Cellular OFDMA-TDMA

- This has become the physical layer in the latest commercial systems based on IEEE 802.16 series of standards for broadband wireless access networks.

- Since, in general, each OFDM carrier can be allocated to different users over time, these systems can be said to employ OFDMA-TDMA.

- We will study resource allocation problems in OFDMA-TDMA cellular systems by formulating various constrained optimization problems, the decision variables being the carriers assigned to various users, the powers used on each carrier by each user, how much data each user sends in a slot, and so on. The constraint could be on average total power.
Overview

- We focus on the case of a single isolated cell. The basic model is of a sequence of OFDMA-TDMA frames, each with a certain number of symbols.
- Given the user requirements, the problem in each frame is to determine how much of the frame to allocate to the uplink and downlink data, and how much data of each user to carry in each direction.
- Since the subcarriers have time varying fading there is also the problem of determining the transmission powers.
- We begin with the simplest problem of one user transmitting over a single carrier with time varying channel gains that stay constant over frames.
The General Model

- The unit of time over which the OFDMA-TDMA scheduling takes place is the **frame**, whose duration is denoted by \( \tau \).
- Each frame has \( k \) OFDMA symbols.
- In **time division duplex** version, each frame is partitioned into a **downlink subframe** and an **uplink subframe**.
- We denote the number of MSs by \( M \) and the number of carriers by \( N \).
- The downlink part of each frame is used to send data to the MSs; such data would be queued in per MS queues at the BS. The uplink part of the frame is used to send data from MSs to the destinations, via the BS.
- If **Internet access** is a major application, then the downlink part of the frame would be substantially larger than the uplink part.
The General Model (ctd.)

- Resource allocation decisions are made only at the frame boundaries, as the information required for resource allocation (i.e., the channel gains, and the backlogs of the various MSs) is updated at this slow scale. The nth frame is the interval \(((n-1)\tau, n\tau)\).

- Let us consider one direction of transmission; say, the downlink. The power gain for MS i on the j-th carrier is \(H_{j,n-1}^{(i)}\) during the nth frame. Thus, we are assuming the channel coherence time \(T_c > \tau\) in order that the channel gains can be assumed to be constant over a frame.

- The data buffer occupancy at the downlink queue for MS i at the beginning of frame n is denoted by \(Q_{n-1}^{(i)}\).
The General Model (ctd.)

- We assume the new data arrive at the end of each frame. Let the arrivals in the nth frame be denoted by $B_n^{(i)}$ which are assumed to arrive at the frame boundary between Frame n and Frame n+1.

- The amount of data transmitted in the downlink to MS i in Frame n is denoted by $S_n^{(i)}$. Thus, the evolution equation for downlink buffer of MS i is

$$Q_{n+1}^{(i)} = Q_n^{(i)} - S_n^{(i)} + B_{n+1}^{(i)}$$

where $Q_0^{(i)} = B_0^{(i)}$. 
downlink queues at the BS

MS 1
MS 2
MS M

K symbols per frame

uplink queues at the MSs

MS 1
MS 2
MS M

Figure 6.1
Figure 6.2
Resource Allocation over a Single Carrier

- We consider the situation in which a single channel is being used to transport data for a single MS. We use the model

\[ Y_k = G_k X_k + Z_k \]

where \( k \) indexes the symbol, and \( H_k (=|G_k|^2) \) is the sequence of power gains.

- Now consider those frames in which \( H_n = h \), for some nonnegative number \( h \). Suppose in these frames we use the transmit power \( P(h) \), that is

\[
\lim_{t \to \infty} \frac{\sum_{n=0}^{t-1} \sum_{\{\text{symbol } k \in \text{ frame } n\}} |X_k|^2}{\sum_{n=0}^{t-1} I_{\{H_n=h\}} K} = p(h)
\]

- Then, over such frames, the channel capacity achievable (in bits per frame) is

\[
C = \frac{K}{\ln 2} \ln \left(1 + \frac{h P(h)}{\sigma^2}\right) = k \ln \left(1 + \frac{h P(h)}{\sigma^2}\right)
\]
Let the channel power gain process $H_k$ be stationary and ergodic, taking values in the finite set $H = \{h_1, h_2, \ldots, h_J\}$, where $J = |H|$. Let us denote by $P_j = P(h_j)$, the power used when the channel power gain is $h_j$, write power control as the $J$ vector $P$. Then, for this power control, the average channel capacity over the first $t$ frames (for large $t$) is given by

$$C_t(P) = \sum_{h_j \in H} \left( \frac{1}{t} \sum_{n=0}^{t-1} I_{(H_n = h_j)} \right) k \ln(1 + \frac{h_j P(h_j)}{\sigma^2})$$

bits per frame. But this is exactly the same as writing

$$C_t(P) = \frac{1}{t} \sum_{n=0}^{t-1} k \ln(1 + \frac{H_n P(H_n)}{\sigma^2})$$

Now, taking $t$ to $\infty$, using the Ergodic Theorem, and defining the resulting limit as $C(P)$, we find that

$$C(P) := kE(\ln(1 + \frac{HP(H)}{\sigma^2}))$$

where $H$ denotes the marginal random variables $H_n$. 
Power Control for Optimal Service Rate (ctd.)

- We wish to ask the question: “What is the best power control?” First we impose a power constraint on the input symbols,

\[
\lim_{k' \to \infty} \frac{1}{k'} \sum_{k=1}^{k'} |X_k|^2 \leq P
\]

- To calculate the left-hand side of this expression, we consider the average symbol power over the first \( t \) frames, and let \( t \to \infty \).

\[
\lim_{t \to \infty} \frac{1}{Kt} \sum_{n=0}^{t-1} \sum_{\text{symbol } k \in \text{ frame } n} |X_k|^2 = \lim_{t \to \infty} \frac{1}{Kt} \sum_{h_j \in H} \sum_{n=0}^{t-1} I_{\{H_n = h_j\}} \sum_{\text{symbol } k \in \text{ frame } n} |X_k|^2 = \lim_{t \to \infty} \sum_{h_j \in H} \left( \frac{1}{K} \sum_{n=0}^{t-1} I_{\{H_n = h_j\}} \sum_{\text{symbol } k \in \text{ frame } n} |X_k|^2 \right) \frac{1}{t} \sum_{n=0}^{t-1} I_{\{H_n = h_j\}} = \sum_{h_j \in H} P_j g_j
\]

- Where \( g_j \) is the fraction of symbols that find the channel power attenuation \( h_j \).
We have also used the fact that in those symbols in which the power gain is $h_j$, the average transmitter power used is $P_j$. Then, we can write

$$C(P) = k \sum_{h_j \in H} g_j \ln(1 + \frac{h_j P_j}{\sigma^2})$$

This leads to the following optimization problem.

$$\max \sum_{h_j \in H} g_j \ln(1 + \frac{h_j P_j}{\sigma^2})$$

subject to

$$\sum_{h_j \in H} P_j g_j \leq \bar{P}$$

$$P_j \geq 0 \text{ for every } h_j \in H.$$ 

This is a nonlinear optimization problem with a concave objective function and linear constraint.
For each power control $P$, and a number $\lambda \geq 0$, consider the function defined as follows:

$$L(P, \lambda) := \sum_{h_j \in H} g_j \ln(1 + \frac{h_j P_j}{\sigma^2}) - \lambda \sum_{h_j \in H} g_j P_j$$

The strict concavity of $L(P, \lambda)$ in $P$ implies that a locally maximizing power vector will also provide a global maximum over $[0, \infty)$. Rewriting

$$L(P, \lambda) := \sum_{h_j \in H} g_j (\ln(1 + \frac{h_j P_j}{\sigma^2}) - \lambda P_j)$$

We can maximize the expression term by term for each channel gain $h_j$. It then follows that the power vector $P_{\lambda}$ that maximizes $L(P, \lambda)$ has the form

$$P_{\lambda, j} = \left( \frac{1}{\lambda} - \frac{\sigma^2}{h_j} \right)^+$$
Thus, for chosen $\lambda$, if we maximize the function $L(P, \lambda)$ the capacity with maximizing power control, $P_{\lambda,j}$, is given by

$$C(P_{\lambda}) = k \sum_{h_j \in H} g_j \ln(1 + \frac{h_j P_{\lambda,j}}{\sigma^2})$$

bits per frame, and the average power is given by

$$\bar{P}_{\lambda} := \sum_{h_j \in H} P_{\lambda,j} g_j$$

Given the set of values taken by the channel power gain process (i.e., $H$), let $h^{(1)}$, $h^{(2)}$, ..., $h^{(J)}$, be these values in descending order. Thus, $h^{(1)}$ is the best channel gain and $h^{(J)}$ the worst.
Figure 6.3

Power allocated to the best channel state
Choosing $\lambda$

- Consider another vector $P$, with
  \[ \sum_{h \in H} P_j g_j \leq \bar{P}_\lambda \]

- Since $P_\lambda$ maximizes $L(P,\lambda)$ for a given $\lambda$, over all nonnegative power controls, we have, for the particular $P$ just chosen,
  \[ C_\lambda (P_\lambda) - \lambda \bar{P}_\lambda \geq C (P) - \lambda \sum_{h \in H} P_j g_j \]

- Which implies that
  \[ C(P_\lambda) \geq C(P) + \lambda (\bar{P}_\lambda - \sum_{h \in H} P_j g_j) \geq C(P) \]

- We conclude $P_\lambda$ is optimal for the original constraint optimization problem for all power controls that satisfy the power constraint $\bar{P}_\lambda$. 
Choosing $\lambda$ (ctd.)

- We can conclude that $P_\lambda$ is optimal among all power controls that satisfy the power constraint $P_\lambda$. It follows that if we can choose $\lambda$ such that $P_\lambda = P$, then the resulting power control will be optimal with the power constraint $P$. That is, we need $\lambda$ so that

$$\sum_{h_j \in H} \left( \frac{1}{\lambda} - \frac{\sigma^2}{h_j} \right)^+ g_j = P$$

- Let us denote this particular $\lambda$ by $\lambda(P)$. Then, the optimal power control becomes $P_{\lambda(P)}$, with

$$P_{\lambda(P),j} = \left( \frac{1}{\lambda(P)} - \frac{\sigma^2}{h_j} \right)^+$$

- For optimal capacity, we also get

$$C_{opt}(P) = k \sum_{\{h_j \in H, h_j > \sigma^2 \lambda(P)\}} g_j \ln\left( \frac{h_j}{\sigma^2 \lambda(P)} \right)$$
Consider the single MS version of the buffer evolution. We have

\[ Q_{n+1} = Q_n - S_n + B_{n+1} \]

where \( Q_0 = B_0 \), and \( S_n (\leq Q_n) \) is the number of bits served from the buffer in frame \( n \). Let us assume that arbitrary number of bits are served in the \( n \)th frame, with \( S_n \leq Q_n \), and such that the power constraint \( P \) is respected.

Let \( A_t, \ t \geq 0 \), be defined by

\[ A_t = \sum_{n=0}^{t} B_n \]

\( A_t \) is the cumulative arrivals until the end of the \( t \)-th frame. Let

\[ \lim_{t \to \infty} \frac{A_t}{t} = a \]
Stability of the Buffer Process

If $a < C_{opt}(P)$ then the buffer process $Q_n$ will be stable, that is, it converges in distribution to a random variable that is finite with probability 1.

- On the other hand, if $a > C_{opt}(P)$ then $Q_n$ will “blow up”.

- Note that

$$C_{opt} = \lim_{t \to \infty} \frac{1}{t} \sum_{n=0}^{t} C_n$$

- Consider any service sequence $S_n$, $n \geq 0$. It is clear that

$$\lim_{t \to \infty} \frac{1}{t} \sum_{n=0}^{t} S_n \leq C_{opt}$$

- On the other hand, for $t \geq 0$,

$$Q_t = \sum_{j=0}^{t} B_j - \sum_{j=0}^{t-1} S_j$$
Stability of the Buffer Process (ctd.)

- Now, if \( a > C_{opt}(P) \), we have

\[
a = \lim_{t \to \infty} \frac{1}{t} \sum_{n=0}^{t} B_n > C_{opt}(P) \geq \lim_{t \to \infty} \frac{1}{t} \sum_{n=0}^{t-1} S_n
\]

- Thus, we get

\[
\lim_{t \to \infty} \frac{1}{t} (\sum_{n=0}^{t} B_n - \sum_{n=0}^{t-1} S_n) > 0
\]

which means as \( t \to \infty \), \( \sum_{n=0}^{t} B_n - \sum_{n=0}^{t-1} S_n \to \infty \), that is, \( Q_n \) goes to \( \infty \).
Stability of the Buffer Process (ctd.)

It follows that, for a given arrival rate $a$, there is a minimum power ($P_a = \inf\{P: C_{opt}(P) > a\}$) that is needed to ensure stability of link buffers.

However, if $P$ is greater than but very close to $P_a$ then the delays will be large.

If it is required that all data are transmitted in the next frame after which they arrive, the rate required in the $n$-th frame will be $S_n = B_n$, $n \geq 0$. Then, the power $P_n$, required in the $n$-th frame will be obtained by solving

$$B_n = k \ln(1 + \frac{H_n P_n}{\sigma^2})$$
Figure 6.4

The graph shows a downward sloping curve representing the relationship between mean delay and a parameter $P$. The axes are labeled as $\bar{P}(d)$ and $P$, with specific points $\bar{P}_a$, $\bar{P}(d)$, and $\bar{P}_{a,\text{max}}$ indicated on the vertical axis and $P$ on the horizontal axis. The graph illustrates how mean delay varies with $P$. 

Copyright © 2008, Anurag Kumar, D. Manjunath and Joy Kuri. Published By Elsevier Inc. All rights reserved.
Stability of the Buffer Process (ctd.)

- From which we see

$$P_n = \frac{\sigma^2}{H_n} (e^{\frac{B_n}{k}} - 1)$$

- It follows that the average power required if all data need to be sent in the frame after the one in which they arrive is given by

$$\bar{P}_{a,\text{max}} := \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \frac{\sigma^2}{H_j} (e^{\frac{B_j}{k}} - 1) = E(\frac{\sigma^2}{H} (e^{\frac{B}{k}} - 1))$$

where $H$ and $B$ denote the random variables that have ergodic distributions of the fading process and the arrival process.
Delay Minimizing Power Control

- For the queue evolution
  \[ Q_{n+1} = Q_n - S_n + B_{n+1} \]

- And a power constraint \( P \), we wish to determine a sequence of power allocations, \( P_n \), so that the buffer delay is minimized. We will assume that the number of bits transmitted, \( S_n (\leq Q_n) \), and the power required, \( P_n \), are related by Shannon capacity formula,
  \[ S_n = k \ln(1 + \frac{H_n P_n}{\sigma^2}) \]

- If \( S_n (\leq Q_n) \) bits need to be transmitted in a frame in which the channel gain is \( H_n \) then the power required in that frame is
  \[ P_n = \frac{\sigma^2}{H_n} \frac{S_n}{e^{k} - 1} \]
Delay Minimizing Power Control (ctd.)

- We will assume that the channel gain at each frame is known at the transmitter at the beginning of the frame. Also, the sequence of arrivals, $B_i$, $0 \leq i \leq n$, until the beginning of Frame $n+1$ is, of course, known to the transmitter. A control policy, $\pi$, prescribes for each $n$, and possible history up to the beginning of Frame $n+1$ (i.e., $(Q_0, B_0, H_0, P_0)$, $(Q_1, B_1, H_1, P_1)$, ..., $(Q_n, B_n, H_n)$) the power control $P_n$ to be used during Frame $n+1$. Let $E^{\pi}(.)$ denote the expectations when the policy $\pi$ is used. Define

\[
\hat{Q}(\pi) = \limsup_{n \to \infty} \frac{1}{n} E^{\pi}(\sum_{k=0}^{n-1} Q_k) \quad \text{and} \quad \hat{P}(\pi) = \limsup_{n \to \infty} \frac{1}{n} E^{\pi}(\sum_{k=0}^{n-1} P_k)
\]

- The problem is

\[
\min_{\{ \pi : \hat{P}(\pi) < P \}} \hat{Q}(\pi)
\]

- It reduces to minimizing

\[
L(\pi, \beta) = \limsup_{n \to \infty} \frac{1}{n} E^{\pi}(\sum_{k=0}^{n-1} (Q_k + \beta P_k))
\]

over all policies $\pi$. 
Delay Minimizing Power Control (ctd.)

- Suppose $\pi^*$ minimizes $L(\pi, \beta)$, for a given $\beta$, Let $E^\pi(.)$ denote the expectations when the policy $\pi$ is used. We have:
  \[ L(\pi^*, \beta) = \hat{Q}(\pi^*) + \beta \hat{P}(\pi^*) \leq L(\pi, \beta) \leq \hat{Q}(\pi) + \beta \hat{P}(\pi) \]

- Suppose $\hat{P}(\pi^*) = P$, then using the inequality just derived, we get
  \[ \hat{Q}(\pi^*) \leq \hat{Q}(\pi) - \beta(P - \hat{P}(\pi)) \leq \hat{Q}(\pi) \]

  whenever $\hat{P}(\pi) \leq P$. It follows that $\pi^*$ is optimal for the original optimization problem for $Q_n$.

- When the arrival process and channel gain process are Markov then this is an average cost Markov decision problem.
Figure 6.5

The diagram illustrates the relationship between power, delay, and their mean values. The axes are labeled as follows:

- \( P_a,_{\text{max}} \) on the vertical axis
- \( \beta(P) \) on the horizontal axis
- \( P_a \) on the right vertical axis
- \( \beta_{\text{max}} \) on the right horizontal axis

Key features include:

- The average power is represented by \( \bar{P} \).
- The mean delay is represented by \( \bar{\beta}(P) \).
- The optimum mean delay is shown between the two curves.

The diagram highlights the trade-off between power and delay, indicating that as power increases, delay decreases, and vice versa, with an optimum point \( P_a \) and \( \beta_{\text{max}} \) where both are balanced.
Figure 6.6
Multicarrier Resource Allocation: Downlink Single MS Case

- We have for \(1 \leq j \leq N\) (the carriers), and \(k \geq 1\) (OFDM symbols),
  \[
  Y_{j,k} = G_{j,k} X_{j,k} + Z_{j,k}
  \]
  where \(Z_{j,k}\) are Gaussian random variables that are i.i.d. in \(j\) and in \(k\),
  with variance \(\sigma^2\). There is a total average power constraint
  \[
  \lim_{l \to \infty} \frac{1}{l} \sum_{k=1}^{l} \sum_{j=1}^{N} |X_{j,k}|^2 \leq \bar{P}
  \]

- We wish to maximize the total bit rate achieved by the MS, subject to this power constraint.

- Let us assume that \(H_{j,k} = h_j, 1 \leq j \leq N\), so that the power gain on each carrier is constant over time. This may be a good approximation to use when the fading is slow compared to an OFDMA frame size.
Multicarrier Resource Allocation: Downlink Single MS Case (ctd.)

The problem is now to split the power \( \bar{P} \) over the \( N \) carriers. Let \( \mathbf{P} = (P_1, P_2, \ldots, P_N) \) be the vector of powers assigned to the carriers, such that \( \sum_{j=1}^{n} P_j \leq \bar{P} \). The total capacity achievable with this power allocation is

\[
C(\mathbf{P}) = k \sum_{j=1}^{N} \ln(1 + \frac{h_j P_j}{\sigma^2})
\]

The optimal power allocation becomes

\[
\max \sum_{j=1}^{N} \ln(1 + \frac{h_j P_j}{\sigma^2})
\]

subject to \( \sum_{j=1}^{N} P_j \leq \bar{P} \)

\( P_j \geq 0 \) for every \( j, 1 \leq j \leq N \)
Similar to the approach used in the previous optimization problem, there is a power price $\lambda(\bar{P})$ defined by

$$\sum_{j=1}^{N} \left( \frac{1}{\lambda(\bar{P}) h_j} - \frac{\sigma^2}{h_j} \right)^+ = \bar{P}$$

Then the optimal power allocation becomes

$$P_{\lambda(P),j} = \left( \frac{1}{\lambda(\bar{P}) - \frac{\sigma^2}{h_j}} \right)^+$$

For the optimal total capacity over the carriers, $C_{opt}(\bar{P})$, we have

$$\frac{\partial C_{opt}(\bar{P})}{\partial \bar{P}} = \lambda(\bar{P})$$
Case of Time Varying Fading Process

We now include a time varying fading process, $H_{j,k}$, in the model. The stationary and ergodic process $H_{j,k}$, $k \geq 1$, has marginal probabilities $g_j(h), h \in H, 1 \leq j \leq N$

Let $P = (P_1, P_2, \ldots, P_N)$, with $\sum_{j=1}^{N} P_j \leq P$. If power $P_j$ is utilized optimally on Carrier $j$, let the corresponding rate be denoted by $C_{opt,j}(P_j), 1 \leq j \leq N$. The problem then becomes

$$\max \sum_{j=1}^{N} C_{opt,j}(P_j)$$

subject to

$$\sum_{j=1}^{N} P_j \leq P, \quad P_j \geq 0, \quad 1 \leq j \leq N$$

Again, defining

$$L(P, \lambda) = \sum_{j=1}^{N} C_{opt,j}(P_j) - \lambda \sum_{j=1}^{N} P_j$$

We obtain, for $j, 1 \leq j \leq N$,

$$\frac{\partial C_{opt,j}(P_j)}{\partial P_j} = \lambda$$

Thus, the solution is to use the same power price on each carrier, and vary this common price until all the power is utilized.
Figure 6.7
Multiple MSs

- We have the model

\[ Y_{i,j,k} = G_{i,j,k} X_{j,k} + Z_{i,j,k} \]

for the i-th MS on the j-th carrier. Suppose a fraction \( \alpha_{i,j} \) of the OFDM blocks carry data for the i-th MS on the j-th carrier. Let us also assume that the channel power gain for MS i on Carrier j (i.e, \( H_{i,j,k} \)) does not vary with k, and thus can be written as \( h_{i,j} \). Then it suffices to assume that the transmissions to the i-th MS on Carrier j are done at the constant power \( P_{i,j} \). Thus, for each carrier j,

\[ \sum_{i=1}^{M} \alpha_{i,j} = 1 \]

and the power constraint becomes

\[ \sum_{j=1}^{N} \sum_{i=1}^{M} \alpha_{i,j} P_{i,j} \leq P \]
Subject to the previous constraint, we wish to maximize the total capacity of the system,

\[
\sum_{j=1}^{N} \sum_{i=1}^{M} \alpha_{i,j} \ln(1 + \frac{h_{i,j}P_{i,j}}{\sigma^2})
\]

From Jensen’s inequality, we observe that

\[
\sum_{i=1}^{M} \alpha_{i,j} \ln(1 + \frac{h_{i,j}P_{i,j}}{\sigma^2}) \leq \ln(1 + \sum_{i=1}^{M} \frac{h_{i,j}P_{i,j}}{\sigma^2}) \leq \ln(1 + \frac{h_{i,j}P_{i,j}}{\sigma^2})
\]

where \( i_j := \arg \max_{1 \leq i \leq M} h_{i,j} \) and \( P_j = \sum_{i=1}^{M} \alpha_{i,j} P_{i,j} \).

Thus, to maximize the total capacity, the approach should be to split the power \( P \) over the \( N \) carriers and use all the power allocated to each channel to serve the MS with the best power gain on that channel. Then we may as well think of a single “MS” with the power gains \( h_j = h_{i_j,j} \), and allocate the power \( P \) so as to maximize this “MS’s” capacity.
WiMAX: The IEEE 802.16 Broadband Wireless Access Standard

- OFDMA has been adopted by IEEE 802.16 series of standards for wireless broadband wireless access. The much talked about WiMAX system, which is being commercially adopted, is a subset of IEEE 802.16 standard. This system provides broadband wireless access for fixed stations. Mobile to provide WiMAX is a subset of IEEE 802.16e that is designed broadband wireless access for mobile stations.

- Mobile WiMAX systems can occupy system bandwidth (i.e., W) from 1.25 MHz to 20 MHz, in the 2.3 GHz, 3.3 GHz, and 3.5 GHz bands. In the initial releases of WiMAX, the downlink traffic and uplink traffic will share the same bandwidth in a time-division-duplex (TDD) fashion.

- As an example, one of the WiMAX profiles is w=5MHz, number of carriers N = 512, number of symbols per frame K = 48 symbols, with a frame time of 5 ms. Various modulation schemes (such as QPSK, 16 QAM, and 64 QAM, etc.) are available for putting the MSs’ bits onto the OFDM symbols.
Figure 6.8