \ldots, L-1\}$

$$
\left\{ |h_l - \tilde{h}_l|^2 \dot{<} \mathsf{SNR}^{-\alpha} \right\} \Rightarrow \left\{ |\tilde{h}_l - \bar{h}_l|^2 \dot{<} \mathsf{SNR}^{-\alpha} \right\}. \quad (54)
$$

Let $h_l = \tilde{h}_l + \beta_l$ with $|\beta_l| < \mathsf{SNR}^{-\alpha}$. Then

$$
\begin{aligned}
|\tilde{h}_l - \bar{h}_l|^2 &= \left| \tilde{h}_l + \sum_{i=l+1}^{L-1} \binom{i}{l} h_i e^{j(i-l)\tilde{\omega}} \right|^2 \\
&= \left| - \sum_{i=l+1}^{L-1} \binom{i}{l} \tilde{h}_i e^{j(i-l)\tilde{\omega}} \right. \\
&\qquad \left. + \sum_{i=l+1}^{L-1} \binom{i}{l} (\tilde{h}_i + \beta_l) e^{j(i-l)\tilde{\omega}} \right|^2 \\
&= |\beta_l|^2 \left(\sum_{i=l+1}^{L-1} \binom{i}{l} \right)^2 \\
&\dot{<} \mathsf{SNR}^{-\alpha}.
\end{aligned}
$$

Implication (54) enables us to bound P_e using events that are conditionally independent of each other given $h_{u(r)+1}, \ldots, h_{L-1}$ and θ_0. This is done in (55), shown at the top of the following page. We now complete the bound, treating each regime separately.

$1/2 < r < 1$

In this regime $\epsilon(r) = \mathsf{SNR}^{r-1}$. Let $\delta > 0$ be an arbitrarily small constant. The bound for P_e is given in (56), also on the following page.The second step follows from two observations. The first is the Gaussian tail diminishes exponentially fast so that the constraints on the $l > 0$ terms vanish. The second is that the constraint involving $|h_0^*|$ does not have any asymptotic effect (this is evident from the presence of the constraint involving $|\tilde{h}_0|$). The third step follows from the fact that $\theta_0 = \angle \tilde{h}_0$, the fourth from Lemma VII.5. The last follows from the fact that $\{\delta < |\tilde{h}_0|^2 < \delta^{-1}\}$ is an order SNR^0 event.

$\frac{1}{2(L-1)} < r < \frac{1}{2}$

In this regime, $\epsilon(r) = \mathsf{SNR}^{-r}$. The bound is given in (57) which continues through to (58) (see the subsequent page). The justification for the second step is the same as for the case $1/2 < r < 1$. The third step follows from the conditional independence of the random variables $\tilde{h}_0, \ldots, \tilde{h}_{L-1}$ given $h_{u(r)+1}, \ldots, h_{L-1}$ and θ_0, and the fact $\theta_0 = \angle \tilde{h}_0$. The fifth step follows from Lemma VII.5. The sixth follows from the fact that the $|h_l|^2$ are exponential random variables and thus have a monotone decreasing probability density function (pdf). The third last step follows from the independence of the h_l and θ_l and from the fact that $\bigcap_{l=0}^{u(r)} \{\delta < |\tilde{h}_l|^2 < \delta^{-1}\}$ is an order SNR^0 event. The second last step follows from Lemma VII.4 and the fact that θ_l is uniformly distributed on $[-\pi, \pi)$.

$0 < r < \frac{1}{2(L-1)}$

Again, in this regime $\epsilon(r) = \mathsf{SNR}^{-r}$. The bound is given in (59) (see the subsequent page). The justification for most of these steps is the same as for the case of $1/2(L-1) < r < 1/2$. The fifth step follows from the fact that $|\tilde{h}_l| \doteq \mathsf{SNR}^{2(L-1)r-1}$

$$\begin{aligned}
P_e &\dot{\geq} \Pr\left(\bigcap_{l=0}^{u(r)}\left\{|h_l-\tilde{h}_l|^2<\epsilon(r)^{(1-2l)}\mathsf{SNR}^{r-1}\right\}, ||h_0|-|h_0^*||^2>\epsilon(r)^{(1-2l)}\mathsf{SNR}^{r-1}/2\right)\\
&=\Pr\left(\bigcap_{l=0}^{u(r)}\left\{\left||h_l|-|\tilde{h}_l|+|\tilde{h}_l|\left(1-e^{j(\angle\tilde{h}_l-\angle h_l)}\right)\right|^2<\epsilon(r)^{(1-2l)}\mathsf{SNR}^{r-1}\right\},\right.\\
&\qquad\left. ||h_0|-|h_0^*||^2>\epsilon(r)^{(1-2l)}\mathsf{SNR}^{r-1}/2\right)\\
&\dot{\geq} \Pr\left(\bigcap_{l=0}^{u(r)}\left\{||h_l|-|\tilde{h}_l||^2+|\tilde{h}_l|^2|\angle\tilde{h}_l-\angle h_l|^2<\epsilon(r)^{(1-2l)}\mathsf{SNR}^{r-1}\right\},\right.\\
&\qquad\left. ||h_0|-|h_0^*||^2>\epsilon(r)^{(1-2l)}\mathsf{SNR}^{r-1}/2\right)\\
&\geq \Pr\left(\bigcap_{l=0}^{u(r)}\left\{||h_l|-|\tilde{h}_l||^2<\epsilon(r)^{(1-2i)}\mathsf{SNR}^{r-1}, |\theta_l-\angle\tilde{h}_l|^2<|\tilde{h}_l|^{-2}\epsilon(r)^{(1-2l)}\mathsf{SNR}^{r-1}\right\},\right.\\
&\qquad\left. ||h_0|-|h_0^*||^2>\epsilon(r)^{(1-2i)}\mathsf{SNR}^{r-1}/2\right). \qquad (55)
\end{aligned}$$

$$\begin{aligned}
P_e &\dot{\geq} \Pr\left(\bigcap_{l=0}^{u(r)}\left\{||h_l|-|\tilde{h}_l||^2<\mathsf{SNR}^{2(1-l)(r-1)},\quad |\theta_l-\angle\tilde{h}_l|^2<|\tilde{h}_l|^{-2}\mathsf{SNR}^{2(1-l)(r-1)}\right\},\right.\\
&\qquad\left. ||h_0|-|h_0^*||^2>\mathsf{SNR}^{2(1-l)(r-1)}/2\right)\\
&\doteq \Pr\left(||h_0|-|\tilde{h_0}||^2<\mathsf{SNR}^{2(r-1)}, |\theta_0-\angle\tilde{h}_0|^2<|\tilde{h}_0|^{-2}\mathsf{SNR}^{2(1-l)(r-1)}\right)\\
&=\Pr\left(||h_0|-|\tilde{h_0}||^2<\mathsf{SNR}^{2(r-1)}\right)\\
&\doteq \Pr\left(||h_0|^2-|\tilde{h_0}|^2|^2<\mathsf{SNR}^{2(r-1)}(|h_0|^2+|\tilde{h_0}|^2)\right)\\
&=\int \Pr(||h_0|^2-|\tilde{h_0}|^2|^2 \quad <\mathsf{SNR}^{2(r-1)}(|h_0|^2+|\tilde{h_0}|^2)|\boldsymbol{h},\theta_0)p(\boldsymbol{h})p(\theta_0)d\boldsymbol{h}d\theta_0\\
&\geq \int_{\delta<|\tilde{h}_0|^2<\delta^{-1}} \Pr(||h_0|^2-|\tilde{h_0}|^2|^2<\mathsf{SNR}^{2(r-1)}(|h_0|^2+|\tilde{h_0}|^2)|\boldsymbol{h},\theta_0)p(\boldsymbol{h})p(\theta_0)d\boldsymbol{h}d\theta_0\\
&\geq \int_{\delta<|\tilde{h}_0|^2<\delta^{-1}} \Pr(||h_0|^2-|\tilde{h_0}|^2|<\mathsf{SNR}^{r-1}\delta|\boldsymbol{h},\theta_0)p(\boldsymbol{h})p(\theta_0)d\boldsymbol{h}d\theta_0\\
&\doteq \mathsf{SNR}^{r-1}P(\delta<|\tilde{h}_0|^2<\delta^{-1})\\
&\doteq \mathsf{SNR}^{r-1}. \qquad (56)
\end{aligned}$$

for all l. The second last step follows from Lemmas VII.4 and VII.3. Combining these bounds we have

$$d(r)\leq\min\left\{1-r, L\left(1-\frac{1}{\frac{L}{(L-1)(L+1)}}r\right), \min_{k\in\{1,\ldots,L-2\}}(k+1/2)\left(1-\frac{1}{\frac{k+1/2}{k(k+1)}}r\right)\right\}$$

for $0<r<1$. □

E. DFE-ZF Equalizer

Proof of Theorem III.10: Ignoring error propagation, the effective SNR at the output of this equalizer is (see [3, p. 475])

$$\mathsf{SNR}_{\text{eff}}^{\text{DFE}-\text{ZF}}=e^{\frac{1}{2\pi}\int_0^{2\pi}\log(\mathsf{SNR}|H(\omega)|^2)d\omega}.$$

Let $X:[0,2\pi)\to\mathbb{R}$ and define the set

$$\mathcal{U}_\epsilon(X)\triangleq\{\omega\in[0,2\pi)\big|X(\omega)\geq\epsilon\}.$$

$$
\begin{aligned}
P_e \quad &\dot{\geq} \Pr\left(\bigcap_{l=0}^{u(r)} \left\{ ||h_l| - |\tilde{h_l}||^2 < \mathsf{SNR}^{2lr-1},\ |\theta_l - \angle\tilde{h}_l|^2 < |\tilde{h}_l|^{-2}\mathsf{SNR}^{2lr-1} \right\},\right. \\
&\left. ||h_0| - |h_0^*||^2 > \mathsf{SNR}^{2lr-1}/2 \right) \\
&\doteq \int \Pr\left(\bigcap_{l=0}^{u(r)} \left\{ ||h_l| - |\tilde{h_l}||^2 < \mathsf{SNR}^{2lr-1}, |\theta_l - \angle\tilde{h}_l|^2 < |\tilde{h}_l|^{-2}\mathsf{SNR}^{2lr-1} \right\} |\boldsymbol{h}^L_{u(r)+1}, \theta_0 \right) \\
&p\left(\boldsymbol{h}^L_{u(r)+1}\right) p(\theta_0) d\boldsymbol{h}^L_{u(r)+1} d\theta_0 \\
&= \int \Pr\left(||h_0| - |\tilde{h_0}||^2 < \mathsf{SNR}^{-1} \middle| \boldsymbol{h}^L_{u(r)+1}, \theta_0 \right) \\
&\prod_{l=1}^{u(r)} \Pr\left(||h_l| - |\tilde{h_l}||^2 < \mathsf{SNR}^{2lr-1}, |\theta_l - \angle\tilde{h}_l|^2 < |\tilde{h}_l|^{-2}\mathsf{SNR}^{2lr-1} \middle| \boldsymbol{h}^L_{u(r)+1}, \theta_0 \right) \\
&p\left(\boldsymbol{h}^L_{u(r)+1}\right) p(\theta_0) d\boldsymbol{h}^L_{u(r)+1} d\theta_0 \\
&\geq \int_{\bigcap_{l=0}^{u(r)} \delta < |\tilde{h}_l|^2 < \delta^{-1}} \Pr\left(||h_0| - |\tilde{h_0}||^2 < \mathsf{SNR}^{-1} \middle| \boldsymbol{h}^L_{u(r)+1}, \theta_0 \right) \\
&\prod_{l=1}^{u(r)} \Pr\left(||h_l| - |\tilde{h_l}||^2 < \mathsf{SNR}^{2lr-1}, |\theta_l - \angle\tilde{h}_l|^2 < |\tilde{h}_l|^{-2}\mathsf{SNR}^{2lr-1} \middle| \boldsymbol{h}^L_{u(r)+1}, \theta_0 \right) \\
&p\left(\boldsymbol{h}^L_{u(r)+1}\right) p(\theta_0) d\boldsymbol{h}^L_{u(r)+1} d\theta_0 \\
&\doteq \int_{\bigcap_{l=0}^{u(r)} \delta < |\tilde{h}_l|^2 < \delta^{-1}} \Pr\left(||h_0|^2 - |\tilde{h_0}|^2|^2 < \mathsf{SNR}^{-1}(|h_0|^2 + |\tilde{h_0}|^2) \middle| \boldsymbol{h}^L_{u(r)+1}, \theta_0 \right) \\
&\prod_{l=1}^{u(r)} \Pr\left(||h_l|^2 - |\tilde{h_l}|^2|^2 < \mathsf{SNR}^{2lr-1}(|h_l|^2 + |\tilde{h_l}|^2), |\theta_l - \angle\tilde{h}_l|^2 < |\tilde{h}_l|^{-2}\mathsf{SNR}^{2lr-1} \middle| \boldsymbol{h}^L_{u(r)+1}, \theta_0 \right) \\
&p\left(\boldsymbol{h}^L_{u(r)+1}\right) p(\theta_0) d\boldsymbol{h}^L_{u(r)+1} d\theta_0 \\
&\geq \int_{\bigcap_{l=0}^{u(r)} \delta < |\tilde{h}_l|^2 < \delta^{-1}} \Pr\left(||h_0|^2 - \delta^{-1}|^2 < \mathsf{SNR}^{-1}\delta \middle| \boldsymbol{h}^L_{u(r)+1}, \theta_0 \right) \\
&\prod_{l=1}^{u(r)} \Pr\left(||h_l|^2 - \delta^{-1}|^2 < \mathsf{SNR}^{2lr-1}\delta, |\theta_l - \angle\tilde{h}_l|^2 < \delta\mathsf{SNR}^{2lr-1} \middle| \boldsymbol{h}^L_{u(r)+1}, \theta_0 \right) \\
&p\left(\boldsymbol{h}^L_{u(r)+1}\right) p(\theta_0) d\boldsymbol{h}^L_{u(r)+1} d\theta_0.
\end{aligned}
\tag{57}
$$

$$
\begin{aligned}
&= \Pr\left(\bigcap_{l=0}^{u(r)} \{\delta < |\tilde{h}_l|^2 < \delta^{-1}\}\right) \Pr\left(||h_0|^2 - \delta^{-1}|^2 < \mathsf{SNR}^{-1}\delta \right) \\
&\prod_{l=1}^{u(r)} \Pr(||h_l|^2 - \delta^{-1}|^2 < \mathsf{SNR}^{2lr-1}\delta, |\theta_l - \angle\tilde{h}_l|^2 < \delta\mathsf{SNR}^{2lr-1}) \\
&\doteq \prod_{l=0}^{u(r)} \Pr(||h_l|^2 - \delta^{-1}|^2 < \mathsf{SNR}^{2lr-1}) \\
&\quad \times \prod_{l=1}^{u(r)} \Pr(|\theta_l - \angle\tilde{h}_l|^2 < \mathsf{SNR}^{2lr-1}) \\
&\doteq \mathsf{SNR}^{\sum_{l=0}^{u(r)}(lr-1/2)} \mathsf{SNR}^{\sum_{l=1}^{u(r)}(lr-1/2)} \\
&= \mathsf{SNR}^{u(r)(u(r)+1)r - u(r) - 1/2}.
\end{aligned}
\tag{58}
$$

$$
\begin{aligned}
P_e &\dot{\geq} \Pr\left(\bigcap_{l=0}^{u(r)}\{||h_l|-|\tilde{h}_l||^2<\mathsf{SNR}^{2lr-1},\ |\theta_l-\angle\tilde{h}_l|^2<|\tilde{h}_l|^{-2}\mathsf{SNR}^{2lr-1}\},\ ||h_0|-|h_0^*||^2>\mathsf{SNR}^{-1}/2\right)\\
&\doteq \Pr\left(\bigcap_{l=0}^{u(r)}\{||h_l|-|\tilde{h}_l||^2<\mathsf{SNR}^{2lr-1},|\theta_l-\angle\tilde{h}_l|^2<|\tilde{h}_l|^{-2}\mathsf{SNR}^{2lr-1}\}\right)\\
&= \int \Pr(||h_0|-|\tilde{h_0}||^2<\mathsf{SNR}^{-1}|\theta_0,\theta_{L-1})\\
&\prod_{l=1}^{L-2}\Pr(||h_l|-|\tilde{h}_l||^2<\mathsf{SNR}^{2lr-1},|\theta_l-\angle\tilde{h}_l|^2<|\tilde{h}_l|^{-2}\mathsf{SNR}^{2lr-1}|\theta_0,\theta_{L-1})\\
&\Pr\left(\left||h_{L-1}|-\mathsf{SNR}^{(L-1)r-1/2}\right|^2<\mathsf{SNR}^{2(L-1)r-1}\right)p(\theta_0)p(\theta_{L-1})d\theta_0 d\theta_{L-1}\\
&\doteq \int\prod_{l=0}^{L-2}\Pr(||h_l|^2-|\tilde{h}_l|^2|^2<\mathsf{SNR}^{2lr-1}(|h_l|^2+|\tilde{h}_l|^2)|\theta_0,\theta_{L-1})\\
&\prod_{l=1}^{L-2}\Pr(|\theta_l-\angle\tilde{h}_l|^2<|\tilde{h}_l|^{-2}\mathsf{SNR}^{2lr-1}|\theta_0,\theta_{L-1})\\
&\Pr\left(|h_{L-1}|^2<\mathsf{SNR}^{2(L-1)r-1}\right)p(\theta_0)p(\theta_{L-1})d\theta_0 d\theta_{L-1}\\
&\geq \int\prod_{l=0}^{L-2}\Pr(||h_l|^2-|\tilde{h}_l|^2|<\mathsf{SNR}^{(l+L-1)r-1}|\theta_0,\theta_{L-1})\\
&\prod_{l=1}^{L-2}\Pr\left(|\theta_l-\angle\tilde{h}_l|<\mathsf{SNR}^{(l-L+1)r}|\theta_0,\theta_{L-1}\right)\\
&\Pr\left(|h_{L-1}|^2<\mathsf{SNR}^{2(L-1)r-1}\right)p(\theta_0)p(\theta_{L-1})d\theta_0 d\theta_{L-1}\\
&\doteq \mathsf{SNR}^{2(L-1)r-1}\mathsf{SNR}^{\sum_{l=0}^{L-2}(l+L-1)r-1}\mathsf{SNR}^{\sum_{l=1}^{L-2}(l-L+1)r}\\
&= \mathsf{SNR}^{(L-1)(L+1)r-L}.
\end{aligned}
\tag{59}
$$

Also define $\bar{H}(\omega)\triangleq H(\omega)/||\boldsymbol{h}||^2$. Then by Lemma VII.6

$$
\begin{aligned}
P_e &\doteq \Pr\left(\mathsf{SNR}_{\text{eff}}^{DFE-ZF}<\mathsf{SNR}^r\right)\\
&= \Pr\left(e^{\frac{1}{2\pi}\int_0^{2\pi}\log(\mathsf{SNR}|H(\omega)|^2)d\omega}<\mathsf{SNR}^r\right)\\
&= \Pr\left(\frac{1}{2\pi}\int_0^{2\pi}\log(\mathsf{SNR}|H(\omega)|^2)d\omega<R\right)\\
&= \Pr\left(\frac{1}{2\pi}\int_0^{2\pi}\log(\mathsf{SNR}||\boldsymbol{h}||^2|\bar{H}(\omega)|^2)d\omega<R\right)\\
&\leq \Pr((1-|\mathcal{U}_\epsilon(\bar{H})|)\cdot\log(\mathsf{SNR}||\boldsymbol{h}||^2\epsilon)<R)\\
&\leq \Pr\left(\left(1-\sup_{\boldsymbol{h}\in\mathbb{C}^L}|\mathcal{U}_\epsilon(\bar{H})|\right)\cdot\log(\mathsf{SNR}||\boldsymbol{h}||^2\epsilon)<R\right)\\
&= \Pr\left(||\boldsymbol{h}||^2<\frac{1}{\epsilon}\cdot\mathsf{SNR}^{\frac{r}{1-\sup_{\boldsymbol{h}\in\mathbb{C}^L}|\mathcal{U}_\epsilon(\bar{H})|}-1}\right)\\
&\doteq \frac{1}{\epsilon^L}\cdot\mathsf{SNR}^{-L\left(1-\frac{r}{1-\sup_{\boldsymbol{h}\in\mathbb{C}^L}|\mathcal{U}_\epsilon(\bar{H})|}\right)}.
\end{aligned}
$$

In the fourth step we have used a lower bound illustrated in Fig. 15. In the last step we have invoked Lemma VII.2 from the Appendix. To complete the proof we must show $|\mathcal{U}_\epsilon(\bar{H})|\to 0$ uniformly as $\epsilon\to 0$.

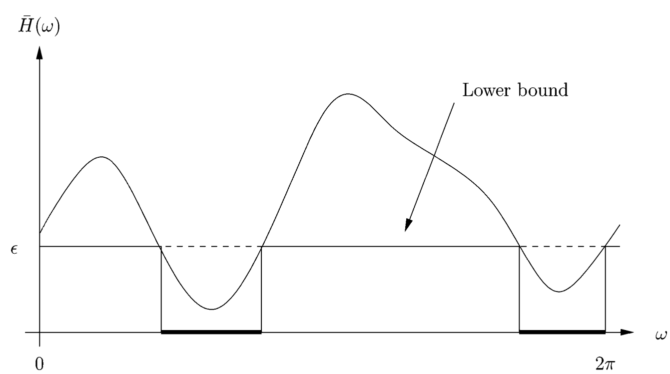

Fig. 15. An illustration of $\bar{H}(\omega)$ for a particular channel realization. In this example $L=3$. For the value of ϵ indicated, $|\mathcal{U}_\epsilon(\bar{H})|$ is equal to the sum of the widths of the two darkened intervals (roughly $\pi/2$). Notice that if ϵ decreases to zero, $|\mathcal{U}_\epsilon(\bar{H})|$ will also decrease to zero.

For a given $\boldsymbol{h}$, $\bar{H}(\omega)$ is an analytic, nonzero function with a finite number of minima so by the identity theorem it is zero only at a finite number of points. Thus, for each $\boldsymbol{h}$, $|\mathcal{U}_\epsilon(\bar{H})|\to 0$ as $\epsilon\to 0$. This establishes pointwise convergence. Demonstrating this convergence is in fact uniform requires a bit more work. First observe

$$\sup_{\boldsymbol{h}\in\mathbb{C}^L}|\mathcal{U}_\epsilon(\bar{H})|=\max_{\boldsymbol{h}\in\mathcal{S}}|\mathcal{U}_\epsilon(H)|.$$

where the set $\mathcal{S}$ is defined in (32). This set is compact. Second, observe that $|\mathcal{U}_\epsilon(H)|$ is defined by the roots of the polynomial $H(\omega) - \epsilon$, the coefficients of which are continuous functions of $\boldsymbol{h}$. As the roots of a polynomial vary continuously with the coefficients, $|\mathcal{U}_\epsilon(H)|$ is a continuous function of $\boldsymbol{h}$. Third, observe that $|\mathcal{U}_\epsilon(H)|$ is monotone decreasing in ϵ.

Combining these observations we have that for every $\boldsymbol{h}$ in the compact set $\mathcal{S}$, $|\mathcal{U}_\epsilon(H)|$ is a continuous monotone decreasing function with pointwise limit 0. Thus, $|\mathcal{S}_\epsilon(H)|$ converges uniformly to 0 as $\epsilon \to 0$ (Dini's theorem).

Thus, $\lim_{\epsilon\to 0} \sup_{\boldsymbol{h}\in\mathbb{C}^L} |\mathcal{U}_\epsilon(\bar{H})| = 0$ and by making ϵ arbitrarily small we have

$$P_e \dot{\leq} \mathsf{SNR}^{-L(1-r)}.$$

The outage bound in Theorem III.2 demonstrates that this upper bound is tight.

We now incorporate error propagation. We have shown that the probability of decoding symbol $x[k]$ incorrectly given the symbols $x[0],\ldots,x[k-1]$ were decoded correctly is $\doteq \mathsf{SNR}^{-L(1-r)}$. The probability that one of $x[0],\ldots,x[k-1]$ is decoded incorrectly is therefore $\doteq 1-(1-\mathsf{SNR}^{-L(1-r)})^k \leq k\mathsf{SNR}^{-L(1-r)} \doteq \mathsf{SNR}^{-L(1-r)}$. Thus, error propagation does not affect the asymptotic error probability and $d(r) = L(1-r)$. □

Proof of Theorem III.5: The error probability of any equalizer upper bounds the error probability of the MLSD. Theorem III.10 states that for the no trailing zeros transmission scheme the DFE-ZF achieves $d(r) = L(1-r)$. This provides a lower bound to the tradeoff curve achievable by the infinite length MLSD. The fact that it also matches the outage bound establishes the desired result. □

Proof of Theorem III.3: The existence of a communication scheme that achieves this outage bound in the limit $N \to \infty$ (e.g., Theorem III.4), in conjunction with Theorem III.2, proves this theorem. □

V. Universality of Equalizers

Throughout this work the channel taps were assumed to be Rayleigh distributed, though the question arises as to whether the results presented here hold more generally. The conventional formulation of this problem is to ask if communication will be error-free whenever the channel is not in outage. Communication schemes possessing this property are called *universal*. We ask this question with regards to those equalizers that we have demonstrated achieve the optimal tradeoff for Rayleigh statistics, in the limit of infinite block length.

From (26) we see that at high SNR, the outage event is $\{||\boldsymbol{h}||^2 < \mathsf{SNR}^{r-1}\}$. Notice that in all proofs for those equalizers that achieve the optimal tradeoff in the limit of infinite block length, the error event is always contained within an event of the form $\{||\boldsymbol{h}||^2 < \mathsf{SNR}^{r-1}\}$. Thus, we see that these equalizers are indeed universal.

VI. Conclusion

With continuous transmission, complex receiver architectures such as MLSD (Viterbi algorithm) or DFE (linear equalization with successive cancellation) are required to achieve reliable communication over multipath fading channels at high SNR. Linear equalizers fall far short of optimal diversity versus multiplexing performance. OFDM systems utilize a prefix (of duration matching the coherence time of the channel) to obtain the optimal diversity–multiplexing tradeoff in limit of large block length, but they require an additional linear encoding architecture at the transmitter. By merely inserting a period of silence every so often, equal in duration to the coherence time of the channel, the optimal diversity–multiplexing tradeoff can be achieved using only a linear equalizer at the receiver. Thus, in the high-SNR limit, a simple precoding strategy can facilitate reliable communication with low complexity.

Appendix

Lemma VII.1: For any $X, Y \in \mathbb{C}$

$$|X+Y|^2 \dot{\leq} |X|^2 + |Y|^2.$$

Proof:

$$\begin{aligned}|X+Y|^2 &\leq 2(|X|^2+|Y|^2)\\ &\doteq |X|^2+|Y|^2.\end{aligned}$$ ■

Lemma VII.2: For any constant $\alpha > 0$

$$\Pr(||\boldsymbol{h}||^2 < \mathsf{SNR}^{-\alpha}) \doteq \mathsf{SNR}^{-L\alpha}.$$

Proof: $||\boldsymbol{h}||^2$ is the sum of L independent exponential random variables with mean 1 and thus has a gamma distribution with mean L

$$\begin{aligned}\Pr(||\boldsymbol{h}||^2 < \mathsf{SNR}^{-\alpha}) &= \frac{1}{(L-1)!}\int_0^{\mathsf{SNR}^{-\alpha}} x^{L-1}e^{-x}dx\\ &\doteq \int_0^{\mathsf{SNR}^{-\alpha}} x^{L-1}dx\\ &\doteq \mathsf{SNR}^{-L\alpha}.\end{aligned}$$ ■

Corollary VII.3: For any integer l and constant $\alpha > 0$

$$\Pr(|h_l|^2 < \mathsf{SNR}^{-\alpha}) \doteq \mathsf{SNR}^{-\alpha}.$$

Lemma VII.4: For any integer l and constants $\alpha > 0$ and $X > 0$

$$\Pr(||h_l|^2 - X| < \mathsf{SNR}^{-\alpha}) \doteq \mathsf{SNR}^{-\alpha}.$$

Proof: $|h_l|^2$ is an exponential random variable with mean 1 so

$$\begin{aligned}\mathsf{SNR}^{-\alpha}e^{-(X+\mathsf{SNR}^{-\alpha})} &\leq \Pr(||h_l|^2 - X| < \mathsf{SNR}^{-\alpha})\\ &\leq \mathsf{SNR}^{-\alpha}e^{-(X-\mathsf{SNR}^{-\alpha})}.\end{aligned}$$

Thus

$$\Pr(||h_l|^2 - X| < \mathsf{SNR}^{-\alpha}) \doteq \mathsf{SNR}^{-\alpha}.$$ ■

Lemma VII.5: For any $X \geq 0$ and $Y \geq 0$

$$(X-Y)^2 \doteq \frac{(X^2-Y^2)^2}{X^2+Y^2}.$$

Proof:

$$(X-Y)^2 = \frac{(X^2-Y^2)^2}{(X+Y)^2}$$

$$\Longrightarrow \frac{(X^2-Y^2)^2}{2(X^2+Y^2)} \le (X-Y)^2 \le \frac{(X^2-Y^2)^2}{X^2+Y^2}. \quad \blacksquare$$

Lemma VII.6: Consider the channel in (1) with $x[k]$ chosen from a QAM constellation of SNR^x points uniformly and independently (across k). Suppose a linear equalizer is used at the receiver such that the filtered estimate can be expressed in the form

$$\hat{y}[k] = x[k] + z[k] \tag{60}$$

with $z[k] \sim \mathcal{CN}(0, \mathsf{SNR}/\mathsf{SNR}_{\text{eff}})$. Then for each symbol k the error probability satisfies

$$P_e \doteq \Pr(\mathsf{SNR}_{\text{eff}} < \mathsf{SNR}^x).$$

Proof: The spacing between QAM constellation points in the receiver space is then $\doteq \mathsf{SNR}^{(1-x)/2}$ and the error probability conditioned on the channel realization satisfies

$$\Pr(e|\boldsymbol{h}) \dot{\le} Q\left(\sqrt{\frac{\mathsf{SNR}^{1-x}}{\mathsf{SNR}/\mathsf{SNR}_{\text{eff}}}}\right) \le e^{-\mathsf{SNR}_{\text{eff}}/\mathsf{SNR}^x}.$$

Let $\epsilon > 0$ be an arbitrarily chosen constant. Taking the expectation over the channel

$$\begin{aligned} P_e &= E_{\boldsymbol{h}}[e^{-\mathsf{SNR}_{\text{eff}}/\mathsf{SNR}^x}] \\ &= \int_0^1 \Pr(e^{-\mathsf{SNR}_{\text{eff}}/\mathsf{SNR}^x} > v)dv \\ &= \int_0^\infty e^{-x} \Pr(\mathsf{SNR}_{\text{eff}}/\mathsf{SNR}^x < x)dx \\ &= \int_0^{\mathsf{SNR}^\epsilon} e^{-x} \Pr(\mathsf{SNR}_{\text{eff}}/\mathsf{SNR}^x < x) \\ &\quad + \int_{\mathsf{SNR}^\epsilon}^\infty e^{-x} \Pr(\mathsf{SNR}_{\text{eff}}/\mathsf{SNR}^x < x)dx \\ &\le \int_0^{\mathsf{SNR}^\epsilon} e^{-x} \Pr(\mathsf{SNR}_{\text{eff}}/\mathsf{SNR}^x < \mathsf{SNR}^\epsilon) + \int_{\mathsf{SNR}^\epsilon}^\infty e^{-x}dx \\ &\le \Pr(\mathsf{SNR}_{\text{eff}}/\mathsf{SNR}^x < \mathsf{SNR}^\epsilon) + e^{-\mathsf{SNR}^\epsilon} \\ &\doteq \Pr(\mathsf{SNR}_{\text{eff}} < \mathsf{SNR}^{x+\epsilon}). \end{aligned}$$

Taking $\epsilon \to 0$ yields the desired result. Conversely, an outage occurs if $\{h : I(x[k]; \hat{y}[k]|\boldsymbol{h} = h) < R\}$ and hence

$$\begin{aligned} P_{\text{out}} &= \Pr(\log(1+\mathsf{SNR}_{\text{eff}}) < x \log \mathsf{SNR}) \\ &\doteq \Pr(\mathsf{SNR}_{\text{eff}} < \mathsf{SNR}^x). \end{aligned}$$

By Lemma III.1, $P_e \dot{\ge} P_{\text{out}}$ which establishes the desired result. ■

Lemma VII.7: For the trailing zeros transmission scheme

$$P_e \dot{\ge} \mathsf{SNR}^{-L(1-r')}$$

where r' is defined in (2).

Proof: This is a matched filter bound argument similar to that in the Proof of Theorem III.2. Suppose the receiver has individual access to the L multipaths via the receiver vector in (25). The receiver can recover $y[n]$ via Section IV-A **(Is SectionIV-A OK here?)**. We now use (16) to compute the outage probability

$$\begin{aligned} P_{\text{out}}(R) &= \inf_{p_X} \Pr\left\{h : I(X;Y|\boldsymbol{h}=h) < R\left(1-\frac{L-1}{N}\right)\right\} \\ &= \Pr\left(\log(1+\|\boldsymbol{h}\|^2\mathsf{SNR}) < R\left(1-\frac{L-1}{N}\right)\right) \\ &= \Pr(\|\boldsymbol{h}\|^2 < \mathsf{SNR}^{-1}(\mathsf{SNR}^{r'}-1)) \\ &\doteq \Pr(\|\boldsymbol{h}\|^2 < \mathsf{SNR}^{r'-1}) \\ &\doteq \mathsf{SNR}^{-L(1-r')} \end{aligned}$$

where the last step follows by Lemma VII.2. ■

Lemma VII.8:

$$\Pr\left(||h_0| - |\bar{h_0}||^2 < \mathsf{SNR}^{2(r-1)}, \|\boldsymbol{h}\|^2 < 1\right) \dot{\ge} \mathsf{SNR}^{r-1}.$$

Proof: By Lemma VII.5

$$\begin{aligned} &\Pr\left(||h_0| - |\bar{h_0}||^2 < \mathsf{SNR}^{2(r-1)}, \|\boldsymbol{h}\|^2 < 1\right) \\ &\quad\doteq \Pr\left(||h_0|^2 - |\bar{h_0}|^2|^2 < \mathsf{SNR}^{2(r-1)}(|h_0|^2 + |\bar{h_0}|^2),\right. \\ &\qquad\qquad \left.\|\boldsymbol{h}\|^2 < 1\right) \\ &\quad\ge \Pr\left(||h_0|^2 - |\bar{h_0}|^2|^2 < \mathsf{SNR}^{2(r-1)}(|h_0|^2 + |\bar{h_0}|^2),\right. \\ &\qquad\qquad \left.|h_0|^2 > 1, \|\boldsymbol{h}\|^2 < \right) \\ &\quad\ge \Pr\left(||h_0|^2 - |\bar{h_0}|^2|^2 < \mathsf{SNR}^{2(r-1)},\right. \\ &\qquad\qquad \left.|h_0|^2 > 1, \|\boldsymbol{h}\|^2 < 1\right). \end{aligned}$$

But $\|\boldsymbol{h}\|^2 < 1 \Rightarrow |\bar{h}_0|^2 < L-1$ and $|h_0|^2 \sim \exp(1)$ so its pdf is a monotone decreasing function and

$$\begin{aligned} &\Pr\left(||h_0| - |\bar{h_0}||^2 < \mathsf{SNR}^{2(r-1)}, \|\boldsymbol{h}\|^2 < 1\right) \\ &\quad\ge \Pr\left(||h_0|^2 - (L-1)^2|^2 < \mathsf{SNR}^{2(r-1)}, |h_0|^2 > 1,\right. \\ &\qquad\qquad \left.\|\boldsymbol{h}\|^2 < 1\right) \\ &\quad= \Pr\left(||h_0|^2 - (L-1)^2|^2 < \mathsf{SNR}^{2(r-1)}\right) \\ &\qquad \times \Pr(\|\boldsymbol{h}\|^2 < 1) \\ &\quad\doteq \Pr\left(||h_0|^2 - 1|^2 < \mathsf{SNR}^{2(r-1)}\right) \Pr(\|\boldsymbol{h}\|^2 < 1) \\ &\quad\doteq \int_{1-\mathsf{SNR}^{r-1}}^{1+\mathsf{SNR}^{r-1}} e^{-x}dx \\ &\quad\doteq \mathsf{SNR}^{r-1}. \end{aligned}$$

■

Lemma VII.9: Let $X \sim \mathcal{CN}(0, \sigma_X^2)$. For every $a \in \mathbb{C}$ and $x \ge 0$

$$\Pr(|X-a| > x) \ge \Pr(|X| > x).$$

Proof: $|X-a|$ is Ricean distributed with noncentrality parameter $|a|^2$. Thus[5, p. 46]

$$\Pr(|X-a| > x) = Q_1\left(\frac{|a|}{\sigma_X}, \frac{x}{\sigma_X}\right)$$

where Q_1 is the Marcum Q-function. By $I_k(\cdot)$ we denote the kth-order modified Bessel function of the first kind. By definition

$$Q_1(\alpha,\beta) = e^{-(\alpha^2+\beta^2)/2} \sum_{k=0}^\infty \left(\frac{\alpha}{\beta}\right)^k I_k(\alpha\beta)$$

$$
\begin{aligned}
&= e^{-(\alpha^2+\beta^2)/2} \sum_{k=0}^{\infty} \left(\frac{\alpha}{\beta}\right)^k \sum_{n=0}^{\infty} \frac{\left(\frac{\alpha\beta}{2}\right)^{k+2n}}{n!(k+n)!} \\
&= e^{-(\alpha^2+\beta^2)/2} \sum_{n=0}^{\infty} \frac{\beta^{2n}\alpha^{2n}}{n!2^{2n}} \sum_{k=0}^{\infty} \frac{\left(\frac{\alpha^2}{2}\right)^k}{(k+n)!} \\
&\geq e^{-(\alpha^2+\beta^2)/2} \sum_{k=0}^{\infty} \frac{\left(\frac{\alpha^2}{2}\right)^k}{k!} \\
&= e^{-(\alpha^2+\beta^2)/2} e^{\alpha^2/2} \\
&= e^{-\beta^2/2} \\
&= Q_1(0, \beta).
\end{aligned}
$$

The last step follows from the fact $I_0(0) = 0$. The inequality step results from just considering the elements of the double summation for which $n = 0$ and noting that all terms in the double summation are positive. Thus

$$\Pr(|X - a| > x) \geq Q_1\left(0, \frac{x}{\sigma_X}\right) = \Pr(|X| > x). \qquad \blacksquare$$

REFERENCES

[1] J. H. Winters, J. Salz, and R. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1740–1751, Feb./Mar./Apr. 1994.

[2] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[3] E. A. Lee and D. G. Messerschmitt, *Digital Communication*, 2nd ed. Boston, MA: Kluwer Academic, 1994.

[4] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 3, pp. 363–378, May 1972.

[5] J. G. Proakis, *Digital Communications*, 4th ed. Boston, MA: McGraw-Hill, 2001.

Leonard H. Grokop (S'02–M'08) received the B. Eng. and B. Sc. degrees from the University of Melbourne, Melbourne, Australia in 2001, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 2005 and 2008, respectively.

He is currently a Senior Engineer in Corporate Research and Development at Qualcomm Inc. His research interests are in wireless communications and information theory.

David N. C. Tse (M'96–SM'07) received the B.A.Sc. degree in systems design engineering from University of Waterloo, Waterloo, ON, Canada in 1989, and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1991 and 1994, respectively.

From 1994 to 1995, he was a Postdoctoral Member of Technical Staff at AT&T Bell Laboratories. Since 1995, he has been at the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley, where he is currently a Professor. His research interests are in information theory, wireless communications, and networking.

Prof. Tse received a 1967 NSERC 4-year graduate fellowship from the Government of Canada in 1989, a NSF CAREER award in 1998, the Best Paper Awards at the Infocom 1998 and Infocom 2001 conferences, the Erlang Prize in 2000 from the INFORMS Applied Probability Society, the IEEE Communications and Information Theory Society Joint Paper Award in 2001, and the Information Theory Society Paper Award in 2003. He was the Technical Program Co-Chair of the International Symposium on Information Theory in 2004, and was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2001 to 2003. He is a coauthor, with Pramod Viswanath, of the text *Fundamentals of Wireless Communication* (Cambridge, U.K.: Cambridge University Press, 2005).