Introduction to Wireless Data Systems

Our discussion so far in this course has primarily focused on 2G/3G voice systems. We will now move on to data systems, which have become the emphasis in 4G/5G. We will see that many differences exist between the traffic characteristics and network topologies of data vs. voice, which lead to different considerations for multiple access.

To summarize, here are the key differences between voice and data network systems:

Voice	**Data**
circuit-based	packet-based
fixed, low rate	variable, high rate
regular cell patterns	more ad-hoc cell patterns
mostly single hop	can be multi-hop
work against channel variations	exploit channel variations

Multi-access: Co-exist or Time-share

Let us first revisit the question of multiple access. For voice, we have seen that CDMA has an advantage due to statistical multiplexing. Is the same true for data?

In general, we could have two approaches for coexistence of data users, like we saw for voice:

(1) TDMA/FDMA-oriented

- Users interleave their transmissions
- Only one user transmits at a time (within one cell and within one frequency channel)

- The active user sends at a high bitrate per Hz

(2) CDMA-oriented

- Use multiple codes to maintain a large number of concurrent transmissions
- Each user/channel sends at a lower bitrate per Hz

In fact, for high data rate services, a TDMA/FDMA-oriented approach (i.e., scheduling only one user at a time) is preferable – the opposite of what we saw for voice.

To see this, consider an example of two users in the same cell, each using a transmit power P. The **Shannon-Hartley theorem** tells us that the channel capacity C (in bits/sec) is related to the channel bandwidth B (in Hz) and the signal-to-interference-noise-ratio SINR as $C = B\log_2(1+\text{SINR})$:

- Under TDMA, only one user transmits at a time, so there is no interference. This leads to an achievable capacity of

  $$C_1 = B\log_2\left(1 + \frac{P}{N}\right), \tag{1}$$

  where N is the noise. At $\frac{P}{N} = 64$, we would have $C_1 \approx 6B$. Each user would get half of this, $3B$.

- Under CDMA, the two users transmit together. This leads to

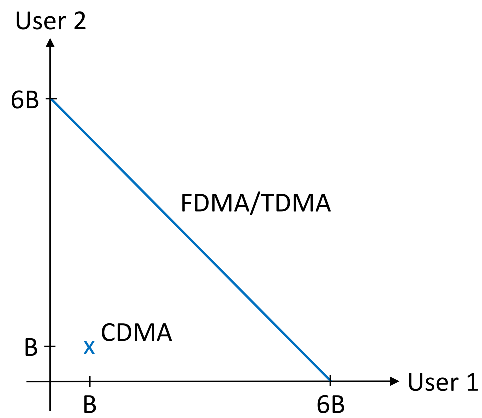
  $$C_2 = B\log_2\left(1 + \frac{P}{P+N}\right) \approx B \tag{2}$$

  for each user, significantly lower than in the TDMA case.

  In CDMA, we also had introduced the processing gain W, which reduced interference for voice substantially. But this processing gain cannot beat the Shannon capacity: the interference will lower to $\frac{P}{W}$, but the usable bandwidth for symbol demodulation is also reduced to $\frac{B}{W}$. In aggregate, we have

  $$\frac{B}{W}\log_2\left(1 + \frac{P}{\frac{P}{W} + N}\right) \leq B\log_2\left(1 + \frac{P}{P+N}\right). \tag{3}$$

Considering the achievable trade-offs between capacities of different users, we can visualize the **capacity region** of a network. In this case, we see that time-interleaving (TDMA) achieves a much larger capacity region than simultaneously transmitting both users' traffic (CDMA):



This is why, even though all 3G standards use CDMA, they also switched to time-interleaving within a cell. CDMA is instead only used to suppress interference from other cells. Then, in 4G, CDMA was completely abandoned in favor of OFDMA.

So, why do the same conclusions not apply to voice? We must keep in mind that:

- Voice is usually at a low, constant rate.
- Voice requires stringent delay guarantees.

Adaptive Modulation and Coding (AMC)

Note that the active user's channel condition will change, e.g., when the user is closer/further away from the base station. As this happens, the SINR will change, which causes the maximum achievable rate in (1) to change. We can employ adaptive modulation and coding (AMC) to attain an instantaneous rate close to $B \log_2(1 + P/N)$.

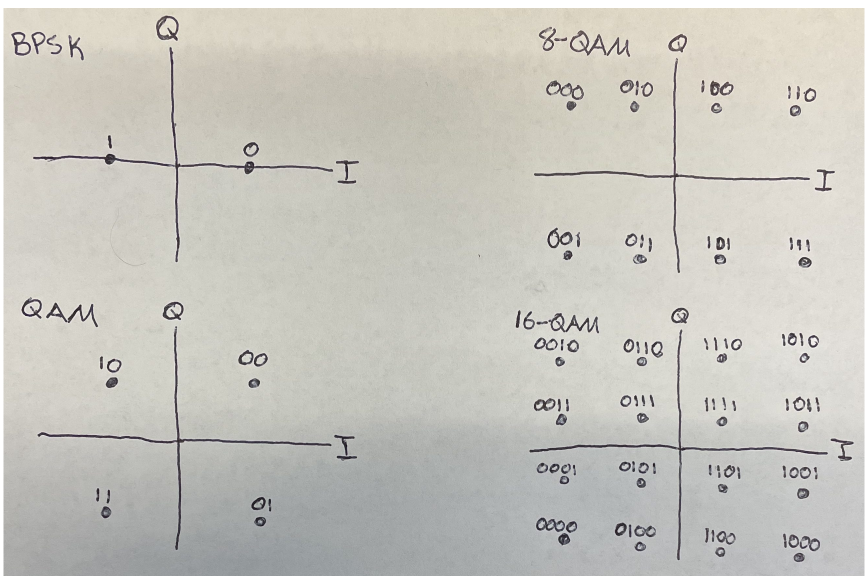
Such adaptivity can be achieved in a few ways:

- *Different processing gain:* Earlier in CDMA, we discussed the impact of the chip rate $1/T_c$ and symbol rate $1/T_s$ in terms of the processing gain $W = T_s/T_c$. We can use a fixed chip rate $1/T_c$ for different symbol rates $1/T_s$, with a higher symbol rate leading to a higher effective data rate. However, a higher symbol rate will also lead to a lower processing gain. Since the demodulator needs a fixed effective SNR,

$$\frac{E_b}{I_0} = \frac{P}{I/W}, \tag{4}$$

it means that the raw SNR P/I needs to be high.

- *Different modulation:* We can change the number of bits per symbol being transmitted by adapting the modulation scheme. **Quadrature Amplitude Modulation** (QAM) employs two carrier waves in each transmission, one in phase $I(t)$ and one quadrature $Q(t)$, which are out of phase by 90°. The resulting transmitted wave is $s(t) = \sin(\omega_c t)I(t) + \cos(\omega_c t)Q(t)$. By varying the amplitudes of $I(t)$ and $Q(t)$, we can create different points in a **constellation diagram** that encode multiple bits in each symbol. In general, N-QAM has N constellation points, with each point containing $\log_2 N$ bits:



  We can increase N further, e.g., to 64-QAM or even 256-QAM. The downside to doing this is that we require a higher SINR at the receiver to differentiate between constellation points. Thus, in AMC, we adapt the modulation format according to the SINR.

- Different amounts of error correction code.

Finally, as the channels of multiple users vary in time and frequency, opportunistic scheduling becomes essential.

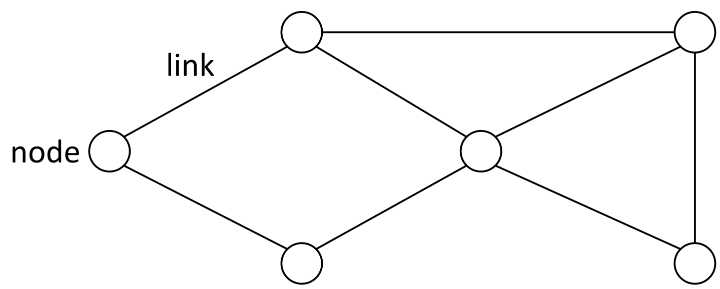
Towards a General Model

Due to these differences that we experience in data networks, the previous design and analysis methodology for voice will not work well anymore. We thus need a new model, that allows us to capture:

- Variable data rates through AMC
- Ad-hoc or even multi-hop topologies
- Opportunistic scheduling

Consider a wireless ad-hoc network with N **nodes** and L **links**. A node is an entity in the network which can transmit and/or receive data (e.g., base stations, mobile devices, and relays in a cellular network), while a link is a transmitter-receiver pair. These links correspond to wireless channels, which may experience time-varying fading.

We commonly represent the logical topology of a network as a **graph**:



In general, packets of data can traverse multiple **hops** wirelessly from source to destination, meaning there are intermediate nodes and links. Also, some nodes may be connected to a wired network.

We will assume for the moment that the node locations are fixed. Such a scenario can occur in many settings:

- *Wireless mesh networks:* Connections between multiple wireless access point nodes, e.g., for a home WiFi network.
- *Wireless sensor networks:* Spatially dispersed sensors collecting information on the environment and reporting it to a central location, e.g., for the Internet of Things (IoT).

As we will se soon, what we really need is that the "link" relationship is fixed (rather than node locations). Cellular/wireless local area networks

(LANs) then become a special case of our model, provided that users are not changing cells.

There are a multitude of design problems to consider:

1. *What is the transmission power and rate that each link should use at each time?* Nearby transmissions can interfere with one another, which makes this challenging to solve optimally.
2. *How to regulate multiple access?* Should nodes transmit simultaneously, or interleaved in time/frequency? Also, which set of links should be active at each time? This leads to the link scheduling problem.
3. *How to find routes from source to destination?* This is important for ad-hoc networks, but it may not be applicable for cellular/wireless LANs when the source and destination have a direct link.
4. *How to make sure that users received the level of service requested?* Relevant metrics for service requirements could include the data rate, service delay, packet loss, fairness across multiple applications, and energy conservation across the network.

We will formulate these questions somewhat independently from the specific technology.

Before proceeding, we should also make a few other remarks about the system model:

1. We assume a **decode-store-and-forward** scheme where data are decoded at the receiving end of each link, and stored in a queue until being forwarded.

   An alternative scheme would be **amplify-and-forward**, in which the intermediate node directly amplifies the received signal and lets the destination node decode it directly. In **cooperative relaying**, the destination combines received signals from the transmitter and intermediate nodes to improve decoding. These techniques can have useful information-theoretic guarantees.
2. We do not handle the question of mobility, i.e., when links come and go. As a result, we also do not consider handoff.

Even with these simplifications, we are still left with a difficult set of problems. Compared to wireline networks:

- Links no longer have fixed capacity.

- The results are highly dependent on the physical layer: transmission signal strength, interference levels, coding/modulation/multi-access, and channel fading can all have significant impacts on the resulting performance.
- Power remains a crucial constraint.