ECE 361: Digital Communications

Lecture 17: Orthogonal Frequency Division Modulation (OFDM) and Capacity of the Wireline Channel

Introduction

In this lecture we will see in detail the OFDM method to convert the wireline channel into a parallel AWGN channel. We also see that this achieves the *capacity* of the wireline channel – in other words, the largest possible data rate of communication is achieved by the OFDM method.

OFDM

Consider the frequency selective model that we have been working with as a good approximation of the wireline channel:

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

We will convert the ISI channel in Equation (1) into a *collection* of AWGN channels, each of different noise energy level:

$$\hat{y}[N_c k + n] = \hat{h}_n \hat{x}[N_c k + n] + \hat{w}[N_c k + n], \quad k \geq 0, \quad n = 0 \ldots N_c - 1. \tag{2}$$

We will be able to make this transition by some very simple signal processing techniques. Interestingly, these signal processing techniques are *universally* applicable to every wireline channel, i.e., they do not depend on the exact values of channel coefficients $h_0, \ldots, h_{L-1}$. This makes OFDM a very *robust* communication scheme over the frequency-selective channel.

Cyclic Prefix

Suppose we have mapped our information bits into N_c voltages. We will revisit the issue of how these voltages were created from the information bits at a slightly later point in this lecture. For now, we write them as a vector:

$$\mathbf{d} = [d[0], d[1], \ldots, d[N_c - 1]]^t.$$

We use these N_c voltages to create an $N_c + L - 1$ block of *transmit* voltages as:

$$\mathbf{x} = [d[N_c - L + 1], d[N_c - L + 2], \ldots, d[N_c - 1], d[0], d[1], \ldots, d[N_c - 1]]^t, \tag{3}$$

i.e., we add a *prefix* of length $L-1$ consisting of data symbols rotated cyclically (Figure 1). The first $L-1$ transmitted symbols contain the "data" symbols $d[N_c-(L-1)], \ldots, d[N_c-1]$. The next N_c transmitted voltages or symbols contain the "data" symbols $d[0], d[1], \ldots, d[N_c-1]$. In particular, for a 2-tap frequency-selective channel we have the following result of cyclic precoding:

$$\begin{aligned} x[1] &= d[N_c-1] \\ x[2] &= d[0] \\ x[3] &= d[1] \\ &\vdots \\ x[N_c+1] &= d[N_c-1] \end{aligned}$$

With this input to the channel (1), consider the output

$$y[m] = \sum_{\ell=0}^{L-1} h_\ell x[m-\ell] + w[m], \qquad m = 1, \ldots, N_c + 2(L-1).$$

The first $L-1$ elements of the transmitted vector $\mathbf{x}$ were constructed from circularly wrapped elements of the vector $\mathbf{d}$, which are included in the last N_c-1 elements of $\mathbf{x}$. The receiver hence ignores the first $L-1$ received symbols $y[1], \ldots, y[L-1]$. The ISI extends over the first $L-1$ symbols and the receiver ignores it by considering only the output over the time interval $m \in [L, N_c+L-1]$. Let us take a careful look at how the N receive voltages (received at times L through N_c+L-1) depend on the transmit voltages $d[0], \ldots, d[N_c-1]$:

$$y[m] = \sum_{\ell=0}^{L-1} h_\ell d[(m-L-\ell) \text{ modulo } N_c] + w[m]. \tag{4}$$

See Figure (1).

Denoting the received voltage vector of length N_c by

$$\mathbf{y} = [y[L], \ldots, y[N_c+L-1]]^t,$$

and the channel by a vector of length N_c

$$\mathbf{h} = [h_0, h_1, \ldots, h_{L-1}, 0, \ldots, 0]^t, \tag{5}$$

(4) can be written as

$$\mathbf{y} = \mathbf{h} \otimes \mathbf{d} + \mathbf{w}. \tag{6}$$

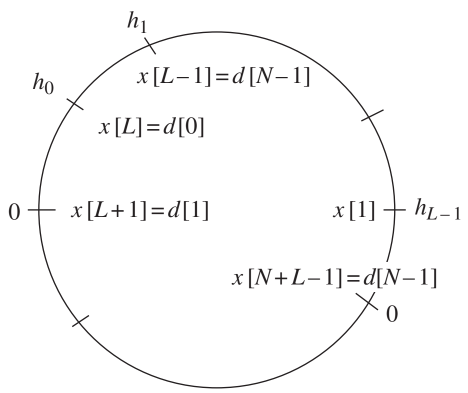

Figure 1: Convolution between the channel (**h**) and the input (**x**) formed from the data symbols (**d**) by adding a cyclic prefix. The output is obtained by multiplying the corresponding values of **x** and **h** on the circle, and outputs at different times are obtained by rotating the x-values with respect to the h-values. The current configuration yields the output $y[L]$.

Here we denoted

$$\mathbf{w} = [w[L], \ldots, w[N_{\rm c} + L - 1]]^t, \tag{7}$$

as a vector of i.i.d. $\mathcal{N} \sim (0, \sigma^2)$ random variables. The notation of $\otimes$ to denote the *cyclic convolution* in (6) is standard in signal processing literature. The point of all this manipulation will become clear when we review a key property of cyclic convolution next.

Discrete Fourier Transform

The discrete Fourier transform (DFT) of a vector (such as **d**) is also another vector of the same length (though the entries are in general, complex numbers). The different components of the discrete Fourier transform of the vector **d**, denoted by DFT(**d**), are defined as follows:

$$\tilde{d}_n := \frac{1}{\sqrt{N_{\rm c}}} \sum_{m=0}^{N_{\rm c}-1} d[m] \exp\left(\frac{-{\rm j}2\pi nm}{N_{\rm c}}\right), \qquad n = 0, \ldots, N_{\rm c} - 1. \tag{8}$$

Even though the voltages $d[\cdot]$ are real, the DFT output $\tilde{d}_n$ are complex. Nevertheless, there is *conjugate symmetry*:

$$\tilde{d}_n = \tilde{d}^*_{N_{\rm c}-1-n}, \quad n = 0, \ldots, N_{\rm c} - 1. \tag{9}$$

DFTs and circular convolution are crucially related through the following equation (perhaps the most important of all in discrete time digital signal processing):

$$\text{DFT}(\mathbf{h} \otimes \mathbf{d})_n = \sqrt{N_{\rm c}}\text{DFT}(\mathbf{h})_n \cdot \text{DFT}(\mathbf{d})_n, \qquad n = 0, \ldots, N_{\rm c} - 1. \tag{10}$$

The vector $[\tilde{h}_0, \dots, \tilde{h}_{N_c-1}]^t$ is defined as the DFT of the L-tap channel $\mathbf{h}$, multiplied by $\sqrt{N_c}$,

$$\tilde{h}_n = \sum_{\ell=0}^{L-1} h_\ell \exp\left(\frac{-j2\pi n\ell}{N_c}\right). \tag{11}$$

Thus we can rewrite (6) as

$$\tilde{y}_n = \tilde{h}_n \tilde{d}_n + \tilde{w}_n, \qquad n = 0, \dots, N_c - 1. \tag{12}$$

Here we have denoted $\tilde{w}_0, \dots, \tilde{w}_{N_c-1}$ as the N_c-point DFT of the noise vector $w[1], \dots, w[N_c]$. Observe the following.

- Even though the received voltages $y[\cdot]$ are real, the voltages at the output of the DFT $\tilde{y}_n$ are complex. Thus it might seem odd that we started out with N real numbers and ended up with $2N$ real numbers. But there is a redundancy in the DFT output $\tilde{y}[\cdot]$. Specifically,

$$\tilde{y}_n = \tilde{y}^*_{N_c-1-n}. \tag{13}$$

  In other words, the real parts of $\tilde{y}_n$ and $\tilde{y}_{N_c-1-n}$ are the same. Further, the imaginary parts of $\tilde{y}_n$ and $\tilde{y}_{N_c-1-n}$ are negative of each other.

- Even though the noise voltages $w[\cdot]$ are real, the noise voltages at the output of the DFT $\tilde{w}_n$ are complex. Just as before, there is a redundancy in the noise voltages:

$$\tilde{w}_n = \tilde{w}^*_{N_c-1-n}, \quad n = 0, \dots, N_c - 1. \tag{14}$$

  We know that the noise voltages $w[m]$ were white Gaussian. What are the statistics of the DFT outputs? It turns out that they are *also* white Gaussian: the real and imaginary parts of $\tilde{w}_0, \dots, \tilde{w}_{\frac{N_c}{2}-1}$ are all independent and identically distributed as Gaussian with zero mean and variance σ^2 (here we supposed for notational simplicity that N_c is even, so $\frac{N_c}{2}$ is an integer).

- Even though the channel coefficients h_ℓ are real (and zero for $\ell = L, \dots, N_c - 1$), the values at the output $\tilde{h}_n$ of the DFT are complex (and, in general, non-zero for all values $n = 0, \dots, N_c - 1$). Again, observe that

$$\tilde{h}_n = \tilde{h}[N_c - 1 - n]^*, \quad , n = 0, \dots, N_c - 1. \tag{15}$$

The result that we get from this precoding is the following: the DFT of the received vector $\mathbf{y}$ and the DFT of our initial "data" vector $\mathbf{d}$ have the relationship that the received and transmitted vectors have in an AWGN channel with no ISI (given a suitable definition for the noise vector and it's DFT as given earlier). This seems to suggest that if we put the

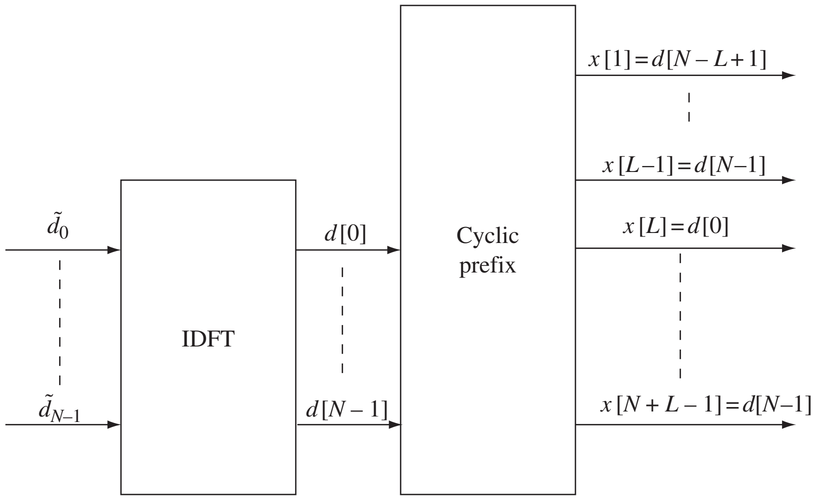

Figure 2: The cyclic prefix operation.

actual data that we want to transmit on the DFT, and take the DFT of what we receive, then we can perform something similar to traditional AWGN style decoding. Note that this scheme uses $L-1$ extra time instants. This yields the block diagram in Figure (3). A careful and detailed derivation of this step is carried out next. At the end of that calculation, we will have also shown how we arrive at the parallel AWGN channel (cf. Equation (2)).

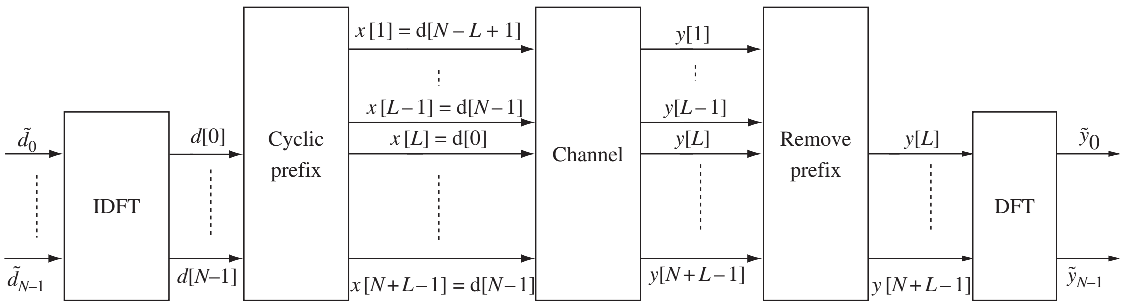

Figure 3: The OFDM transmission and reception schemes.

Packaging the Data

In this section we will see how to connect the AWGN modulation techniques with the OFDM transmission scheme. Suppose we start out with N_c real voltages $\hat{x}[0], \ldots, \hat{x}[N_c-1]$. These are the transmit voltages on the N_c sub-channels using the separate communication architecture (they are generated by efficient coding techniques – such as LDPC codes – for the AWGN channel). Let us suppose that N_c is an even number. This will simplify our notations. We generate *half* of the data vector $\tilde{\mathbf{d}}$ as follows:

$$\begin{aligned}
\Re\left[\tilde{d}_n\right] &\stackrel{\text{def}}{=} \hat{x}[2n] && (16)\\
\Im\left[\tilde{d}_n\right] &\stackrel{\text{def}}{=} \hat{x}[2n+1], \quad n = 0,\dots,\frac{N_\text{c}}{2}-1. && (17)
\end{aligned}$$

The second half is simply conjugate symmetric of the first part (so as to respect Equation (9)):

$$\tilde{d}_n = \tilde{d}^*_{N_\text{c}-1-n}, \quad n = \frac{N_\text{c}}{2},\dots,N_\text{c}-1. \tag{18}$$

Since $\tilde{\mathbf{d}}$ is conjugate symmetric by construction, the inverse discrete Fourier transform (IDFT) vector $\mathbf{d}$ is composed only of real numbers. The cyclic prefix is then added on and the transmit voltages $x[m]$ are generated. Observe that we need an extra $L-1$ time instants to send over the N_c voltages $\hat{x}[0],\dots,\hat{x}[N_\text{c}-1]$.

Unpacking the Data

At the output of the DFT of the received voltage vector $\mathbf{y}$ we have the complex vector $\tilde{\mathbf{y}}$. Taking complex conjugate operation on both sides of Equation (10):

$$\begin{aligned}
\tilde{y}^*_n &= \tilde{h}^*_n \tilde{d}^*_n + \tilde{w}^*_n && (19)\\
&= \tilde{h}_{N_\text{c}-1-n}\tilde{d}_{N_\text{c}-1-n} + \tilde{w}_{N_\text{c}-1-n} && (20)\\
&= \tilde{y}_{N_\text{c}-1-n}. && (21)
\end{aligned}$$

Here we used Equations (9),(15), (14) to verify Equation (13). This means that *half* the DFT outputs are *redundant* and can be discarded. Using the first half, we arrive at the following N_c received voltages $\hat{y}[0],\dots\hat{y}[N_\text{c}-1]$:

$$\begin{aligned}
\hat{y}[2n] &\stackrel{\text{def}}{=} \Re\left[\frac{\tilde{h}^*_n}{|\tilde{h}_n|}\tilde{y}_n\right] && (22)\\
&= |\tilde{h}_n|\hat{x}[2n] + \hat{w}[2n] && (23)\\
\hat{y}[2n+1] &\stackrel{\text{def}}{=} \Im\left[\frac{\tilde{h}^*_n}{|\tilde{h}_n|}\tilde{y}_n\right] && (24)\\
&= |\tilde{h}_n|\hat{x}[2n+1] + \hat{w}[2n+1], \quad n = 0,\dots,\frac{N_\text{c}}{2}-1. && (25)
\end{aligned}$$

Here $\hat{w}[\cdot]$ is also white Gaussian with zero mean and variance σ^2 (explored in a homework exercise). Putting together Equations (23) and (25), we can write

$$\hat{y}[n] = \hat{h}_n \hat{x}[n] + \hat{w}[n], \quad n = 0,\dots,N_\text{c}-1 \tag{26}$$

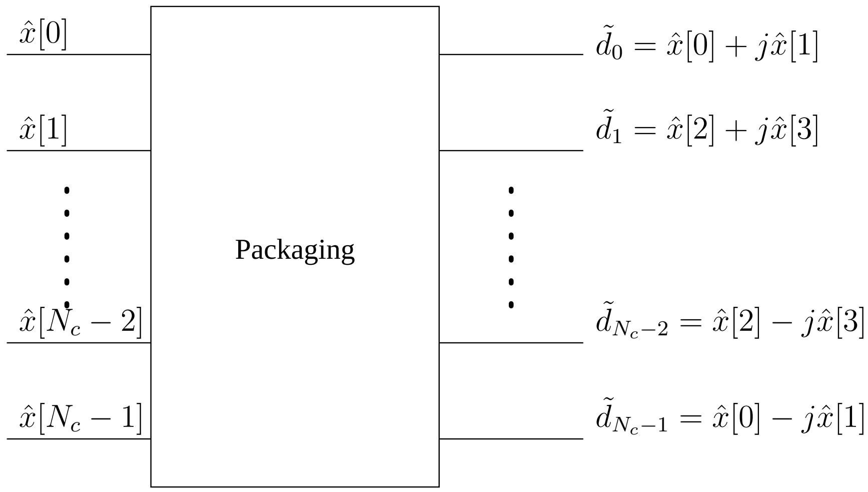

Figure 4: The packaging at the transmitter maps AWGN coded voltages of the sub-channels to the transmit voltages on the ISI channel.

where we have written

$$\hat{h}_n \stackrel{\text{def}}{=} \begin{cases} |\tilde{h}_{\frac{n}{2}}| & n \text{ even} \\ |\tilde{h}_{\frac{n-1}{2}}| & n \text{ odd.} \end{cases} \tag{27}$$

We can repeat the OFDM operation over the next block of N_c symbols (taking up an extra $L-1$ time instants, as before) and since the wireline channel stays the same, we have the end-to-end relation (as in Equation (26)):

$$y[N_\text{c} + n] = \hat{h}_n \hat{x}[N_\text{c} + n] + \hat{w}[N_\text{c} + n], \quad n = 0, \ldots, N_\text{c} - 1. \tag{28}$$

By repeating the OFDM operation over multiple N_c blocks, we have thus created the parallel AWGN channel promised in Equation (2). This packaging and unpacking of data as appended to the basic OFDM scheme (cf. Figure 3) is depicted in Figures 4 and 5.

Prefix: Converting a Topelitz Matrix into a Circulant Matrix

The OFDM scheme uses $L-1$ extra time samples to transmit the block of N_c symbols. The resulting linear transformation between the input and output voltage vectors can now be written as:

$$\mathbf{y} = \mathbf{H}_\text{c}\mathbf{x} + \mathbf{w}, \tag{29}$$

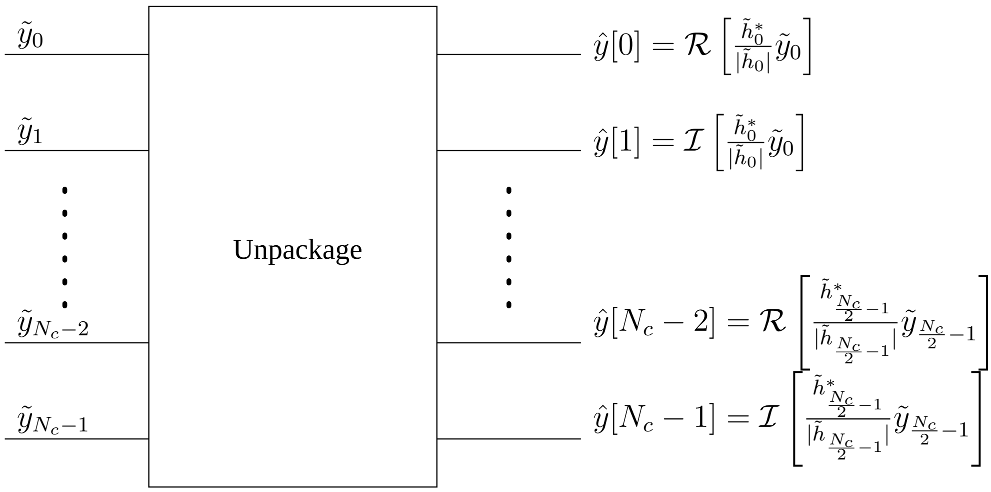

Figure 5: The unpacking at the receiver maps the DFT outputs into the outputs suitable for decoding the coded voltages of the sub-channels.

where $\mathbf{H}_c$ is the *circulant* version of the original Toeplitz matrix $\mathbf{H}$. For example, with $L = 3, N_c = 5$, the Toeplitz and its corresponding circulant version are:

$$\mathbf{H} = \begin{bmatrix} h_0 & 0 & 0 & 0 & 0 \\ h_1 & h_0 & 0 & 0 & 0 \\ h_2 & h_1 & h_0 & 0 & 0 \\ 0 & h_2 & h_1 & h_0 & 0 \\ 0 & 0 & h_2 & h_1 & h_0 \end{bmatrix}, \quad \mathbf{H}_c = \begin{bmatrix} h_0 & 0 & 0 & h_2 & h_1 \\ h_1 & h_0 & 0 & 0 & h_2 \\ h_2 & h_1 & h_0 & 0 & 0 \\ 0 & h_2 & h_1 & h_0 & 0 \\ 0 & 0 & h_2 & h_1 & h_0 \end{bmatrix}. \tag{30}$$

Now the important fact:

All circulant matrices are *universally* diagonalized by the same $\mathbf{Q}_1, \mathbf{Q}_2$.

Furthermore, if we allow for linear transformations with complex entries, $\mathbf{Q}_1, \mathbf{Q}_2$ are the IDFT and DFT matrices! Here the DFT matrix is defined as usual: the (k, l) entry is

$$d_{kl} = \frac{1}{\sqrt{N_c}} e^{-j2\pi\frac{kl}{N_c}} \quad k, l = 0, 1, \cdots, N_c - 1. \tag{31}$$

Similarly, the (k, l) entry of IDFT matrix is given by

$$\frac{1}{\sqrt{N_c}} e^{j2\pi\frac{kl}{N_c}} \quad k, l = 0, 1, \cdots, N_c - 1. \tag{32}$$

A practical implication is that neither transmitter nor the receiver needs to know the channel matrix in order to convert the ISI channel into a parallel AWGN channel.

Capacity of the Wireline Channel

The cyclic prefix added a penalty of $L-1$ time samples over a block of length N_c. By choosing the block length very large, we can amortize the cost (in data rate) of the cyclic prefix overhead. So, the cyclic prefix operation can be thought of one without loss of much generality. Further, the IDFT and DFT operations (at the transmitter and receiver, respectively) are invertible matrix operations. So, they can be done without loss of generality as well (in the sense they can always be undone, if need be). So, the OFDM way of converting the wireline channel into a parallel AWGN channel entails no loss of reliable data rate of communication. We conclude:

> The capacity of the wireline channel is the same as that of the corresponding parallel AWGN channel for large values of N_c.

Now, we know the capacity and the corresponding reliable communication techniques that achieve it in the context of the basic AWGN channel. A natural strategy would be to exploit this deep understanding in this slightly different take on the basic version of the AWGN channel. One such strategy to transmit a packet over the parallel AWGN channel would be to divide the data into the subpackets, one for each subchannel and then code and transmit each subpacket over the corresponding subchannel separately. However, since the channel coefficients of the subchannels are different, if we divide the total power equally, they will have different SNRs and hence different capacities. Intuitively, we should allot more power to the channel with the larger coefficient. It turns out that coding separately over each subchannel is indeed the best strategy, i.e., if the powers are allocated properly, this strategy will achieve the capacity of the parallel channel with N_c subchannels.

Let P_k be the power used on the subchannel k. The rate we can achieve over the k^{th} subchannel is

$$R_k = \frac{1}{N_c + L - 1} \frac{1}{2} \log_2 \left(1 + \frac{\tilde{h}_k^2 P_k}{\sigma^2} \right) \quad \text{bits/channel use} \quad k = 0, 1, \cdots, N_c - 1 \tag{33}$$

where $\tilde{h}_k$ is the coefficient of the k^{th} subchannel. We normalize the rate by $N_c + L - 1$ as we use each subchannel once in each block of $N_C + L - 1$ channel uses. The overhead of $L - 1$ is due to the cyclic prefix used in OFDM. We have the total power constraint that the total power used over all subchannels should not exceed P, i.e.,

$$\sum_{k=0}^{N_c - 1} P_k \leq P \tag{34}$$

The total rate is the sum of the individual capacities of the subchannels. Thus, the capacity is achieved by the solution of the following optimization problem.

$$\begin{aligned} &\max_{P_0,P_1,\cdot,P_{N_c-1}} \tfrac{1}{2(N_c+L-1)} \textstyle\sum_{k=0}^{N_c-1} \log_2\left(1+\tfrac{P_k\tilde{h}_k^2}{\sigma^2}\right) \\ &\text{subject to} \quad \textstyle\sum_{k=0}^{N_c-1} P_k \leq P \end{aligned} \tag{35}$$

The solution of this optimization problem is

$$P_k^* = \left(\frac{1}{\lambda} - \frac{\sigma^2}{|\tilde{h}_k|^2}\right)^+ \tag{36}$$

where the function $(x)^+ \stackrel{\text{def}}{=} \max(x,0)$ and λ is chosen so that it satisfies

$$\sum_{k=0}^{N_c-1} P_k^* = P. \tag{37}$$

Thus, the capacity of the parallel channel is

$$C_{N_c} = \frac{1}{2(N_c+L-1)} \sum_{k=0}^{N_c-1} \log_2\left(1+\frac{\tilde{h}_k^2}{\sigma^2}\left(\frac{1}{\lambda}-\frac{\sigma^2}{\tilde{h}_k^2}\right)^+\right). \tag{38}$$

How do the parallel channel coefficients $\tilde{h}_k$, the DFT (Discrete Fourier Transform) of the tap coefficients h_l of the original wireline channel,

$$\tilde{h}_k = \sum_{l=0}^{L-1} h_l e^{-j2\pi\frac{kl}{N_c}} \quad k = 0,1,\cdots N_c-1, \tag{39}$$

change with the block size N_c? To understand this, let us first consider the *discrete time Fourier transform* $H(f)$ of the channel: it is given by

$$H(f) = \sum_{l=0}^{L-1} h_l e^{-j2\pi\frac{lf}{W}} \quad f \in [0,W]. \tag{40}$$

Thus the channel coefficient $\tilde{h}_k$ is $H(f)$ evaluated at $f = \frac{kW}{N_c}$. Now we see that as the number of subcarriers N_c grows, the frequency width $\frac{W}{N_c}$ of each subcarrier goes to zero and they represent finer and finer sampling of the continuous spectrum. Further, the overhead $L-1$ becomes negligible as compared to N_c. Thus, as N_c goes to infinity, the capacity achieved by the OFDM channel reaches the capacity of the wireline channel (cf. Equation (38) and the substitution $\sigma^2 = N_0/2$):

$$C = \int_0^W \frac{1}{2}\log_2\left(1+\frac{2P^*(f)|H(f)|^2}{N_0}\right) \mathrm{d}f \quad \text{bits/s}. \tag{41}$$

Here we have denoted the optimal power allocation function

$$P^*(f) = \left(\frac{1}{\lambda} - \frac{N_0}{2|H(f)|^2}\right)^+, \tag{42}$$

where the constant λ satisfies

$$\int_0^W P^*(f)\mathrm{d}f = P. \tag{43}$$

Looking Ahead

This brings us to an end to our study of reliable communication over the wireline channel. In the next several lectures we turn to reliable wireless communication. We will see that though wireless communication has its own unique challenges, our understanding of the wireline channel forms a basic building block. We will also see that OFDM will come to play an important role in reliable wireless communication.