ECE 361: Digital Communication

Lecture 13: Intersymbol Interference Management: High SNR Regime

Introduction

We have seen that the key aspect of the wireline channel is the intersymbol interference (ISI), and our focus is to develop receiver techniques that will mitigate and potentially harness the effects of interference.

In Lecture 10, we focused on the low SNR regime. Here the noise power dominates the total signal power and hence the interference power. Essentially, we saw that in this regime, interference is not the issue and treating interference as noise is indeed a good strategy. The matched filter, which focusses on collecting the signal energy and ignores the interference, performs well in this regime.

The focus of this lecture is on high SNR regime. Now the total signal power, and hence the interference power, dominates the noise power. Hence, the main aspect of the receiver design is on how to handle interference. It makes no sense in this regime to naively ignore interference and treat it as noise. Matched filter will not perform well in this regime and we need to develop new techniques to deal with interference.

In this lecture, we will again adopt the receiver centric approach and will focus on simple linear processing schemes that will handle interference and convert the ISI channel into an effective AWGN channel. We can then use our analysis and codes developed for AWGN channel. In particular, we will study two schemes, viz., zero forcing equalizer (ZFE) and successive interference cancellation (SIC), that perform well in high SNR regime. Both of these schemes pay attention to the fact that interference has a structure and given the knowledge of present symbol, interference can be predicted. We will begin with a simple calculation to understand why matched filter fails in High SNR regime.

Performance of Matched Filter at High SNR

The received voltage of the wireline channel is

$$y[m] = \sum_{\ell=0}^{L-1} h_\ell x[m-\ell] + w[m], \quad m > 0. \tag{1}$$

Recall that the SINR at the output of the matched filter is

$$\mathsf{SINR}_{MF} = \frac{\left(\sum_{l=0}^{L-1} h_l^2\right) \mathsf{SNR}}{\frac{\sum_{\ell=0}^{L-1}\sum_{k=0;k\neq\ell}^{L-1} h_\ell^2 h_k^2}{\sum_{\ell=0}^{L-1} h_\ell^2}\mathsf{SNR}+1}, \tag{2}$$

where

$$\mathsf{SNR} = \frac{\mathbb{E}\left[(x[m])^2\right]}{\sigma^2}. \tag{3}$$

Note that the SINR at the output of the matched filter depends on the operating SNR. Two different regimes are of keen interest:

- *Low* SNR: For small values of SNR, the SINR is close to linear in SNR. In other words, a doubling of the operating SNR also doubles the $\mathsf{SINR}_{\rm MF}$. Mathematically, this regime happens when the interference

$$\frac{\sum_{l=0}^{L-1}\sum_{k=0;k\neq l}^{L-1} h_l^2 h_k^2}{\sum_{l=0}^{L-1} h_l^2}\mathsf{SNR} \ll 1. \tag{4}$$

  In this regime, the interference is much smaller than the background noise and it makes fine sense to just ignore the interference. The channel is almost like an AWGN one and hence the linear relationship between SNR and $\mathsf{SINR}_{\rm MF}$.

- *High* SNR: For large values of SNR, the $\mathsf{SINR}_{\rm MF}$ is almost constant and hardly changes. Mathematically, this regime kicks in when the interference

$$\frac{\sum_{l=0}^{L-1}\sum_{k=0;k\neq l}^{L-1} h_l^2 h_k^2}{\sum_{l=0}^{L-1} h_l^2}\mathsf{SNR} \gg 1 \tag{5}$$

  and

$$\mathsf{SINR}_{\rm MF} \approx \frac{\left(\sum_{l=0}^{L-1} h_l^2\right)^2}{\sum_{l=0}^{L-1}\sum_{k=0;k\neq l}^{L-1} h_l^2 h_k^2}. \tag{6}$$

  In this regime, the interference is much larger than noise and we pay a steep price by just ignoring its presence. The interference level is directly proportional to the transmit signal energy and hence the $\mathsf{SINR}_{\rm MF}$ saturates to a fixed value and is insensitive to increase in SNR in this regime.

To ameliorate the deleterious presence of interference in the high SNR regime, it helps to try to eliminate the source of interference as much as possible. The zero forcing equalizer, the next topic, precisely does this: it entirely removes the interference (or forces the interference to zero, hence its name).

The Zero Forcing Equalizer

Let's consider communication over n channel uses on the L-tap wireline channel. The received voltage sequence of length $n+L-1$ is given by Equation 1. We can collect together all the n transmissions and receptions and represent them compactly in vector form. Let $\mathbf{x}$ and $\mathbf{y}$ be the vector of transmitted and received voltages respectively. Let $\mathbf{w}$ be the noise vector of i.i.d. Gaussian noises. The convolution of the channel vector $\mathbf{h}$ with the transmitted vector $\mathbf{x}$ can be written as a product of channel matrix $\mathbf{H}$ and vector $\mathbf{x}$. Rows of the channel matrix $\mathbf{H}$ are obtained from shifting $\mathbf{h}$ appropriately. To illustrate this, let's consider a concrete example of transmission of 4 symbols over a 3 tap channel, i.e., $n=4$ and $L=3$. We can write the Equation (1) in matrix notation as

$$\begin{bmatrix} y[1] \\ y[2] \\ y[3] \\ y[4] \\ y[5] \\ y[6] \end{bmatrix} = \begin{bmatrix} h_0 & 0 & 0 & 0 \\ h_1 & h_0 & 0 & 0 \\ h_2 & h_1 & h_0 & 0 \\ 0 & h_2 & h_1 & h_0 \\ 0 & 0 & h_2 & h_1 \\ 0 & 0 & 0 & h_2 \end{bmatrix} \begin{bmatrix} x[1] \\ x[2] \\ x[3] \\ x[4] \end{bmatrix} + \begin{bmatrix} w[1] \\ w[2] \\ w[3] \\ w[4] \\ w[5] \\ w[6] \end{bmatrix}. \tag{7}$$

Thus, we can write Equation 7 in short hand notation as,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{8}$$

where $\mathbf{x}$ is a n dimensional vector, $\mathbf{y}$ and $\mathbf{w}$ are $n+L-1$ dimensional vectors. The channel matrix $\mathbf{H}$ is an $n+L-1 \times n$ dimensional matrix.

If there were no noise, then we have $n+L-1$ equations in n unknowns. This is an over specified system of equations (more equations than unknowns). Then we can choose any n independent equations to solve for $\mathbf{x}$. We can choose these n equations in many ways and all the choices are equivalent.

With noise $\mathbf{w}$ present in the channel, however, the *choice* of the independent equations in solving for $\mathbf{x}$ matters. Each of the choices will still be able to remove the interference completely, but leave behind noises of *different* variances (the noise left behind is just the linear combinations of $w[m]$s). The SNR of the resulting AGN channel now depends on our choice of the equations we use to obtain the interference-free observations.

The zero forcing equalizer (ZFE) with respect to one of the transmit voltages, say $x[1]$, is that linear combination of the observed voltages which satisfies two properties:

- it leads to an interference-free observation (but still corrupted by additive Gaussian noise) of $x[1]$;
- it has the largest output SNR in the resulting AWGN channel with respect to $x[1]$.

Now we see why the ZFE is much better than the matched filter at high SNRs: ZFE cancels the interference completely (at the cost of potentially changing the noise variance; but this change is independent of the operating transmit SNR). Thus, as we increase the signal power, the effective received SNR scales *linearly* with the signal power. This is in stark contrast to the performance of the matched filter which had a ceiling for received SNR no matter how much the transmit power is increased.

Example

To get a concrete feel, let us consider the ISI channel model by considering just a single interfering symbol, i.e., $L = 2$:

$$y[m] = h_0 x[m] + h_1 x[m-1] + w[m], \quad m \geq 1. \tag{9}$$

Let's also restrict the transmission to 3 symbols, i.e., $n = 3$. So we are considering

$$\begin{aligned} y[1] &= h_0 x[1] + w[1] && (10) \\ y[2] &= h_1 x[1] + h_0 x[2] + w[2] && (11) \\ y[3] &= h_1 x[2] + h_0 x[3] + w[3] && (12) \\ y[4] &= h_1 x[3] + w[4]. && (13) \end{aligned}$$

Again, let us consider sequential communication starting at $m = 1$. Since $x[0] = 0$, there is *no* ISI at the first time sample:

$$y[1] = h_0 x[1] + w[1]. \tag{14}$$

The first transmit voltage $x[1]$ also features in the voltage received at the second time sample:

$$y[2] = h_1 x[1] + h_0 x[2] + w[2], \tag{15}$$

but this time it is corrupted by interference $(h_0 x[2])$. At high operating SNRs, the interference dominates the additive Gaussian noise and $y[1]$ is far more valuable in decoding the information contained in $x[1]$ than $y[2]$. We could use $y[1]$ alone to decode the information bits at the first time sample: this is just a simple AWGN channel with SNR equal to

$$h_0^2 \mathsf{SNR}. \tag{16}$$

The important conclusion is that it increases linearly with the operating SNR level regardless of whether the SNR value is large or small. Indeed, we can expect this simple receiver to outperform the matched filter at large SNR operating levels.

Well, getting to $x[1]$ without any interference was easy, but how about $x[2]$? Now we could use the two linear equations

$$\begin{aligned} y[1] &= h_0 x[1] + w[1] && (17) \\ y[2] &= h_1 x[1] + h_0 x[2] + w[2], && (18) \end{aligned}$$

to "eliminate" the interfering term $x[1]$: consider the linear combination

$$h_0 y[2] - h_1 y[1] = h_0^2 x[2] + \left(h_0 w[2] - h_1 w[1]\right). \tag{19}$$

This leads to a plain AWGN channel with respect to the transmit voltage $x[2]$. Since we do not have any interference from the previous symbol $x[1]$, the SNR of this AWGN channel is simply

$$\frac{h_0^4}{h_0^2 + h_1^2} \mathsf{SNR} \tag{20}$$

where SNR is, as usual, the ratio of the transmit energy to that of the additive Gaussian noise. Again, observe the linear relation to SNR regardless of whether it is high or low.

Now, the pattern is clear: to decode the information in $x[3]$ we consider the linear combination

$$h_0^2 y[3] - h_1 \left(h_0 y[2] - h_1 y[1]\right). \tag{21}$$

Substituting from Equation (19) we see that

$$h_0^2 y[3] - h_1 \left(h_0 y[2] - h_1 y[1] \right) = h_0^3 x[3] + \left(h_0^2 w[3] + h_1^2 w[1] - h_0 h_1 w[2] \right), \tag{22}$$

an AWGN channel with respect to the transmit voltage $x[3]$. The SNR at the output of this AWGN channel is

$$\frac{h_0^6}{h_0^4 + h_1^4 + h_0^2 h_1^2} \mathsf{SNR}, \tag{23}$$

again depending linearly with respect to the operating SNR.

As we have already noted, it is not the only way to do this. Specifically, we could take *different* linear combinations of the received voltages that also lead to the desired zero interference characteristic. We could then pick that linear combination which yields the largest output SNR. To see this concretely, suppose we wanted to use the ISI channel only for three time instants $m = 1, 2, 3$. This means that we observe $x[3]$ without interference from $y[4]$:

$$y[4] = h_1 x[3] + w[4], \tag{24}$$

an AWGN channel with output SNR equal to

$$h_1^2 \mathsf{SNR}. \tag{25}$$

Depending on whether this SNR value is larger or smaller than the one in Equation (23), we could use either just $y[4]$ alone or $h_0^2 y[3] - h_1 \left(h_0 y[2] - h_1 y[1] \right)$ to decode the information in $x[3]$; both of them are interference-free observations of $x[3]$. To be concrete, this comparison comes down to the following: if

$$\left(h_1^2 + h_0^2 \right) \left(h_1^4 + h_0^4 \right) \geq 2 h_0^6, \tag{26}$$

then we would prefer to use $y[4]$ to decode the information in $x[3]$. Else, we prefer $h_0^2 y[3] - h_1 h_0 y[2] + h_1^2 y[1]$.

Analogously, we could have used the linear combination

$$h_1 y[3] - h_0 y[4] = h_1^2 x[2] + \left(h_1 w[3] - h_0 w[4] \right) \tag{27}$$

to decode the information in the transmit voltage $x[2]$. The corresponding output SNR is

$$\frac{h_1^4}{h_0^2 + h_1^2} \mathsf{SNR}. \tag{28}$$

This is an alternative to the linear combination used in Equation (19) and we see that the output SNRs of the two linear combinations are not the same (cf. Equation (20)). We would use the linear combination in Equation (19) if

$$h_0^2 \geq h_1^2 \tag{29}$$

and the linear combination in Equation (27) otherwise.

Finally, we could have used the linear combination

$$h_1^2 y[2] - h_0 \left(h_0 y[4] - h_1 y[3] \right) = h_1^3 x[1] + \left(h_1^2 w[2] + h_0 h_1 w[3] - h_0^2 w[4] \right) \tag{30}$$

to decode the information in $x[1]$. The output SNR of this AWGN channel is

$$\frac{h_1^6}{h_0^4 + h_0^2 h_1^2 + h_1^4} \mathsf{SNR}. \tag{31}$$

Comparing this with the earlier choice of just using $y[1]$ to decode the information in $x[1]$, we see that we would prefer this simple choice (as compared to the linear combination in Equation (30)) provided

$$\frac{h_1^6}{h_0^4 + h_0^2 h_1^2 + h_1^4} \leq h_0^2. \tag{32}$$

Indeed, due to the similarity of the transmit voltages $x[1]$ and $x[3]$ (the first and last transmissions), this condition is also similar to that in Equation (26):

$$\left(h_1^2 + h_0^2\right)\left(h_1^4 + h_0^4\right) \geq 2h_1^6. \tag{33}$$

Successive Interference Cancelation

The ZFE is a linear processing technique to remove interference. If we allow *nonlinear* processing, there are alternative interference removal strategies. The most prominent of them is the so-called *successive interference cancelation* (SIC) strategy. It is different from the ZFE in two important ways:

1. it is a nonlinear operation on the received voltage symbols;
2. while ZFE works on removing *any* interference voltages, the SIC will only work when the interference voltages are the outputs of a good coding and modulation scheme. The SIC harnesses the fact that the transmit symbols are not any choice of voltages, but there is a clear structure and redundancy built into them.

We start out with a naive version. Once we fully understand it, we can build on top of the key idea. Observe that the very first transmitted symbol does not see any interference. If we can decode the first symbol reliably (probability of which is high in the high SNR regime), then we can subtract the interference caused by the first symbol from the successive received voltages. Thus, the second symbol, which had interference from the first symbol alone, now sees an AWGN channel. This symbol can be decoded reliably and interference that it has caused to the further symbols can be canceled. Thus, each symbol now sees only the additive noise.

Let's consider an example to see how successive interference cancelation (SIC) works. Consider L-tap wireline channel. The first few received voltages are

$$y[1] = h_0 x[1] + w[1] \tag{34}$$
$$y[2] = h_0 x[2] + h_1 x[1] + w[2] \tag{35}$$
$$y[3] = h_0 x[3] + h_1 x[2] + h_2 x[1] + w[3] \tag{36}$$

and so on.

Since, the first symbol does not see any ISI, we can simply decode it using ML receiver. Let $\hat{x}[1]$ be the decoded symbol. Note that $x[1]$ sees an AWGN channel with

$$\mathsf{SNR}_{effective} = h_0^2 \mathsf{SNR}. \tag{37}$$

Assuming that we have decoded $x[1]$ correctly, we subtract its effect from $y[2]$ before decoding $x[2]$.

$$\tilde{y}[2] = y[2] - h_1\hat{x}[1] = h_0 x[2] + h_1(x[1] - \hat{x}[1]) + w[2] \tag{38}$$

Thus, if $\hat{x}[1] = x[1]$, we have successfully removed the interference. $x[2]$ now sees an AWGN channel with

$$\mathsf{SNR}_{effective} = h_0^2 \mathsf{SNR}. \tag{39}$$

We now use $\hat{x}[1]$ and $\hat{x}[2]$ to cancel the interference from $y[3]$. Thus, each of the symbol sees an interference-free channel with

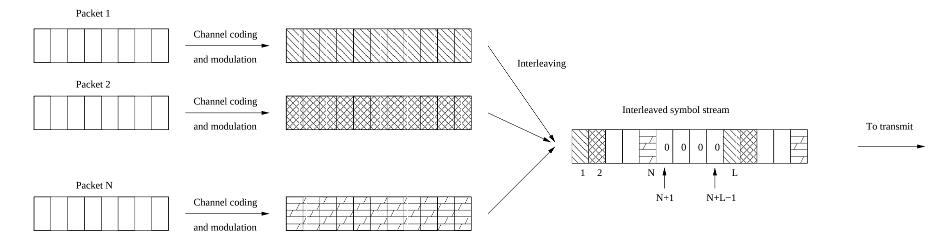
$$\mathsf{SNR}_{SIC} = h_0^2 \mathsf{SNR}. \tag{40}$$

Note that the effective SNR for SIC grows linearly with SNR.

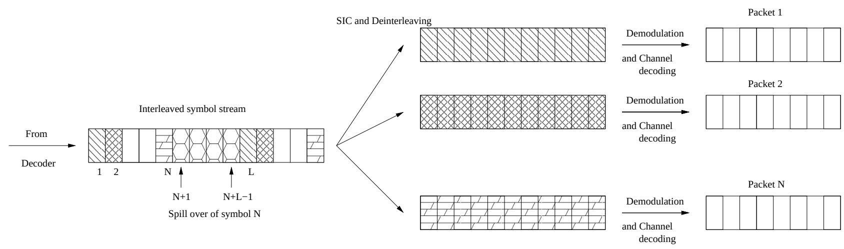
However, this simple scheme has a caveat. Here we have assumed that we always decode the symbols correctly. If $\hat{x}[1] \neq x[1]$, then we are actually *adding* interference to $y[2]$ and thus, the probability of error in decoding $x[2]$ is now quite high. This is particularly true in the high SNR regime. Now given that $x[2]$ is decoded wrongly, an analogous argument suggests that $x[3]$ will also be decoded wrongly with a very high probability. Thus, the errors propagate in this scheme. So no matter how small $P(\varepsilon)$ is, as long as $P(\varepsilon) > 0$, we cannot achieve arbitrary reliability.

However, this problem can be easily overcome by using long block codes along with interleaving. We know that for an AWGN channel, we can achieve arbitrary reliability with long block codes. We will now split the data into N ($N > L$) subpackets which are coded and modulated separately using capacity achieving codes. We then interleave one symbol from each of the N subpackets and pad these N symbols with $L-1$ zeros. Because of this zero padding, all the symbols from the first subpackets do not see any interference. We can then decode the entire first subpacket with arbitrary reliability. We then use the decoded subpacket to remove the interference from the received symbols of subsequent subpackets. Since each of the symbols of each subpacket is decoded with arbitrary reliability, the risk of error propagation is not there anymore. Note that this scheme has the overhead associated with zero padding. We pad $L-1$ zeros to each of the N symbols. Thus, the channel is used only for the $\frac{N}{N+L-1}$ fraction of time. But we can use $N \gg L-1$ and $\frac{N}{N+L-1} \approx 1$ as N grows. Figure 1 explains this scheme pictorially.

Note that SIC is different from the other equalizers we have studied so far. Note that matched filter and ZFE involve processing the received symbols *before* decoding. Decoding takes place only with the output of the equalizers. But SIC goes back and forth between decoding and processing received voltages. SIC can be coupled with the other equalizers to improve the performance. We will see the full power of the SIC receiver in the next lecture.



(a) Interleaving



(b) Deinterleaving

Figure 1: Interleaving and zero padding for SIC

Looking Ahead

We have seen that the matched filter worked well at low values of SNR, while the zero forcing equalizer worked well at high values of SNR. In the next lecture, we will focus on the metric of the SNR at the output of the equalizer to derive the *optimal* filter: this is the so-called MMSE (minimum mean squared error) equalizer. It strictly outperforms both the matched filter and the ZFE and is commonly used in practical communication systems. It turns out that MMSE combined with SIC achieves the *capacity* (the fundamental limit on the rate of reliable communication) of the wireline channel.