Precoding and the MMSE-DFE

Richard D. Wesel and John M.Cioffi
wesel@isl.stanford.edu and *cioffi@isl.stanford.edu*
Information Systems Laboratory, Stanford University
Stanford, CA 94305

Abstract

This paper provides upper and lower bounds on the information rates achievable with a minimum mean square error (MMSE) Tomlinson-Harishima precoder (THP) assuming ideal interleaving. We also give an exact formula for zero-forcing (ZF) THP achievable information rates. At high SNR, the performance of the MMSE-THP and the ZF-THP identically suffer only the 2.55 bit or 1.53 dB "shaping" loss from capacity. At low SNR, shaping loss is negligible, but the THP systems again do not achieve capacity, largely due to the receiver modulo operation (and interleaving in the MMSE case). In general, the MMSE-THP outperforms the ZF-THP at low SNR. We show these effects for an example ISI channel.

1 Introduction

For high SNR, the ideal zero-forcing decision feedback equalizer (ZF-DFE)[1] transforms an AWGN channel with ISI to an AWGN channel with no ISI that has the same channel capacity [5]. The ideal minimum mean square error (MMSE) DFE with interleaving and an optimized transmit spectrum performs this capacity preserving transformation at any SNR [6, 7]. However, this "perfect" performance cannot be obtained by real MMSE-DFE's because real MMSE-DFE's suffer error propagation. In fact, without interleaving the ideal MMSE-DFE achieves rates which are above the true channel capacity [3, 4].

Error propagation can be eliminated with the precoding method developed by Tomlinson [8] and independently by Harishima [9] which essentially moves the feedback section to the transmitter. The Tomlinson-Harishima precoder (THP) can be applied both to the ZF-DFE producing the ZF-THP and to the MMSE- DFE producing the MMSE-THP.

[1] Other ideal ZF-DFE bounds are explored in [1, 2, 3, 4].

While eliminating error propagation, the THP introduces constraints which preclude transmission at rates approaching channel capacity. In this paper we characterize the achievable information rates for the MMSE-THP and the ZF-THP and use these characterizations to compare the MMSE-THP and ZF-THP to each other and to the channel capacity for an example AWGN channel with ISI. We examine the factors which prevent THP from achieving capacity and draw general conclusions about THP performance. We then briefly discuss recent advances in precoding.

We consider only real signals in this paper. We will refer to a sequence $\{x_k\}$ in terms of its formal D-transform $X(D) = \sum_k x_k D^{-k}$. We assume an input signal $X(D)$ is filtered by the channel $H(D)$. Real white gaussian noise $N(D)$ with power spectrum $N_0/2$ is then added to produce $Y(D) = X(D)H(D)+N(D)$.

2 Tomlinson-Harishima Precoding

Figure 1 shows a THP transmitter. $B(D)$ must be a causal and monic filter. The $\Gamma_M[\cdot]$ operator maps $\mathcal{R}$ onto the interval (-M/2, M/2] according to $\Gamma_M[x] = x + mM$ where m is the integer for which $\Gamma_M[x] \in (-M/2, M/2]$. Typically, for IID $W(D)$, $X(D)$ will be approximately IID and approximately uniformly distributed on $(-M/2, M/2]$. If W(D) is IID with a uniform distribution on $(-M/2, M/2]$, then $X(D)$ will be exactly IID and uniform on $(-M/2, M/2]$. As studied by Mazo et al [10], when $W(D)$ is a pulse amplitude modulation (PAM) signal, $X(D)$ will have a slightly larger power than $W(D)$.

Figure 2 shows our channel model followed by a THP receiver. Figure 3 presents a structure equivalent to Figure 2 where $B(D)$ is the same as in Figure 1. A result of Tomlinson and of Harishima [8, 9] is that $\Gamma_M[Y_1(D)] = W(D)$ as long as the inputs are restricted to the interval $(-M/2, M/2]$.

1058-6393/95 $4.00

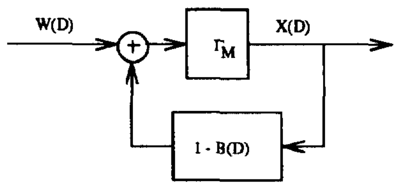

Figure 1: THP transmitter

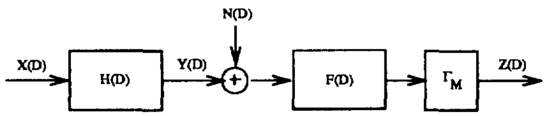

Figure 2: Channel model and THP receiver

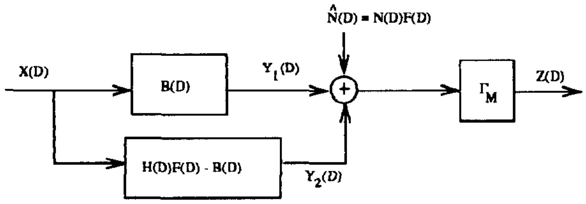

Figure 3: Equivalent to Figure 2

2.1 ZF-THP

Let the ZF-THP be the structure of Figures 1 and 2 with the filters $F(D)$ and $B(D)$ being the feedforward and feedback filters of the ZF-DFE for $H(D)$. Then as required for THP, $B(D)$ is causal and monic. Furthermore $H(D)F(D) - B(D) = 0$, i.e. $F(D) = B(D)$ and $\hat{N}(D)$ is white [11].

As noted in [12] removing the $\Gamma_M[\cdot]$ operators produces a structure with exactly the input-output characteristics of the ideal ZF-DFE. However, this structure would violate the power constraint on the channel input and perhaps be unstable.

2.2 MMSE-THP

Let the MMSE-THP be the structure of Figures 1 and 2 with the filters F(D) and B(D) being the feedforward and feedback filters of the unbiased MMSE-DFE [6, 7] for $H(D)$ and a specified SNR. Again, as required $B(D)$ is causal and monic.

However, in this case, $H(D)F(D) - B(D)$ is an anticausal filter. As a result, if $\{x_i : i \geq k\}$ depends on $\{x_i : i < k\}$ then $y_{1,k}$ will depend on $y_{2,k}$. Recall that if W(D) is IID uniform, X(D) will be IID uniform making $y_{1,k}$ and $y_{2,k}$ will be independent.

The noise $\hat{n}$ will not be white for the MMSE-THP. Furthermore, the power increase from $W(D)$ to $X(D)$ for PAM inputs discussed earlier will cause the ISI noise power $E[y_{2,k}^2]$ to increase from the ISI noise power observed for the ideal MMSE-DFE with PAM inputs.

Unlike the ZF-THP, removing the $\Gamma_M[\cdot]$ operators does *not* produce a structure equivalent to the ideal MMSE-DFE. For the $\Gamma_M[\cdot]$-removed system to be equivalent to the MMSE-DFE, the precoder filter $B(D)$ must be replaced by the *two-sided* filter

$$\frac{H(D)F(D)}{H(D)F(D) - B(D) + 1}.$$

The THP feedback filter must be causal. Thus it's not possible for a THP to be an ideal MMSE-DFE when the $\Gamma_M[\cdot]$ operators are removed.

3 Information Rates for the THP

3.1 Interleaved MMSE-THP bounds

Precoding and equalization in general seek to transform an AWGN channel with ISI to an AWGN channel without ISI. However, the channel produced by the MMSE-THP still has ISI in the signal $Y_2(D)$. Furthermore $\hat{N}(D)$ is not white. As in [7] we will use ideal interleaving to obtain an AWGN channel with no ISI.

Ideal interleaving indicates interleaving deep enough that samples of $\hat{N}(D)$ associated with the same codeword are independent and deep enough that samples of $Y(D) = Y_1(D) + Y_2(D)$ associated with the same codeword are independent. Interleaving in this context is information-lossy since we are neglecting information that future and past received symbols contained both about the value of w_k and the value of $\hat{n}_k$ corresponding to the current symbol. Note that interleaving has no effect on the dependence of $y_{1,k}$ and $y_{2,k}$.

With interleaving, the inputs and outputs associated with any particular codeword behave as those of a Discrete Memoryless Channel (DMC) with peak limited input w and output

$$\begin{aligned} z &= \Gamma_M[y_1 + y_2 + \hat{n}] \\ &= \Gamma_M[w + y_2 + \hat{n}] \end{aligned}$$

The mutual information of this DMC is given below.

$$\begin{aligned} I(w; z) &= h(z) - h(z|w) \qquad (1) \\ &= h(\Gamma_M[w + y_2 + \hat{n}]) - h(\Gamma_M[y_2 + \hat{n}]|w). \end{aligned}$$

A lower bound on $I_{MMSE-THP}$, the maximum mutual information for the MMSE-THP, can be found by assuming w is IID uniform on $(-M/2, M/2]$. First, note that for such a w we have

$$I(w;z) = \log_2(M) - h(\Gamma_M[y_2 + \hat{n}]) \quad (2)$$

Since $\Gamma_M[y_2 + \hat{n}] = \Gamma_M[y_2 + \Gamma_M[\hat{n}]]$, it's variance is upper bounded by σ^2, the variance of $y_2 + \Gamma_M[\hat{n}]$. Furthermore, its region of support is $(-M/2, M/2]$. Thus an upper bound on the second term in (2) is the maximum differential entropy for a distribution with region of support $(-M/2, M/2]$ and variance σ^2.

This maximum differential entropy is achieved by a truncated Gaussian distribution [13]. Let $\mathcal{T}(0, \sigma^2, M)$ be a zero mean Gaussian truncated to $(-M/2, M/2]$ with variance (after truncation) of σ^2.

$$h(\Gamma_M[y_2 + \hat{n}]) \leq h(\mathcal{T}(0, \sigma^2, M)). \quad (3)$$

Combining (2) and (3) produces the desired lower bound on $I_{MMSE-THP}$.

$$I_{MMSE-THP} \geq \log_2(M) - h(\mathcal{T}(0, \sigma^2, M)) \quad (4)$$

To use (4), the variance γ^2 of the Gaussian before truncation may be found by solving the equation

$$\sigma^2 = \gamma^2 - \frac{\gamma M \exp(-M^2/8\gamma^2)}{\sqrt{2\pi}\,\mathrm{erf}(M/2\gamma)}, \quad (5)$$

where

$$\mathrm{erf}(t) = \int_0^t \exp(-u^2/2)du.$$

The differential entropy (in bits) of $\mathcal{T}(0, \sigma^2, M)$ is

$$\begin{aligned} h(\mathcal{T}(0,\sigma^2,M)) &= \frac{1}{2}\log_2(2\pi\gamma^2 e) \quad + \quad (6) \\ \log_2(2\mathrm{erf}(M/2\gamma)) &- \frac{M\log_2(e)\exp(-M^2/8\gamma^2)}{4\gamma\sqrt{2\pi}\,\mathrm{erf}(M/2\gamma)}. \end{aligned}$$

An upper bound on $I_{MMSE-THP}$ can be found from the following two inequalities:

$$\begin{aligned} h(\Gamma_M[w + y_2 + \hat{n}]) &\leq \log_2(M) \\ h(\Gamma_M[y_2 + \hat{n}]|w) &\geq h(\Gamma_M[\hat{n}]). \end{aligned}$$

Thus

$$I_{MMSE-THP} \leq \log_2(M) - h(\Gamma_M[\hat{n}]). \quad (7)$$

We would like to emphasize that while we employed a uniform distribution in the derivation of the lower bound, the bounds of (4) and (7) apply to maximization over all input distributions.

3.2 ZF-THP Information Rates

For the ZF-THP the Gaussian noise $\hat{n}$ is white and $y_2 = 0$. Even without the interleaving required for the MMSE-THP, the overall structure is a DMC with input w and output $z = \Gamma_M[w + \hat{n}]$. Thus,

$$\begin{aligned} I_{ZF-THP} &= h(z) - h(z|w) \\ &= h(\Gamma_M[w + \hat{n}]) - h(\Gamma_M[\hat{n}]). \end{aligned}$$

I_{ZF-THP} is maximized by choosing w to be IID uniform over the interval $(-M/2, M/2]$ giving

$$I_{ZF-THP} = \log_2(M) - h(\Gamma_M[\hat{n}]). \quad (8)$$

The lower bound for $I_{MMSE-THP}$ in (4) looks identical to (8). However, they are different because $\hat{n}_k$ depends on $F(D)$ which will be different for the ZF-THP than for the MMSE-THP. (See [11] for details.)

We did not explicitly include a power constraint in these bounds. Rather, we assumed that the modulo boundaries would be chosen to be as large as possible while still meeting the power constraint. Note, however, that including the obvious power constraint of $M^2/12$ will change none of our results.

4 Performance Comparisons

Figure 4 shows the information rate curves resulting from the previous section as well as capacity and the IID Gaussian input information rate for an example AWGN channel with ISI. The impulse response of this channel is the first fifty taps of

$$H(D) = \frac{1 - D^2}{1 - 1.5D + 0.54D^2}.$$

Capacity and the IID Gaussian input information rate were computed assuming a power constraint of $M^2/12$. Capacity was approximated by discrete water pouring using an 8096-point FFT. The THP information rates required $h(\Gamma_M[\hat{n}])$ and the variance of $\Gamma_M[\hat{n}]$, which were computed numerically with Mathematica.

At low SNR we see in Figure 4 that the MMSE-THP outperforms the ZF-THP. This result has not been proven to hold for any AWGN channel with ISI. However, if y_2 were Gaussian, the result would follow from (2) and (8) since the variance of $y_2 + \hat{n}$ for the MMSE-THP is always smaller than the variance of $\hat{n}$ for the ZF-THP [11]. Note that $h(\Gamma_M[\mathcal{N}(0, \sigma^2)])$ increases monotonically with the σ^2. y_2 can often be closely approximated by a Gaussian.

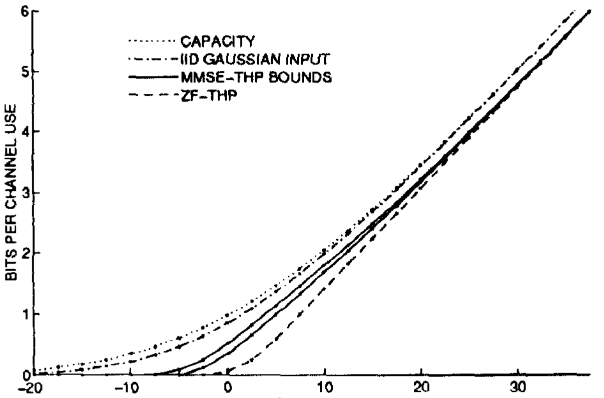

Figure 4: Information rates for example channel

Furthermore, for SNR $>$ 5 dB, the difference between $h(\Gamma_M[\mathcal{N}(0,\sigma^2)])$ and $h(\Gamma_M[\mathcal{T}(0,\sigma^2,M)])$ is neglible. Thus, in this region, at least equivalent performance of the MMSE-THP over the ZF-THP is gauranteed by (4), (8), and the smaller MMSE-THP variance discussed above.

At high SNR, ZF- and MMSE-THP perform identically. This is not surprising since for large SNR the MMSE-DFE and ZF-DFE filters which define the MMSE-THP and ZF-THP become identical.

For all SNR, the MMSE-THP bounds in Figure 4 are tight enough to show qualitatively its performance relative to ZF-THP and capacity. In general, the tightness of the MMSE-THP upper bound depends how large y_2 is relative to $\hat{n}$. This depends on the impulse response and the SNR. For more severe ISI, the upper bound will tend to be less tight. For the $1-D$ channel, the bounds were tighter than for the channel shown in Figure 4.

For high SNR, the bounds become identical regardless of the ISI. As SNR increases and the MMSE-THP approaches the ZF-THP, $Y_2 \to 0$. Thus in the limit the bounds differ by

$$h(\mathcal{T}(0,E[\hat{n}^2],M)) - h(\Gamma_M[\mathcal{N}(0,E[\hat{n}^2])])$$

which converges to zero as $E[\hat{n}^2]$ decreases to zero since both terms converge to $\frac{1}{2}\log(2\pi e E[\hat{n}^2])$. Thus, in the limit of high SNR, the MMSE-THP has

$$I = \log_2(M) - \frac{1}{2}\log_2(2\pi e E[\hat{n}^2]) \qquad (9)$$

5 THP loss from capacity

Neither THP method approaches capacity at any SNR. In this section we examine the constraints imposed by THP which lead to this loss.

5.1 IID Input Constraint

The THP transmitter forces the channel input sequence to be approximately IID as discussed earlier. In fact, we have shown that an IID uniform $W(D)$ and thus an IID $X(D)$ achieves the maximum information rate for the ZF-THP structure. Thus, we do not expect THP exceed the maximum rate achievable using IID inputs. This maximum IID input rate is achieved by an IID Gaussian input. (See [1] and references therein.) The IID Gaussian information rate curve is plotted in Figure 4 for comparison with the other curves.

A properly chosen transmit filter after the precoder can bring the IID Gaussian information rate up to the capacity of $H(D)$ [7]. Such a pulse response would require IID inputs to maintain the original power constraint.

5.2 Shaping Loss

The transmitter modulo forces the channel inputs to be peak limited. The loss due to this peak limitation is often referred to as the "shaping" loss and is asymptotically .255 bits or 1.53 dB [14].

From [7], the maximum information rate for an IID input sequence $X(D)$ with power constraint $M^2/12$ is

$$I_{IID} = \frac{1}{2}\log_2\left(1 + \frac{M^2/12}{E[\hat{n}^2]+E[y_2^2]}\right) \qquad (10)$$

where $\hat{n}$ and y_2 are those that occur in the MMSE-THP. Subtracting (9) from (10) yields

$$\frac{1}{2}\log_2\frac{\pi e}{6} = 0.255 \text{ bits}$$

plus terms which go to zero for high SNR.

Thus we have shown that at high SNR the difference between the THP information rate and the maximum information rate assuming only IID $X(D)$ and a power constraint is the shaping loss of 0.255 bits. Note that for high SNR, IID sequences can approach capacity. Thus at high SNR all the loss incurred by both THP schemes can be attributed to the peak limitation. Figure 4 demonstrates this asymptotic behavior.

5.3 Receiver Γ_M and Interleaving

At low SNR, the loss due to shaping becomes negligible. In the region around zero dB, we see in Figure 4 both THP schemes suffering increased loss from the IID Gaussian information rate even as the shaping loss becomes negligible. The only explanations for this increased penalty are the receiver $\Gamma_M[\cdot]$ and interleaving operations.

6 Conclusion

We provided bounds which allow the achievable information rates for the MMSE-THP and the ZF-THP to be compared with each other and with capacity. At high SNR MMSE-THP and ZF-THP converge to the same algorithm and suffer only the loss from capacity required by the peak limitation. In this sense, THP is optimal given the peak limitation at high SNR.

This result prompts one to ask whether there are schemes like THP which approach capacity even more closely by mitigating the peak limitation. Recently work has been done in this area [12, 15, 16]. The "ISI coding" method described in [16] can approach capacity at high SNR to the extent that trellis codes can approach capacity on AWGN channels.

At low SNR the loss from the IID Gaussian information rate is mostly due to the receiver $\Gamma_M[\cdot]$ (and interleaving for the MMSE-THP). Schemes similar to THP typically employ some type of many–to–one mapping in the receiver. It is difficult to imagine a many–to–one mapping which would not ultimately cause information loss as the noise variance becomes large. The methods described in [16] assume zero forcing and thus would not approach capacity for low SNR in any case because the variance of $\hat{n}$ will be too large.

It is in the low SNR region that questions remain about THP and equalization in general. For THP, one might try to improve our upper bound by incorporating some more information about y_2. For equalization in general, a better understanding is needed about the underlying difficulty of attempting to transform an AWGN channel with ISI to an AWGN channel without ISI while avoiding loss from capacity.

Acknowledgements

We would like to thank Shlomo Shamai for helpful suggestions including improving the MMSE-THP lower bound with the maximum entropy property of the truncated Gaussian. We also enjoyed interesting discussions with Amos Lapidoth and Erik Ordentlich. This work was partially supported by the AT&T Foundation.

References

[1] Shlomo Shamai (Shitz), Lawrence H. Ozarow, and Aaron D. Wyner. Information rates for a discrete-time gaussian channel with intersymbol interference and stationary inputs. *IEEE Trans. on Info. Theory*, 37(6), November 1991.

[2] John R, Barry, Edward A. Lee, and David G. Messerschmitt. Capacity penalty due to ideal tail-cancelling equalization. Submitted to *IEEE Trans. on Info. Theory*.

[3] Shlomo Shamai (Shitz) and Rajiv Laroia. The intersymbol interference channel: Lower bounds on capacity and precoding loss. In *28th Annual Conference on Information Sciences and Systems, Princeton, New Jersey*, March 1994.

[4] Shlomo Shamai (Shitz) and Rajiv Laroia. The intersymbol interference channel: Lower bounds on capacity and precoding loss. Submitted to IEEE Trans. on Comm.

[5] Robert Price. Nonlinearly feedback equalized PAM versus capacity for noisy channel filters. In *Proceedings ICC '72*, pages 22–12 to 22–17, June 1972.

[6] John M. Cioffi, Glenn P. Dudevoir, M. Vedat Eyuboglu, and G. David Forney, Jr. MMSE decision-feedback equalizers and coding – part I: Equalization results. To appear in *IEEE Trans. on Comm.*

[7] John M. Cioffi, Glenn P. Dudevoir, M. Vedat Eyuboglu, and G. David Forney, Jr. MMSE decision-feedback equalizers and coding – part II: Coding results. To appear in *IEEE Trans. on Comm.*

[8] M. Tomlinson. New automatic equalizer employing modulo arithmetic. *Electronic Letters*, 7:138–139, March 1971.

[9] Hiroshi Harishima and Hiroshi Harishima Miyakawa. Matched-transmission technique for channels with intersymbol interference. *IEEE Trans. on Comm.*, 20(4):774–780, August 1972.

[10] James E. Mazo and Jack Salz. On the transmitted power in generalized partial response. *IEEE Trans. on Comm.*, 24(3):348–352, March 1976.

[11] John M. Cioffi. Signals and detection. Stanford University Lecture Notes, 1994.

[12] M. Vedat Eyuboğlu and G. David Forney, Jr. Trellis precoding: Combining coding, precoding and shaping for intersymbol interference channels. *IEEE Trans. on Info. Theory*, 38(2):301–314, March 1992.

[13] Shlomo Shamai and Amir Dembo. Bounds on the symmetric binary cutoff rate for dispersive gaussian channels. *IEEE Trans. on Comm.*, 42(1):39–53, January 1994.

[14] G. David Forney, Jr. Trellis shaping. *IEEE Trans. on Info. Theory*, 38(2):281–300, March 1992.

[15] Rajiv Laroia. A simple and effective precoding scheme for noise whitening on intersymbol interference channels. *IEEE Trans. on Comm.*, 41(10):1460–1463, October 1993.

[16] Rajiv Laroia. Coding for intersymbol interference channels – combined coding and precoding. Submitted to IEEE Trans. on Info. Theory.