Presentation 4

Basics of Wireless Networking

For over 25 years wireless networks are part of our day-to-day experience and the basis for mobile communication that has changed many patterns of behavior. Despite the simplicity of using mobile telephones and tablets, the technology behind mobile communication is more complex by several orders of magnitude. This chapter presents a simplified introduction to radio communication and the two principle forms of wireless mobile networking — Wi-Fi and cellular. We will pay particular attention to the common features of the two contexts — the local and global networks.

Slide 2 **Energy and Power**

Two essential concepts for understanding radio transmission are energy and power. **Energy** is defined as the ability to do work, and may be kinetic (energy of motion) or potential (stored energy). **Power** is the quantity of energy used or transferred per second. Power is measured in units of **Watts** and energy is measured in **Joules**, defined as Watts × seconds. Power purchased from the electric company is paid for in units of kilowatt-hours, where

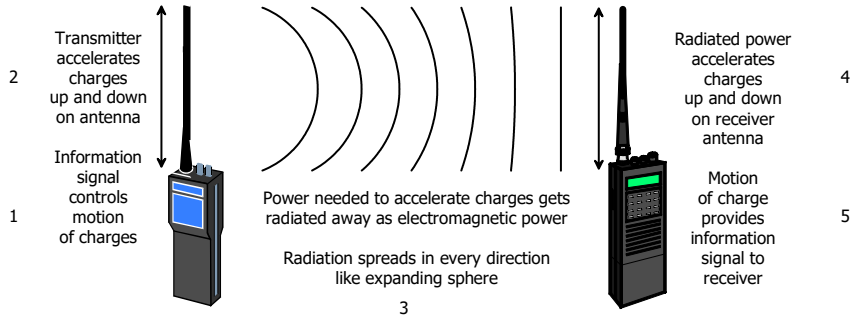
1 kW-hour = 1000 Watts × 3600 seconds/hour
= 3.6×10^6 Joules

Slides 3 – 4 **Electricity and Magnetism**

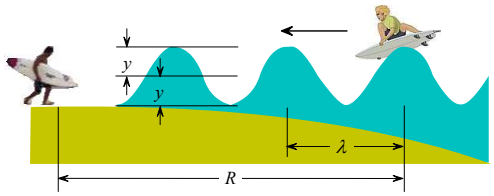
A charged object, such as an electron in motion may create an electric field **E** and a magnetic field **B**. These fields are vector quantities in 3 dimensions, and their magnitude and direction generally depend on the time t and location **x** at which they are measured. **Radiation** is the transfer of power from a charged object through its electromagnetic field. This radiation can perform work when it is absorbed by another object, the process underlying radio transmission. The radiation field associated with an electromagnetic field is proportional to the vector product **E** × **B** and this is the power emitted by a moving charged object.

A motionless charge produces an electric field with magnitude $E = -\frac{\text{charge}}{R^2}$, where R is the distance of the charge at the point of measurement, and magnetic field B = 0. Since **E** × **B** = **0** this system does not radiate.

In a radio, charges are accelerated back and forth along the length of the antenna. Electric charge in motion is called an **electric current**, and in an antenna the charges accelerate to produce the time-dependent current. Both **E** and **B** are non-zero, producing a radiation field that can be detected in a remote antenna.



Electromagnetic radiation satisfies a wave equation, whose solution shows that the radiated power depends on time t and distance R from antenna. Electromagnetic waves can be compared to **ocean waves** rolling onto a beach.



The wave height has peaks and troughs, and we take y to be the height of peak above the average water height, which is also the depth of trough below center.

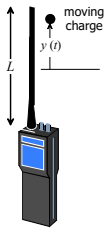
At a fixed distance from the shore, wave rises and falls over time — the period T is time between the arrivals of two wave peaks. The frequency f of wave arrivals is given by 1/T, the number of wave peaks per second.

At some fixed time, multiple wave peaks are present at various distances from the shore. The wavelength λ is the distance between two wave peaks.

A **surfer** can ride the peak of a wave, because the peak depends on the distance from the shore and the time. The speed of the moving peak is given by f × λ.

By analogy, a **charge on a radio antenna** accelerated up and down oscillates from top to bottom over a distance L every T seconds.

The frequency of the oscillation is f = 1/T so that the position y(t) of the charge at time t can be written



$$y(t) = \frac{L}{2}\cos\left(2\pi\frac{t}{T}\right) = \frac{L}{2}\cos(2\pi ft)$$

The properties of electromagnetic field are given by Maxwell's equations, which leads to a wave equation. The solution to the electromagnetic field of a straight antenna is

$$\mathbf{E}(R,t) = \frac{\mathbf{E}_0\cos\left[2\pi f\left(R/c - t\right)\right]}{R} \qquad \mathbf{B}(R,t) = \frac{\mathbf{B}_0\cos\left[2\pi f\left(R/c - t\right)\right]}{R}$$

R = distance from antenna to point of measurement

t = time (measured on some clock)

f = frequency $= 1/T$

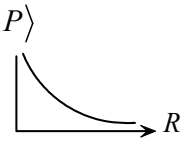
$\mathbf{E}_0, \mathbf{B}_0$ are physical constants

c = speed of light

The radiated power vector and the average power (averaged over one period) are

$$\mathbf{P} = \mathbf{E}\times\mathbf{B} = \frac{\mathbf{E}_0\times\mathbf{B}_0}{R^2}\cos^2\left[2\pi f\left(R/c - t\right)\right] \qquad \langle P\rangle = \frac{1}{T}\int_0^T |P(t)|\,dt = \frac{|\mathbf{E}_0\times\mathbf{B}_0|}{2}\frac{1}{R^2}$$

Because the power is proportional to $1/R^2$, the signal strength **fades** — it decreases with distance from the source antenna.



The electromagnetic wavelength is defined as $\lambda = cT = c / f$, so that $\lambda f = c$, the speed of light. Using this definition, the electric field can be written

$$\mathbf{E}(R,t) = \frac{\mathbf{E}_0 \cos\left[2\pi f\left(R/c - t\right)\right]}{R} = \frac{\mathbf{E}_0 \cos\left[2\pi\left(\frac{fR}{c} - ft\right)\right]}{R} = \frac{\mathbf{E}_0 \cos\left[2\pi\left(\frac{R}{\lambda} - \frac{t}{T}\right)\right]}{R}$$

and the wave peaks are found when $\cos\left[2\pi\left(\frac{R}{\lambda} - \frac{t}{T}\right)\right] = 1$, which is satisfied when

$\left(\frac{R}{\lambda} - \frac{t}{T}\right) = 0, 1, 2, \ldots$

Therefore shifting in time and distance from one wave peak to the next requires $\left(\frac{\Delta R}{\lambda} - \frac{\Delta t}{T}\right) = 0$.

The speed of the wave is the change in distance divided by the change in time from one peak to the next, so that $v = \frac{\Delta R}{\Delta t} = \frac{\lambda}{T} = \lambda f = c$, which shows that the **wave peaks move at the speed of light**.

Looking at the electric field at some fixed distance R_{const} it takes the form

$$\mathbf{E}(R_{const}, t) = \frac{\mathbf{E}_0}{R_{const}} \cos\left[2\pi\left(\frac{R_{const}}{\lambda} - \frac{t}{T}\right)\right] = C_1 \times \cos\left(2\pi ft - C_2\right)$$

where C_1 and C_2 are constants, which rises and falls over time.

Looking at the electric field at some fixed time t_{const} it takes the form

$$\mathbf{E}(R, t_{const}) = \frac{\mathbf{E}_0}{R} \cos\left[2\pi\left(\frac{R}{\lambda} - \frac{t_{const}}{T}\right)\right] = \frac{C_1}{R} \cos\left[2\pi\left(\frac{R}{\lambda} - C_2\right)\right]$$

where C_1 and C_2 are constants, which rises and falls, but also fades over distance R.

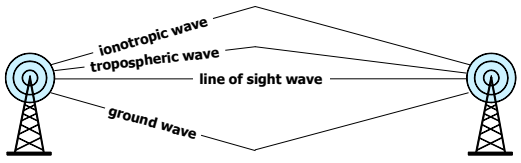
Radio antennas are most effective in the frequency range from 30 kHz (wavelength λ = 10 km) to 300 GHz (wavelength λ = 1 mm). As shown on slide 10, chemical reactions generate higher frequencies: infra-red (IR), visible light, ultra-violet (UV), X-rays (Roentgen) and nuclear reactions generate gamma rays.

A line antenna is most efficient when $L = \lambda / 2$. GSM phones operate at a frequency of about 1 GHz, which is equivalent to a wavelength of $(3\times10^{10}$ cm/sec)/(10^9 Hz) = 30 cm. This is convenient for a phone about 15 cm in size.

Slides 12 – 17 Radio Wave Propagation

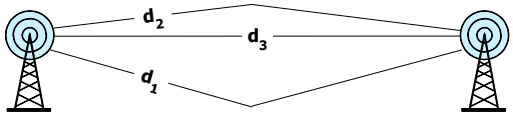
A radio transmitter generates radio waves, which propagate (spread out) through space. Part of the radiated power is generally obstructed or reflected, and another part is detected by receiver.

The radio transmission that propagates directly to the receiver is called the **line of sight wave**. The **ground wave** is a radio signal that reflects and propagates along the ground. Other portions of the transmission **refract** off different layers of the atmosphere, the lower troposphere and higher ionosphere.



The longer the propagation path, the longer the delay of the arriving signal.

In addition to refraction in the atmosphere (bounding off the sky), the propagated wave may be reflected back to the transmitter (as with a mirror) or absorbed in trees, walls, and buildings. Absorption weakens the line of sight wave, making it less effective. Refraction causes the transmitted signal to be received in multiple copies separated by various time delays, because a longer transmission path leads to a longer transmission time. These multiple delay times, compared to the arrival of the line of sight wave, cause a phenomenon called **multipath fading**.



At any time t, the signals arriving at the antenna on the right have traveled over three paths of lengths $d_1 > d_2 > d_3$. The **line of sight wave** travels over the shortest distance d_3 and was the last to be transmitted. The refracted **atmospheric wave** traveled the longer distance d_2 and so it must have been transmitted earlier to arrive together with the line of sight wave. The **ground wave** travels the even longer distance d_3 and so it must have been transmitted even earlier than the atmospheric wave to arrive at the same time as the other two. Since the atmospheric wave and ground wave were transmitted at different times than the line of sight wave, they contain different information and can interfere with the main information stream.

Interference can also cancel the line of sight wave. Slide 15 shows how two waves on a string can propagate in opposite directions and be detected at opposite ends of the string. But in the very center of the string, the two pulses exactly cancel each other — at that precise point no pulses will be detected. This is an example of how moving a transmitter or receiver slightly can change reception quality.

A more detailed demonstration is given on slides 16 and 17. Two waves arrive at a receiving antenna from two slightly separated sources (R,t) and (R+ΔR,t+Δt). Making various approximations and using trigonometric identities, the electric field becomes

$$\mathbf{E} = \frac{\mathbf{E}_0}{R}\left\{2\cos\left[2\pi\left(\frac{R}{\lambda} - ft\right)\right] \times \cos\left[\pi\left(\frac{\Delta R}{\lambda} - f\Delta t\right)\right]\right\}$$

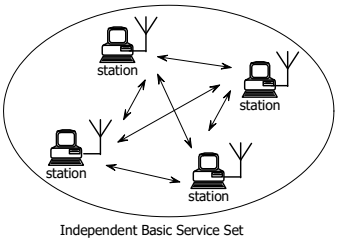
which has the form of a single-source electric field multiplied by a factor depending on the separation of the two sources. In a completely transparent medium (such as empty space) the argument of the second cosine is 0 and the cosine is 1. But under various circumstances it can occur that the cosine is zero and the electric field of the transmission vanishes. This can be a serious problem in WiFi, and for this reason most modern access points have a double or triple antenna that can reduce the multipath fading.

Slides 19 – 22 IEEE 802.11 Protocol Layers

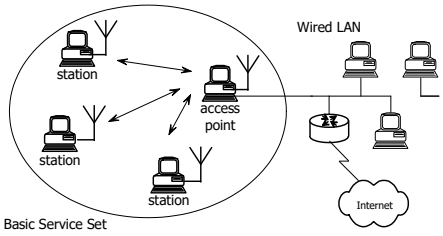
IEEE 802.11 is a large collection of standards for wireless local area networks. The main standard is defined at the physical and data link layers. Within the data link layer, the MAC sublayer defines required and optional features of a wireless LAN, and IEEE 802.2 is used for optional logical link control (common to Ethernet and other 802 standards).

WiFi is a trademark of Wi-Fi Alliance trade association and describes a collection of specifications for mass market, consumer wireless LANs based on 802.11 standards. WiFi can be configured in 3 basic operating modes.

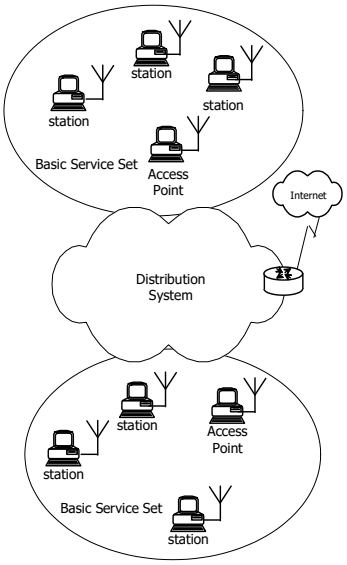
In WiFi **ad hoc mode**, an Independent Basic Service Set (IBSS) is any set of 802.11 STAs (wireless stations) that transmit and receive on the same frequency. This mode provides peer-to-peer service with no access point and no connection to an external network. The STAs communicate directly with one another, a useful service for quick set up, such as emergency response in a remote location. No authentication or registration is required to join the IBSS. Multiple IBSSs are completely independent, with no bridging or hand-off between them.



In WiFi **infrastructure mode**, a Basic Service Set (BSS) is a set of wireless end stations (STA) and an Access Point (AP) that is connected to external (usually wired) network infrastructure. The AP acts as a base station for the wireless network — all traffic flows through the AP with access permission decided by contention (competition) or polling (the AP grants access by some schedule). To join the BSS, stations must first associate with the AP through user authentication and registration. This is the most common configuration for home WiFi systems.



In WiFi **extended infrastructure mode**, an Extended Service Set (ESS) is two or more BSSs interconnected to form a single subnetwork (broadcast domain) that looks like one large BSS to the LLC sublayer. There is one Access Point (AP) in each BSS, and the BSSs are connected with one another via a Distribution System (DS). The DS is the backbone network, performing MAC-level transport of MAC SDUs, but its implementation details are left general (they are not specified in 802.11). The **portal** is a software gateway function in the AP that bridges the BSS to any non-802.11 DS protocol. The DS services permit handoff (a station moving from one BSS to another) by providing coordination between APs.



Slides 23 – 27 WiFi Medium Access

Medium access in WiFi is affected by connectivity problems not seen in cable systems.

The **hidden node problem** describes four wireless stations A, B, C, D in which A begins transmitting to B. C is too far from A to notice its transmission and starts transmitting to D. But C is in range of B so that the C to D transmission interferes with the A to B transmission.

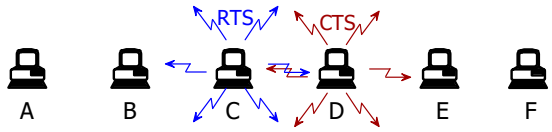
The **exposed node problem** describes four wireless stations A, B, C, D in which B begins transmitting to A. C is in range of B, notices its transmission, and waits before transmitting to D. But D and B are out of range and A and C are out of range, so this wait is inefficient.

WiFi specified two medium access methods.

In **CSMA with Collision Avoidance (CSMA/CA)**, stations listen for transmissions and do not transmit if any carrier signal is detected. This method is exposed to the hidden node problem because no collision detection is possible (the antenna cannot receive while the transmitter active). To avoid collisions, the stations practice non-persistent access with random backoff, which means that when they see an opportunity to transmit they wait a random time interval before beginning.

Multiple Access with Collision Avoidance (MACA) is a connection-oriented technique with additional overhead but better throughput in hidden and exposed nodes. Before data transmission begins a channel is set-up. The source station initiates a Request To Send (RTS) message containing a Net Allocation Vector (NAV) that specifies the predicted data transmission time. The destination station responds with a Clear To Send (CTS) message that echoes the NAV. Following error-free transmission of the data, the destination sends an ACK message.

With 6 stations



C transmits to D — first C sends RTS and then D returns CTS. At the same time B has data for A and E has data for F. The **exposed node problem** (B to A) is solved: B hears the RTS but not the CTS, so it knows that C's destination is out of range (and no interference will occur if it transmits). The **hidden node problem** (E to F) is solved: F hears the CTS but not the RTS, so it knows that the sender to D is out of range but D is in range, so it must wait before transmitting.

Slides 28 – 37 MAC Sublayer Frame Structure

The MAC layer frame in WiFi consists of the header, data, and frame check sequence (FCS).

Frame Control	Duration/ ID	Address 1	Address 2	Address 3	Sequence Control	Address 4	Frame Body	FCS
2 bytes	2 bytes	6 bytes	6 bytes	6 bytes	2 bytes	6 bytes	0-2312 bytes	4 bytes

The header fields are:

Frame Control	Control flags (see slide 29)
Duration/ID	Timing control (NAV)
Addresses	Various MAC entities whose address is required for end-to-end transmission
Sequence Control	Sequence/Fragment number for error/flow control

The MAC header contains up to 4 address fields describing up to 5 MAC entities:

1. BSS Identification Number (BSSID) MAC address of AP (or random in IBSS)
2. Source Address (SA) Station that initiated the message
3. Destination Address (DA) Final destination for the message
4. Transmitting Station Address (TA) Source station for the message on this hop
5. Receiving Station Address (RA) Destination station for the message on this hop

These MAC entity addresses are used in address fields 1 to 4 depending on the transmission scenario (to DS and from DS are flags in the frame control field of the MAC header):

	To DS	From DS	Address 1	Address 2	Address 3	Address 4
No AP in use	0	0	DA	SA	BSSID	—
From STA to AP	0	1	DA	BSSID	SA	—
From AP to STA	1	0	BSSID	SA	DA	—
From AP to AP	1	1	RA	TA	DA	SA

In all cases, the address fields refer to:

- Address 1 Immediate destination address
- Address 2 Immediate source address
- Address 3 Final destination or source when DS performs distribution
- Address 4 Source address for DS to DS messages (802.11 is also DS)

Slide 32 shows address in an IBSS with no AP in use. The to/from DS bits are 0, the source and destination addresses are stations, and the BSSID is included for error protection.

Slide 33 shows addressing in a BSS. When the AP sends data to a station, the immediate destination is the final destination, the immediate source is the AP (with MAC address BSSID) and the original source was the station with address SA. When the AP receives data from a station, the immediate destination is the AP (with MAC address BSSID), the immediate source is the original source with address SA, and the final destination is the station with address DA. Slide 34 shows an example of a frame sent from station A to station B via the AP in two hops.

Slide 35 shows control and management frames in a BSS. These are sent with both DS flags set to 0, but the AP is either source or destination (and included in address 3 for error protection).

Slides 36 and 37 show addressing in an ESS. As in a BSS, all transmissions are sent through an AP, and the same address fields are used in the same way. On the first hop, the addressing is "To AP 1 from Station A for Station B" and on the last hop, the addressing is "To Station B from AP 2 from Station A". WiFi does not describe how the frames are transmitted from AP 1 to AP 2.

Slides 38 – 39 Mobility Management

As seen in Mobile IP (slides 18 – 30 of presentation 3) **mobility management** has certain features common to all infrastructures:

Registration and **authentication** of user credentials is performed when a mobile station (MS) is activated in a given service area. In WiFi, the access point (AP) manages registration by requiring a password. In the cellular system (GSM / 3G / 4G), databases are maintained on individual users. The Home Location Register (**HLR**) maintains account and location information for home customers (users with a contractual relationship to the server provider). A Visitor Location Register (**VLR**) is a cache of HLR data on active roamers in any particular service area, allowing the system to locate the user at any given attachment point.

Call establishment is performed when a user initiates or receives a call in a service area (in the home service area or as a roamer).

Security (encryption) is intended to protects users from fraud and eavesdropping.

Handoff (handover) is performed when the MS changes its attachment point to the network. In WiFi, this means switching AP. In cellular, this is the complex process of moving from cell to cell.

In either case the process can be performed in one of two ways. In **hard handoff**, the old cell transfers control to the new cell in a break-before-make sequence. The transceiver (transmitter/receiver) in the old cell stops transmitting to the user and then the transceiver in new cell begins transmitting to the user. In older systems this cause many calls to drop. In **soft handoff**, the central transceiver (primary connected base station) coordinates with the nearest APs or cells and determines which transmitter is receiving the strongest signal from the user. In a make-before-break sequence, the transceiver in the old cell continues transmitting to the user while the transceiver in new cell begins transmitting to user. When the transceiver in the new cell is safely connected user, the old cell stops transmitting. Newer systems use soft handoff to prevent call drops and other inefficiencies.

Slide 41 1970 — 0G Mobile Phone System (MPS)

The first mobile phone system (zeroth generation) deployed one central transceiver (base station) and mobile telephones that could communicate by radio. The system transmitted at high power for maximum distance, covering 65 to 80 km.

The radio modulation scheme was standard analog FM, supporting 12 simultaneous mobile phone calls. If all 12 channels were busy, other calls were blocked. This required 24 carrier frequencies, with 2 frequencies per phone: one dedicated transmit frequency and a separate dedicated receive frequency.

Slides 42 – 46 Cellular Concept

The essential idea in cellular telephony is to divide the coverage area into small regions called cells. In each cell, a central cell transceiver serves all users in the cell. All mobile stations communicate via the base cell transceiver, so that each active device must be allocated a frequency pair: a **downstream** frequency to receive from the base station and an **upstream** frequency for transmitting to the base station.

The cellular system can transmit at relatively low power, just enough power to cover one cell. Therefore the system can **reuse** the same set of frequencies in many cells with no interference between cells.

Slide 43 shows reuse patterns in cellular systems. In the 7 cell reuse pattern, cells are designated A, B, C, ... , G. The full set of radio channels available for the system are divided into 7 sets and each cell type uses only frequencies of its type. The pattern guarantees that no A cell borders on another A cell, which increases the distance (and reduces the interference) between A cells (using the channels allocated to cells designated A).

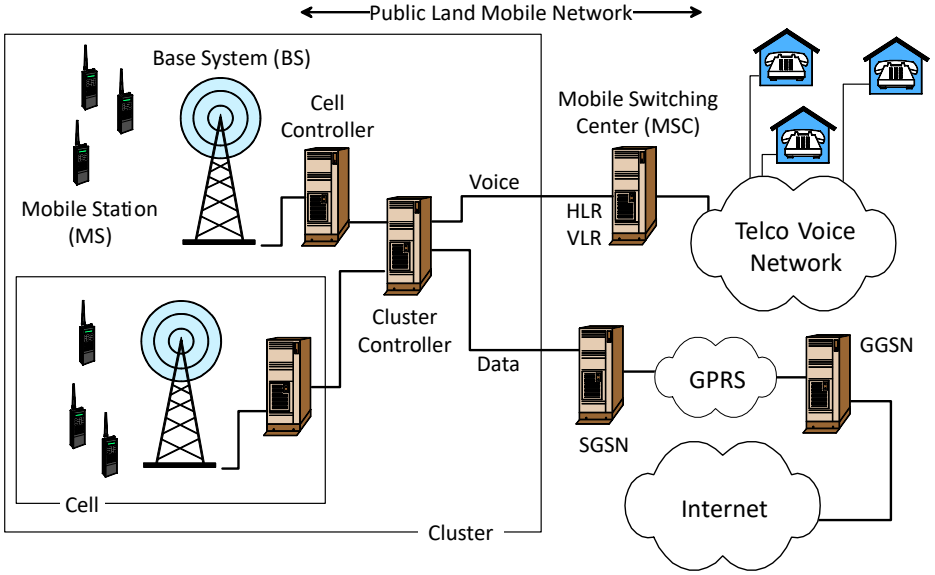
As shown on slide 44, a mobile service provider operates in some set of service areas (in some parts of the world — including the USA — large geographical areas have no cellular coverage). A service area is divided into clusters — each cluster is based on the reuse pattern, so that all available channels are used in each cluster. The cluster is divided into cells.

The generations of cellular networks are shown in slide 45.

- The 1G networks were proprietary systems operated by service providers. These systems used analog radio transmission and could not operate with other systems.
- The 2G networks in the US were upgrades of proprietary 1G systems, providing service to more uses but still no interoperability.

- The 2G systems in Europe were based on standards designed to permit a user to use the same device anywhere in the EU. The GSM system became the basis for most cellular network architectures since the 1990s, using time division multiplex to combine 8 digitized voice channels per frequency band on a frequency division multiplex of the analog radio capacity. CDMA is a more efficient multiplex system in which each user is assigned a different encoding scheme. Both GSM and CDMA provided voice service over traditional circuit switched digital channels. Separate from the voice channel is the GPRS packet switched network that enabled IP-based data to be exchanged with the Internet. In 2G systems, each individual cell was connected to a centralized switching center that connects the user device to the non-mobile telephone network and controls handover.
- Following the success of GSM, 3G systems in most of the world implemented UMTS (a GSM enhancement) or WCDMA (a CDMA enhancement). In addition to higher transmission speeds and higher reliability (fewer dropped calls), 3G systems improved handover when a mobile device moved from one cell to another. Unlike 2G, individual cells were organized into lower level groups that could provide handover without intervention by the centralized switching center.
- 4G systems were defined to satisfy Next Generation Network requirements. LTE and WiMAX are called 4G systems, and although they provide higher speeds, they do not meet the NGN requirements. In 4G networks, the original circuit switched network for voice calls is eliminated. Voice calls are transmitted in packet mode as Voice over IP (VoIP).
- 5G is defined in two stages. The initial implementation is a radio enhancement to 4G LTE. The full design provides for highly abstract Software Defined Networks (SDN).

Slide 46 shows the general architecture of a 2G cellular network, which became the basis for understanding the network functions is all subsequent generations.



- A cell (base system) consists of a radio antenna and a controller called a Base Transmitter System (BTS).
- The BTS of each cell connects to a cluster controller called a Base Controller System (BSC).
- The BTS separates digitized voice traffic from data traffic.

- Voice calls are sent on circuit mode channels to a Mobile Switching Center (MSC) for connection to the PSTN, which interconnects cellular service areas.
- Data is sent to a Serving GPRS Support Node (SGSN), a packet equivalent of the MSC, for connection to a Gateway GPRS Support Node (GGSN) for connection to the general Internet.

Slides 47 – 50 GSM Registration Process

The registration process is similar in GSM and UTMS (3G):

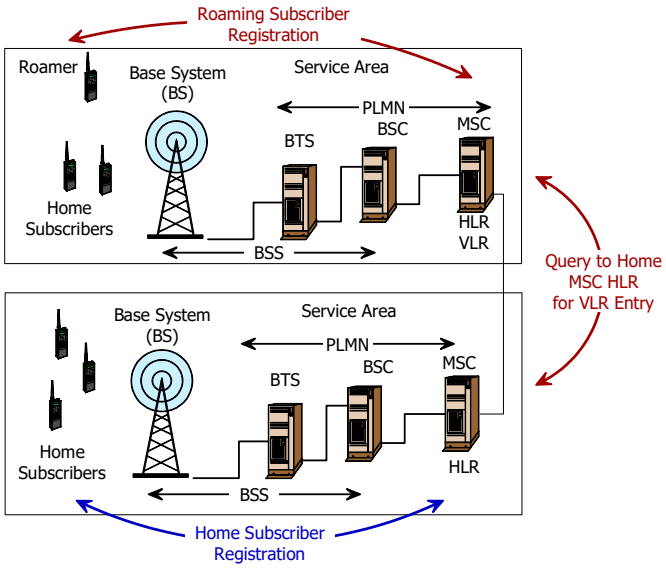
- An MS enters a service area and establishes a low bit-rate control channel with the service provider.
- The MS requests service and the BTS allocates a frequency pair (upstream and downstream) for the registration process.
- The MS reports its location, status, and identity to the Mobile Switching Center (MSC). The user ID consists of a dedicated hardware ID code in the phone and an ID from the **Subscriber Identity Module (SIM)** card identifying the customer. The MS generates an encrypted access code for the network, called a Temporary Mobile Subscriber Identity (TMSI).
- The Mobile Switching Center (MSC) authenticates the customer identity with the HLR (or for roaming subscriber, creates a VLR entry), updates the HLR and billing database.
- The MS releases the frequency pair and maintains contact with the BTS over the low bit-rate control channel.

The registration process is summarized in slide 48.

The process of **call establishment** is shown in slide 49 (for a call initiated by the mobile station) and slide 50 (for a call incoming to the mobile station).

Slide 51 Mobility Elements in GSM

Mobility management in GSM conforms to the pattern shown below.



The user operating in the **home service area** uses a control channel to register with a base station in a cell. The cell consists of a base transceiver (BS), a Base Transmitter System (BTS), the controller for the cell. The cell BTS forwards the registration request to the cluster controller called the Base Controller System (BSC). Together, the BTS (cell) and BSC (cluster) are called the Base Station Subsystem (BSS). The cluster forwards the registration request to the Mobile Switching Center (MSC), which contains the HLR. The MSC accepts the request, registering the user in the cell.

If the **user roams** to the service area of another provider, the user must submit a new registration request. This time, the request is passed to the MSC and from there to the MSC of the home service provider. If the home service provider permits roaming for the user, elements of the user's record are sent from the HLR in the home system to the VLR in the non-home system. The home HLR registers the location of the user so that incoming calls can be routed.

Slides 52 – 53 Handover Types

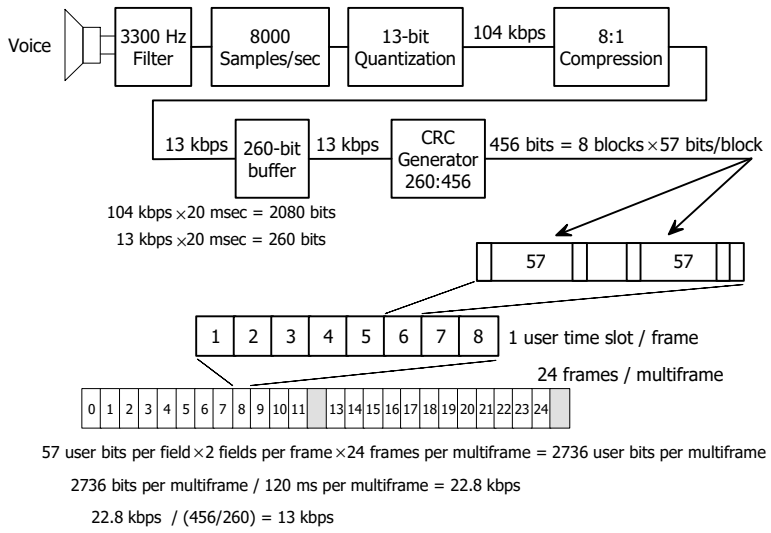
A GSM channel is defined by a frequency pair (upstream and downstream) used by the MS to communicate with the BTS. GSM defines four types of handover:

Intra-cell	The MS changes frequencies to avoid interference.
Inter-cell / Intra-cluster	The MS moves between cells within the control of one BSC (cluster).
Inter-cluster / Intra-MSC	The MS moves between cells controlled by different BSCs (clusters) and the MSC controls the handover.
Inter MSC	The MS moves between cells controlled by different MSCs

The GSM inter-BSC handover procedure is shown in slide 52. The MS monitors the received signal power from neighboring cells and the cluster determines that a handover is necessary. The request is forwarded to the MSC, which requests the handover from a new cluster. A cell in the new cluster activates a channel, passing the ACK of the request back to the new cluster, to the MSC to the old cluster, the old cell, and the MS. The MS begins to operate on the new channel, and confirmation of completion is sent to the MSC, old cluster and old cell.

Slide 54 GSM Voice Transmission Summary

The digital transmission of voice in GSM is as follows:



- The voice is band-limited to an analog range of 300 to 3300 Hz and sampled at a rate of 8000 samples per second (as in a Bezeq phone).
- The samples are rounded off to a 13-bit number (8192 levels instead of 256 levels on a Bezeq phone). 8000 samples/second × 13 bits/sample = 104 kbps.
- The digital stream is compressed at a ratio of 8:1, producing a 13 kbps stream.
- The bits are fed into a 260-bit buffer (one 20 msec sample). For each 260 bits a CRC error correction code of 196 bits is added, producing 456 bits. The transmission rate is now 456 bits / 0.020 seconds = 22.8 kbps.
- Each 456-bit block is divided into 8 blocks of 57 bits. A user time slot is constructed which includes control information and a pair of 57-bit blocks.
- Time slots from 8 users are combined into a frame.
- The GSM system transmits multiframes with 26 frame slots. Frame 12 and frame 25 are not used, so that the multiframe contains 24 frames, which contain 192 user time slots, which contain 2736 bits from any given user.

Slides 55 – 56 GSM Protocol Stack

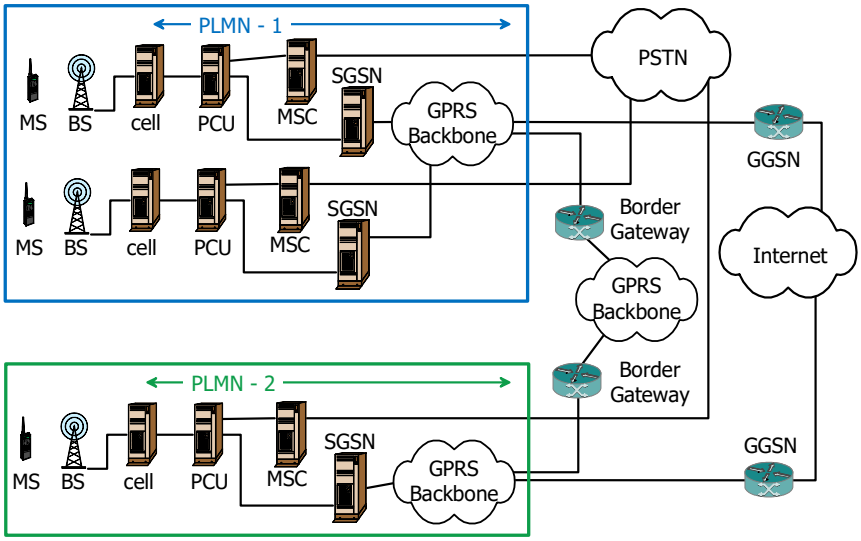
As seen on slide 55, the GSM protocol stack is quite complex. The top layers operate between the MS and the MSC for connection and mobility control. The BSC (cluster) and MSC run three protocols (BSSAP, SCCP, and MTP) to manage telephone call signaling.

The RMM protocol handles channel (frequency pair) allocation to the MS from the BSC. The SCCP signaling control connection between the MSC and BSC is converted to a BTSM signaling control connection between the BSC and BTS, and then to another RMM instance between the BTS and MS. The MTP that handles standard telephone signaling (ring indication, buys signal and so on) between the MSC and BSC (under control of the SCCP connection) is converted to LAPD, a standard signaling protocol for digital telephones.

The taxonomy of channel structures (slide 55) multiplexed over the GSM multiframes is similarly complex.

Slides 57 – 61 General Packet Radio Service (GPRS)

GPRS provides packet mode data access for over the GSM and 3G networks using an IP-based architecture. IP datagrams are separated from circuit mode voice traffic at the cluster by adding new functions to the BSC cluster controllers.



The **Packet Control Unit (PCU)** provides a packet mode function in the BSC to handle IP datagrams. Circuit mode voice/data are routed to MSC for forwarding to another MSC or PSTN (public switched telephone network, such as Bezeq). Packet mode data is routed to a **Serving GPRS Support Node (SGSN)** for forwarding to the Internet. In GSM the PCU to SGSN connection runs IP over Frame Relay, and in 3G it runs IP over ATM.

Mobility management for circuit mode voice traffic uses standard PSTN / PLMN routing, but packet mode traffic uses IP routing. The IP routing requires two new support nodes, SGSN and GGSN.

The **Serving GPRS Support Node (SGSN)** can be seen as a packet-switched version of the MSC. It handles packets to and from the MS and manages MS mobility, keeping track of the MS location and status in the mobile system. The SGSNs operate over a GPRS backbone jointly operated by cellular operators. MS to MS IP traffic can move over the GPRS backbone without exiting to the Internet backbone via a GGSN.

The **Gateway GPRS Support Node (GGSN)** provides an interface SGSNs to external IP networks, maintaining routing information for the SGSN currently serving a user. It is the exterior gateway for GPRS network. It runs DHCP to assign IP addresses to the MS and routes incoming IP datagrams to appropriate PCU. Traffic that leaves the GGSN to the Internet is pure IP traffic with no information about MS location or status. That information must be reattached to replies from the Internet entering the GGSN, so that the SGSN can route data to the appropriate PCU.

The **GPRS connection process** is shown in slide 61. The MS switches on and sends a GPRS attach request. User Registration (1) is required to associate a PLMN address (IMSI) with a Packet Data Protocol (PDP) address (a static or dynamic IP address). The BSC queries the HLR (2) for user authentication, and the HLR updates a VLR (3). Call Admission Control (CAC) determines the required network resources and grants resources if available (4). IP datagram delivery is now performed by hop-by-hop routing using routing tables in the GSN (GGSN or SGSN), by way of the

address conversion and VC mapping of the tunnel protocols. The GSN also handles compression and encryption.

Another 2.75G enhancement was **High Speed Circuit Switched Data (HSCSD)**, which permitted data-only switched circuit channels. Standard GPRS is a packet switched (packet mode) service, which only requires capacity when data is being transmitted (and so **billing is based on download volume**). However, HSCSD is a circuit mode service, like a voice call, and requires dedicated system capacity for the entire call (and so **billing is based on air time**).

3G GRPS now uses **High Speed Downlink Packet Access (HSDPA)**, a fast packet mode service permitting download rates up to 337 Mbps. Slide 72 lists a few methods used in HSDPA in the physical layer to achieve these rates.

Slides 64 – 66 SMS in GSM Architecture

Short Message Service (SMS) is a text-oriented service introduced in 2G cellular. Text is transferred in the control frames of regular messages, so that they require almost no capacity from the network operator.

The SMS system adds 3 new nodes to the GSM system. A **Short Messaging Entity (SME)** is any entity that can receive or send SMS, whether a fixed network element, an MS, or another service center. The **Short Message Service Center (SMSC)** stores and forwards SMS between any SME and an MS. Within the SMSC, the **Gateway Message Service Center (GMSC)** receives an SMS, interrogates the HLR for routing information, and delivers the SMS to the MSC for the destination SME. The **Interworking Message Service Center (IWMSC)** receives an SMS from the MSC and delivers it to a different SMSC for forwarding if necessary.

The stages of delivery of an SMS are shown in slide 66.

Slides 67 – 75 CDMA

Code Division Multiple Access (CDMA) is a commercial system originally developed by Qualcomm. The initial version (now called CDMAone) operates on 1G/2G frequencies and is not compatible with 3G CDMA 2000, which is a collection of incremental enhancements.

CDMA (and CDMA 2000) allocate a 25 MHz radio band in each direction (upstream and downstream). This band is band divided into 1.25 MHz RF channels allowing 20 RF channels per cluster. The system uses DSSS digital transmission (see slide 13 of presentation 1), transmitting 1.2288 Mcps (million chips per second) in each 1.25 MHz radio channel. The chips are used to produce a **code division multiplex** in which each user has a different code (chip sequences for the transmission of 1 bit). All users can transmit simultaneously using independent chip sequences — orthogonal (Walsh) codes or pseudorandom noise (PN) codes that can be unambiguously separated at the receiver.

Orthogonal codes are formed in an m-dimensional vector space with the inner product

$$\mathbf{U} \cdot \mathbf{V} = \frac{1}{m} \sum_{i=1}^{m} U_i \times V_i$$

The **m orthonormal basis vectors** for the vector space are written S_i for i = 1, 2, 3, ... , m and satisfy

$$S_i \cdot S_j = m \times \delta_{ij} = \begin{cases} 0, i \neq j \\ m, i = j \end{cases}$$

Any vector **T** in the vector space can be written in the form

$$\mathbf{T} = \sum_{i=1}^{m} t_i \times S_i$$

where the coefficients t_i can be found in the usual way from

$$t_i = S_i \cdot \mathbf{T} = S_i \cdot \frac{1}{m} \sum_{j=1}^{m} t_j \times S_j = \frac{1}{m} \sum_{j=1}^{m} t_j \times \left(S_i \cdot S_j \right) = \frac{1}{m} \sum_{j=1}^{m} t_j \times m\delta_{ij} = t_i$$

The **coding scheme** for the forward channel is:

- The basis vector S_i is assigned as a chip sequence to station i, for i = 1, 2, 3, ... , m.
- The base station chooses the coefficient $t_i = \begin{cases} -1 \text{ , transmits bit 0 to station i} \\ 0 \text{ , no data for station i} \\ +1 \text{ , transmits bit 1 to station i} \end{cases}$ for each user station i.
- The receiver in each MS receives the vector $\mathbf{T} = \sum_{i=1}^{m} t_i \times S_i$
- Station i finds the data bit from $t_i = S_i \cdot \mathbf{T}$ using its assigned code S_i.

Slide 69 shows an example of 4 stations with basis vectors (codes)

$$S_1 = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \quad S_2 = \begin{bmatrix} -1 \\ +1 \\ +1 \\ -1 \end{bmatrix} \quad S_3 = \begin{bmatrix} -1 \\ -1 \\ +1 \\ +1 \end{bmatrix} \quad S_4 = \begin{bmatrix} -1 \\ +1 \\ -1 \\ +1 \end{bmatrix}$$

so that

$$S_2 \cdot S_2 = \frac{1}{4}\left((-1)^2 + (+1)^2 + (+1)^2 + (-1)^2 \right) = 1$$

$$S_1 \cdot S_2 = \frac{1}{4}\left((-1)(-1) + (-1)(+1) + (-1)(+1) + (-1)(-1) \right) = 0$$

Slide 70 shows a transmission example for which Stations 1, 2, and 4 each receive the two-bit sequence 0 then 1, while Station 3 receives no data then 1.

Multiplying each code by its coefficient and adding the signals produces the total radio signal amplitudes (+3,-1,+1,+1) and (-4,0,0,0).

Slide 72 shows that applying the inner product to the two amplitude sequences and the basis vectors, each station is able to find the correct bit transmitted to it.

Orthogonal codes can be built as Walsh codes as shown on slide 73:

W_0 is a 1 bit and its complement W_0' is a 0 bit.

W_1 is a square matrix of 3 W_0 elements and a W_0'.

W_2 is a square matrix of 3 W_1 elements and a W_1'. This gives the 4 basis codes used in the example.

This can be continued — a W_N matrix has dimensions $2^N \times 2^N$.

For the **reverse channel** (MS to base station), Walsh codes are not practical because they require synchronization in time among many stations. Instead pseudorandom noise codes are

used, which produce "nearly orthogonal" codes. A pseudorandom Bernoulli sequence is equivalent to a sequence of m coin tosses, and produces a nearly equal number of (+1) and (–1) elements in each code. By the central limit theorem, the probability of a (+1) or a (–1) is given by

$$P_{-1} = P(-1) = \frac{1}{2} + \delta \qquad P_{+1} = P(+1) = \frac{1}{2} - \delta \qquad |\delta| < \frac{1}{\sqrt{m}}$$

If two codes, A and B, are produced from as pseudorandom Bernoulli sequences $C^{(A)}$ and $C^{(B)}$, then we expect that

for A = B

$$\frac{1}{m}\sum_{i=1}^{m} C_i^{(A)} \times C_i^{(B)} = \frac{1}{m}\sum_{i=1}^{m} [\pm 1]^2 = 1$$

while for A ≠ B

$$\frac{1}{m}\sum_{i=1}^{m} C_i^{(A)} \times C_i^{(B)} = \frac{1}{m}\sum_{i=1}^{m} \left[P_{+1} \times P_{+1} - P_{+1} \times P_{-1} - P_{-1} \times P_{+1} + P_{-1} \times P_{-1}\right] = 4\delta^2 < \frac{4}{m}$$

So, for m large enough, the inner product will give 1 for A = B, and give a much smaller number for A ≠ B ("nearly orthogonal").

CDMAone uses 64 orthogonal Walsh codes in the forward channel and a long PN sequence in the reverse channel.

Slides 76 – 81 2G to 3G Cellular Transition

Slide 76 shows the changes made between 2G GSM and **3G** (UMTS). Aside from name changes of nodes, 3G is designed for greater coordination between clusters, implemented by a direct interface between the Radio Network Controllers (RNC) that replace the GSM BSC cluster controllers.

Slides 77 to 81 show the handover process in 3G UMTS. The MS is operating in cell 1 and moving toward cell 4. The MS is connected to cell 1 and being monitored by cell 2 in the same cluster (RNC-1). The MS moves to cell 2 within RNC-1 and is moving between cell 2 and cell 3 in the new cluster (RNC-2). The MS is connected to cell 2 and cell 3, with cell 4 monitoring. RNC-1 is in control and coordinates with RNC-2. The MS is then connected to cell 3 and cell 4 under control of RNC-1. Finally, the MS connects to cell 4 under control of RNC-2, with cell3 monitoring.

Slides 82 – 85 4G Cellular

The planning for 4G cellular began in 2000 in the ITU, at the same time that 3G was introduced. A major goal was NGN (next generation network) convergence with general support of mobility, with no distinction between voice and data. The principal objectives were state as:

- Higher network capacity than 3G
- Higher spectral efficiency (higher bit rates (bps) per analog bandwidth (Hz) and higher bps / Hz / per cell site)
- 100 Mbps for any moving client and 1 Gbps for any stationary client
- 100 Mbps between any two points in the world

- Smooth handoff across all heterogeneous network types
- Global roaming across multiple networks
- QoS for multimedia support — audio, HDTV, streaming
- Interoperability with existing wireless standards
- An all IPv6 packet switched network that eliminates circuit mode voice entirely (voice calls to be implemented as Voice over IP).

The world economic crisis of 2000 – 2001, the near economic collapse of 2007 – 2008, and international political turmoil since 2001 have left these goals largely unachieved. In the meantime, some new networking infrastructures have been introduced and manufacturers have been eager to label their products 4G. Recently the ITU withdrew its objection to the use of the term 4G to describe systems that do not meet the original 4G requirements.

In particular **Long Term Evolution (LTE)** is a 3.5G service marketed as 4G. LTE provides an improved radio interface and higher data rates, up to 299.6 Mbps downstream and 75.4 Mbps upstream. LTE also supports improved mobility, supporting an MS moving at 500 km/h.

Voice over LTE (VoLTE) is a packet switched voice system that replaces the circuit switched voice network. It provides enhanced voice encoding as a VBR stream with circuit-switched fallback (CSFB) if the packet mode system does not provide adequate performance. As an intermediate migration path, most cell providers have been installing data over LTE together with a standard circuit switched voice network.

As is the 2G to 3G transition, a number of network elements have been upgraded in LTE. The 3G UTRAN is replaced with the Evolved UTRAN (e-UTRAN). Similarly, the eNodeB replaces the UMTS NodeB in each e-UTRAN.

The general network architecture is called the **Evolved Packet Core (EPC)**. It includes the Mobility Management Entity (MME) that replaces the GSM VLR for management of roaming, and the Home Subscriber Server (HSS) that replaces GSM HLR for management of devices in their home networks.

As an all-IP network, no circuit mode switches are implemented and the hierarchical network of MSC / SGSN / nodes in GSM/UMTS is absent. The Serving Gateway (SGW) acts as an IP router, routing and forwarding user data packets within the cellular network. The Packet Data Network Gateway (PGW) routes and forwards user data packets to the Internet, replacing the GGSN in GSM/UMTS.

Slide 86 5G Cellular

The general plan for 5G cellular specifies a full redefinition of the cellular system as an abstract and global Software Defined Network (SDN). It is conceived as a Service Based Architecture (SBA), in which network functions are defined as stateless services rather than hardware functions. The functions are implemented through extensive virtualization. In general, networks will be designed by standards of software engineering.