Mobile communications

- Mobile communications is wireless – and almost all wireless communications is radio
- Wireless and radio basics: Electromagnetic waves, transmission, reception, bandwidth

EM waves and early radio

- Radio makes use of electromagnetic radiation propagating through free space
- James Clerk Maxwell showed that a changing electric field induces a changing magnetic field, and vice versa (circa 1861-73)
- Maxwell's Equations showed that electromagnetic waves, composed of changing electric and magnetic fields coupled with each other, can propagate in free space
- Heinrich Hertz devised the first experiment proving the existence of EM waves (1886) – a spark gap transmitter and receiver
- Hertz on the applications of EM waves: "It's of no use whatsoever"
- First uses of radio in communication: 1893-1895; various inventors (Tesla, Marconi, Popov)
- Basic radio system

(Fig. 1)

- transmitter induces alternating current in the transmitting antenna; antenna induces alternating magnetic field; EM wave propagates; EM wave induces alternating electric field in the receiving antenna; electric field induces alternating current in the receiver
- Alternation: Relatively high frequencies are needed

Signal transmission

(Fig. 2)

- f_c is the carrier frequency
- x(t) is a waveform to be transmitted (Baseband signal)
- Modulation: multiply by sinusoid at carrier frequency f_c
- i.e. boost frequency components

(Eq. 1)

Signal reception

(Fig. 3)

- y(t) is the current induced in the receiver antenna
- Amplification: $Ay(t) = x(t)\cos(2\pi f_c t)$
- Demodulation: multiplied by the same sinusoid

(Eq. 2)

- We now have x(t) and high-frequency components – lowpass filter to get rid of them
- And we have recovered x(t)
- **Noise:** The received signal y(t) might be extremely weak, so that the random thermal motions of electrons in the amplifier are significant by comparison
- These random motions introduce an additive, random noise term
- In fact the amplifier output is $x(t)\cos(2\pi f_c t) + n(t)$
- More about this later

Bandwidth

- Under the Fourier transform, x(t) can be represented as a collection of sinusoids: X(f) – can go back and forth between the two domains. (i.e., x(t) and X(f) are equivalent)
- The bandwidth of x(t) is the largest f such that X(f) is negligible
- Generally, larger bandwidth = more quickly changing signal = more information (Nyquist sampling)
- Telephone line: 8 kHz; CD-quality audio: 20 kHz; NTSC video: 6 MHz
- Let X(f) and X*(f) represent Fourier transforms of signals each with bandwidth B

(Fig. 4)

- modulate X(f) with carrier frequency f_c
- (example, for each f < B)
- modulate X*(f) with carrier frequency f_c+2B

(Fig. 5)

- Demodulate X(f) with frequency f_c – and lowpass filter
- Demodulate X*(f) with frequency f_c+2B – and lowpass filter

(Fig. 6)

- Thus two (or more) signals can share frequency space, as long as their bandwidth is finite

Wireless Spectrum

- The collection of all wireless devices in the world must work together to share all the available bandwidth (the wireless spectrum).
- The effectiveness of an antenna at a given frequency is highest if its length is proportional to the wavelength (usually $\lambda/4$ or $\lambda/2$), and $\lambda = c/f$ (c = speed of light = 3×10^8 m/s)
- E.g. f = 1 MHz, λ = 300 m; f = 100 MHz, λ = 3 m; f = 1 GHz, λ = 0.3 m
- This sets a practical limit on the lowest frequencies that are usable (also, less bandwidth is available)

- Atmospheric absorption limits the highest frequencies to around 300 GHz (contemporary applications do not go above 100 GHz) – beyond this is infrared

Name	Band	Application
ELF, SLF, VLF LF	Up to 30 kHz 30—300 kHz	Military, navigation
MF	300 KHz—3 MHz	AM radio
HF	3—30 MHz	Shortwave radio, amateur radio
VHF	30—300 MHz	FM radio, TV, point-to-point comm
UHF	300 MHz—3 GHz	TV, cellphones, ISM bands
SHF EHF	3—30 GHz 30—300 GHz	ISM, Satellites, high-capacity links

- Each slice of the wireless spectrum is licensed for a specific task, and by law cannot be used for any other purpose
- E.g. radio stations are licensed to the station using them by the government, and by law can only be used by that station for approved purposes
- An exception is the "ISM band" (ISM = Industrial, Scientific, Medical) – these are unlicensed bands which may be used for any purpose (subject to power restrictions)
- There are several in various parts of the wireless spectrum, but the most popular is from 2.4-2.5 GHz; used for WiFi, Bluetooth, Zigbee, etc.

Path Loss

- Wireless signals decay with distance from the transmitter

- In free space, the decay is proportional to the size of the wave front (conservation of energy)

(Fig. 7)

- in empty three-dimensional space, the surface area of a sphere is proportional to r^2 (more precisely it's $4\pi r^2$), so this is how fast signals decay
- If signal decay is proportional to r^d, then d is called the *path loss exponent*
- Real-world signals can bounce off the ground, buildings, etc., so in practice d can be between 2 and 4
- On the other hand, if the signal is guided (e.g. by a tunnel), d can be less than 2 (doesn't happen often).

(Fig. 8)

- example: 1 W signal, 100m from tx to rx, if rx power is 0.1 μW with d=2, assuming same constant of proportionality, what is rx power with d=3 and d=4?
- Answer: 1 μW, 0.01 μW

More on signal propagation

- Main effects on radio signals other than attenuation: scattering, shadowing, diffraction, reflection

(Fig. 9)

Antenna Design

- Antennas have an antenna pattern, the relative signal strength (either transmit or receive) from all directions
- Some examples: isotropic, dipole, sector

(Fig. 10)

- A directional antenna is designed by creating physical reflectors/blockers (e.g., dish) or by combining many antenna elements (e.g., array)
- Project signal where wanted, reject interference from unwanted sources, eliminate extra paths

The Mobile Radio Environment: Multipath Fading

- As mentioned, wireless signals can follow many paths from transmitter to receiver (reflection, scattering, etc.)
- These signals have different lengths and therefore different phases

(Fig. 11)

- For a sinusoid with frequency f_c, phase at the rx is

(Eq. 3)

- Say you have two paths, one with distance d_1, the other with distance d_2. They combine at the receiver to form

(Eq. 4)

- Trig identity: sin a + sin b = 2 sin((a+b)/2) cos((a-b)/2)

(Eq. 5)

- since sin is always between -1 and 1, the effect of the fading is to diminish the signal amplitude, possibly to nothing!

- Example. d_1-d_2 = 1.5m, f=100 MHz

Power vs. Amplitude

- Amplitude refers to the field strength of an E or M field
- Power is proportional to the square of the field strength
- So power decaying proportional to r^2 (path loss exponent) means field strength decays proportional to r
- Example. Received amplitude as given last class (Eq. 3) has power

(Eq. 6)

Plane Earth Path Loss

- As a practical example of two-ray multipath fading, consider a signal propagating over a flat plane

(Fig. 12)

- h_T is tx height, h_R is rx height, d is the distance from tx to rx, R is the reflection coefficient (i.e. coefficient multiplied by reflection)
- received signal is

(Eq. 7)

- If d is sufficiently large, R = -1 (from physics)
- Trig identity: sin a – sin b = 2 cos((a+b)/2) sin((a-b)/2)

(Eq. 8)

- $d_2 - d_1$ is … using the approximation $(1+u)^{1/2} \approx 1 + u/2$

(Fig. 9)

(Eq. 10)

- furthermore $\sin u \approx u$ for small u

(Eq. 11)

- so the path loss exponent for plane earth propagation is 4.
- Limitations: Earth needs to be very flat and a very good conductor in order for this to hold (e.g., calm ocean)
- On dry ground, path loss exponent is close to 2 as long as line-of-sight path exists from tx to rx

Rayleigh fading

- What about lots and lots of sinusoids?
- Remember that sin(a+b) = sin a cos b + cos a sin b, so

(Eq. 12)

- The A cos θ and A sin θ terms are random (because path length is random, and phase is random)
- Mean zero, variance small (but the same for both)
- So in the end we get

(Eq. 13)

- Thanks to the central limit theorem, sums of large numbers of random variables approach the Gaussian distribution, so U and V are Gaussian random variables

- Another trig identity: U cos a + V sin a = $(U^2+V^2)^{1/2}\sin(a+b)$, where

(Eq. 14)

- The signal strength is then $(U^2+V^2)^{1/2}$, which has the Rayleigh distribution, which has CDF

(Eq. 15)

- The extra phase doesn't matter – can be tracked
- r^2 is the average power of the sinusoid (given)
- tells us the probability that the signal strength will be less than a given value in an environment with lots of paths
- Example: Average power is 1 W. What is the probability that the signal strength falls below 1?

Flat vs. frequency-selective fading

- How different is fading from frequency to frequency?
- Recall Eq. 3

(Eq. 16)

- if d_1-d_2 is large, a small change in frequency will have a huge effect (frequency-selective fading)
- on the other hand if d_1-d_2 is small, a small change in frequency will have a negligible effect (flat fading)

Rayleigh Fading

- Example: Average power is 1 W. What is the probability that the signal amplitude falls below 1?
- Recall Eq. 9 from last week's notes

(Eq. 17)

Flat vs. frequency-selective fading

- How different is fading from frequency to frequency?
- Recall Eq. 3 from last week's notes

(Eq. 18)

- if d_1-d_2 is large, a small change in frequency will have a huge effect (frequency-selective fading)
- on the other hand if d_1-d_2 is small, a small change in frequency will have a negligible effect (flat fading)
- note that $(d_1-d_2)/c$ is equal to the time delay of arrival between the two signals, so if $B(d_1-d_2)/c$ is small, then fading is flat
- $(d_1-d_2)/c$, or maximum delay between first and last significant paths, is called the delay spread

Effect of Motion on Wireless Communication

- Suppose the transmitter antenna is stationary, while the receiver antenna is mounted on a moving vehicle
- Rx antenna starts out at distance d and moves at a velocity of v with respect to tx

- Pure sinusoid is transmitted at carrier frequency f_c

(Fig. 13)

- Signal is now

(Eq. 19)

- collecting terms

(Eq. 20)

- The frequency $f_c v/c$ is called the Doppler frequency – arises from the doppler effect
- Causes periodic dropouts in signal strength if compensation is not used in the receiver

(Eq. 21)

Example. $f_c = 1$ GHz, v = 108 km/hr

- Causes periodic dropouts in signal strength in multipath
- Say two paths exist, one stationary and the other in motion

(Fig. 14)

- Received signal is now

(Eq. 22)

Example. f_c = 1 GHz; one path stationary: d_1 = 100m; second path moving: d_2 = 100m initially, v = 54 km/hr

Fast vs. Slow Fading

- Combining multipath with doppler, we have a signal that changes with time
- How quickly does the signal change?
- Best to consider this from the perspective of a packet

(Fig. 15)

- If the signal changes much more quickly than the packet length, this is called "fast fading"
- If the packet length is much longer than signal changes, this is called "slow fading"
- Generally, fast fading is better – getting stuck with a low signal strength is bad, but in fast fading you will probably get a good signal on average
- Unfortunately most of the fading in the world is slow

Link Budgeting

- How much power do you need to be reasonably assured of good communication? – make a link budget.
- Aside: Decibels

- Gains and losses in wireless communication are multiplicative, e.g. amplifier gain, fading
- Multiplicative gains are hard to deal with intuitively … but … if the gains are ABCD, they can be made additive by taking the log: log(ABCD) = log A + log B + log C + log D
- You can express a quantity x in decibels by taking $10\log_{10}x$
- Examples: x=1=0dB; x=10=10dB; x=100=20dB; x=1000=30dB; …
- Other fun stuff to know: x=2≈3dB
- Adding in dB is equivalent to multiplying in normal domain; subtracting in dB is equivalent to dividing in normal domain
- E.g. 20=2x10=3dB+10dB=13dB; 500=1000/2=30dB-3dB=27dB
- dBm = dB referenced to 1 mW : e.g., 30 dBm = 1000 mW = 1 W.
- Link budgets are usually expressed as POWER (not amplitude) and in terms of dB.
- Take the starting power (at the transmitter), add all the gains, and subtract all the losses

(Eq. 23)

- P_T: Transmitter power; G_T: Transmitter antenna gain; G_R: Receiver antenna gain; L_P: Path loss; L_F: Fading margin; L_O:Other losses; P_R:Receiver power (all in dB)

Link Budgeting

- How much power do you need to be reasonably assured of good communication? – make a link budget.
- Link budgets are usually expressed as POWER (not amplitude) and in terms of dB.

- Take the starting power (at the transmitter), add all the gains, and subtract all the losses

(Eq. 24) $P_R = P_T + G_T + G_R - L_P - L_F - L_O$

- P_T: Transmitter power; G_T: Transmitter antenna gain; G_R: Receiver antenna gain; L_P: Path loss; L_F: Fading margin; L_O:Other losses; P_R:Receiver power (all in dB)

Path Loss

- Path loss term is L_P
- Before we said that the path loss is proportional to d^a, where a is the path loss exponent
- In dB, we have

(Eq. 25) $L_P = 10 \log_{10} kd^a = 10 \log_{10} k + 10a \log_{10} d$

Fading margin

- this amount is allocated to ensure a high probability that fading will not disrupt the signal
- Can use the probability of various amplitudes in Rayleigh fading to obtain an adequate margin

Example. You need -10dBm of power to ensure reliable communication. Required range is 500 m, with a path loss exponent of 3 (ignore the constant of proportionality). The transmit antenna has G_T=3 dB, and the receive antenna is isotropic. Allow 10 dB for the fade margin, and 0 dB for other losses. What is the required power at the transmitter in dBm?

(Answer: $P_R = P_T + G_T + G_R - L_P - L_F - L_O$ … solve for P_T. The only tricky part is that G_R = 0 dB because the antenna is isotropic.)

Data Link Layer: Access Control and Multiple Access

- The role of the data link layer is to ensure reliable communication between two connected terminals, and to allocate access to a shared medium
- E.g., in wired communication, Ethernet is a data link protocol
- These protocols become very important in wireless communications, because everyone in the world is sharing the same medium (i.e., the air)
- Compared to wired systems, the random delays and frequency shifts inherent in wireless systems are problematic and require special attention in protocol design
- Also at this layer is error detection/correction

Fixed multiple access protocols

- There are two traditional multiple access protocols, which divide up the medium in terms of time or frequency – both ensure that the users do not interfere with each other
- Frequency division multiple access (FDMA): The entire frequency band is divided up (equally or not) among the users. In each user's allocated band, the user can do as s/he pleases.
- Guard bands between users are given to prevent doppler frequency shifts from causing interference
- Example. 1 MHz of bandwidth to be shared among 10 users, with a 10 kHz guard band between each user. So

each user gets 90 kHz (don't forget about half a guard band before the first user, and half a guard band after the last user).

- Time division multiple access (TDMA): Time is broken up into frames. Each frame is divided up (equally or not among the users. In each user's allocated time, the user can do as s/he pleases.
- Guard times between users are given to prevent random delays from causing interference
- Example. Frame duration of 10ms, guard time of 0.1ms, 10 users. So each user gets 0.9ms per frame.
- Problem with FDMA and TDMA: They require central control and make it difficult to quickly reuse resources … best for cellphone-like systems, where people need a full channel for a long period of time; bad for packet data systems, where data is sent less frequently

Exposed/hidden terminal problem

- Carrier sense multiple access (CSMA) solves this problem in Ethernet (i.e., sense whether the channel is free and then transmit) … this is decentralized and appropriate for packet data. Can we do the same thing here?
- Not exactly: there are two problems, the hidden terminal problem, and the exposed terminal problem

- Hidden terminal problem:
- There are three nodes: A, B, C
- Say B is in radio range of A and C. However, A and C are NOT in radio range of each other.
- A wants to send to B. A senses the medium and sees that it is clear, so A sends.
- While A is sending to B, C decides to send to B. C senses the medium – C can't tell that A is transmitting because A

and C are not in radio range. Thus, C decides that the medium is clear and transmits.
- Messages from A and C collide at B, corrupting each other.

- Exposed terminal problem:
- There are four nodes: A, B, C, and D.
- A is in range of B, B is in range of A and C, C is in range of B and D, and D is in range of C. (i.e., the nodes are arranged in a line as A B C D, and each node can only see its neighbors)
- B sends to A, and at the same time, C wants to send to D. C senses the medium and decides the medium is busy, so C does not send. However, B's transmission cannot reach D, so in fact it is safe for C to transmit, and that time/bandwidth is wasted.

(For the above, draw figures to help the students understand the relationships among the users)