Wireless Communications & Broadcasting

The rapid development of wireless communications systems and digital broadcasting has been one of the most exciting developments in electronics in recent years. Some of the most important frequency bands and applications of the electromagnetic spectrum are shown in Table 1. Wavelength (λ) is inversely proportional to frequency, and assuming free space propagation the relationship $c=f\,\lambda$ can be used where c is the speed of light in vacuo (3×10^8 m/s). For radio waves, the wavelength is large and the macro effects of electromagnetism and wave propagation readily apply. At the other extreme, light has a typical wavelength of <1μm and quantum effects start to become important: In other words, the packet nature of light – the existence of photons – must be taken into account.

Radio transmission is used for hundreds of applications. The frequency bands ranging from 100 kHz up to 30 GHz (30,000 MHz) are already extensively used for radio, TV, mobile telephones, private mobile radio (Taxis, Police, etc.), satellite TV, international telephony and data communications by satellite, and a whole host of other communications systems. Frequencies from 30 GHz to 300 GHz are already starting to be allocated to new systems, as the lower frequencies become congested. In addition to communications, the radio spectrum has slots allocated for other uses, such as radar, navigation, radio astronomy and remote-sensing. Remote-sensing is very important for environmental monitoring of things like the hole in the ozone layer and the destruction of tropical rainforests.

A key principle in radio communications is that antennas, which convert electrical currents into radio waves (or *vice versa*), need to be of a size comparable to a wavelength. This, essentially, is why high frequency signals are used for radio transmission – the antennas must be of a practical size. In order to send speech, for example, which only goes up to around 4 kHz, the trick is to 'piggy-back' the speech signal onto a radio-frequency ('RF') carrier signal. This process is called *modulating* the carrier.

Table 1 Common applications of various frequencies

Frequency	**Application**
0.5 to 2.6 MHz	Medium wave AM radio band
220MHz	Digital audio broadcasting
500 to 850 MHz	UHF TV band
850, 900 MHz and 1.8, 1.9 GHz	GSM mobile (cell) phones
2.2GHz	3G mobile phones
2.45 GHz	Microwave ovens, Bluetooth, 802.11b/g/n wireless LANs
10 GHz	Military Radar
11-12 GHz	European satellite TV band
77 GHz	Cruise control in some cars
198 THz	Long distance fibre communications
333 THz	TV infra-red remote control
430 THz	Red light
1 Million THz	X-rays in hospital

Managing the Radio Spectrum

In order to prevent different users interfering with one another, the use of the radio spectrum is carefully regulated by government agencies (such as the FCC in the USA and OFCOM in the UK) and international bodies such as the ITU (international Telecommunications Union). They issue licences to bona-fide users and seek out unauthorised users. There are many different types of licence, depending on the transmission frequency that you wish to use and on your application.

To ensure that radio equipment and consumer products are compatible, there is a tireless effort by standards committees to define the hardware and software specifications for different applications. Modern communications systems are so sophisticated that these documents run to hundreds and hundreds of pages. The IEEE standards are particularly important; for example, the IEEE802.11 wireless LAN standards were key in the explosive growth of Wi-Fi around the world.

Frequency allocation charts are available to download, and the number of different bands that have been allocated for diverse purposes is amazing. www.cedmagazine.com has a free wallchart of ITU allocations (pdf file) but there are so many allocations it is unreadable at A4 scale.

Modulation and demodulation

By themselves, sinusoidal radio waves do not carry any information. Modulation is the means whereby signals such as speech, music, TV, or computer data is "piggy backed" onto a radio wave so that the information can be transmitted without wires, cables or optical fibres. The easiest modulation technique to understand is amplitude modulation (AM). In this method, the RF signal amplitude is made to vary at the frequency of the speech signal. Fig. 1 shows the resulting waveform when the music is actually a pure sine-wave (e.g. a pure note from a musical instrument).

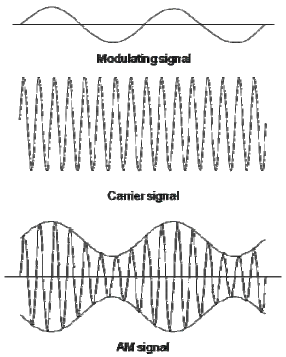


Fig. 1 Amplitude modulation

The amplitude-modulated RF signal can be received and demodulated using a simple tuned radio receiver, shown in Fig. 2. The antenna (or aerial) picks up the RF signal. The signal is now very weak as it has been broadcast in all directions by a distant transmitter. So, it is separated from other signals and amplified. Tuning is needed because in analogue radio systems different RF carrier frequencies are used for different channels, as illustrated in Fig. 3. After some amplification, the carrier *envelope* is detected and the speech is therefore recovered. The envelope detection can use the same rectification property of a diode that is used in a power supply. In very early radios, the diode was formed by connecting a fine pointed wire onto a bulk crystal, giving the name "crystal radio set".

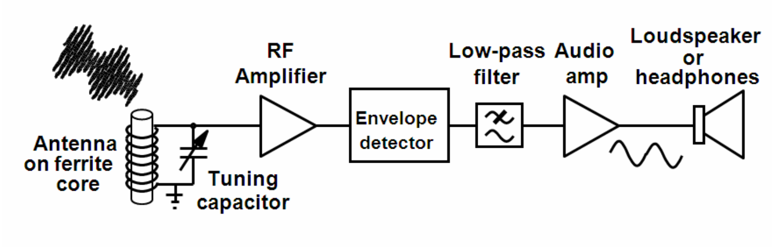


Fig. 2 AM Radio reception and demodulation

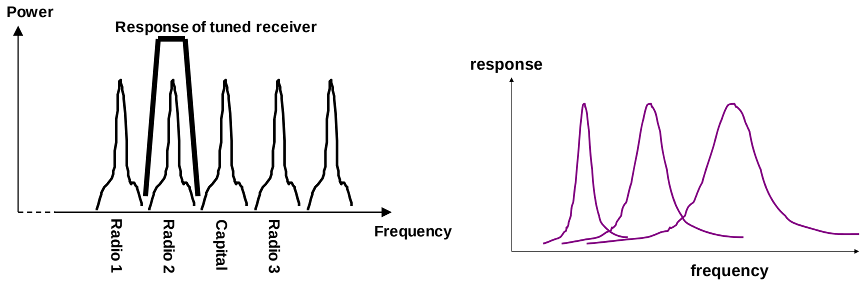


Fig. 3 Spectrum sharing for analogue broadcast radio (a) principle: (b) practical problem of bandwidth increasing as the receiver is tuned to higher frequencies

The superheterodyne receiver

The simple tuned radio frequency (TRF) receiver in Fig. 2 does not give adequate performance for a modern consumer product. Its main limitation is that as it is tuned to higher frequencies, its *selectivity* – the ability to select one channel frequency and reject the others - gets worse. This leads to interference between channels, noise, buzzes and squeaks. A better way of tuning from one station to another is necessary. The solution is the *superheterodyne* receiver, shown in Fig. 4, in which a *mixer* is used to convert the RF signal to a low frequency before filtering. The mixer is a circuit that can be used to frequency-shift an RF signal by multiplying it with a second sinusoidal high frequency – the *local oscillator* (LO) signal. The frequency-translated output signal of a mixer is called the intermediate frequency (IF) and this IF signal retains the information content of the RF input. In the superheterodyne receiver, frequency shifting the RF down to an IF enables the required channel to be filtered more cleanly from the unwanted neighbouring signals.

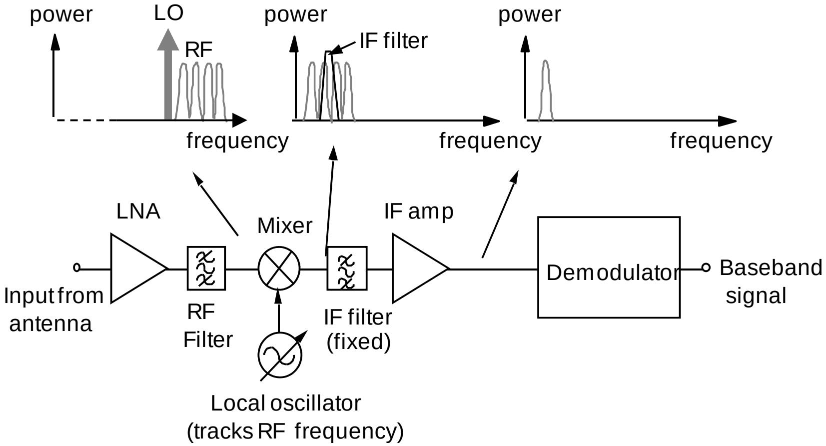


Fig. 4 Superheterodyne (“superhet”) receiver

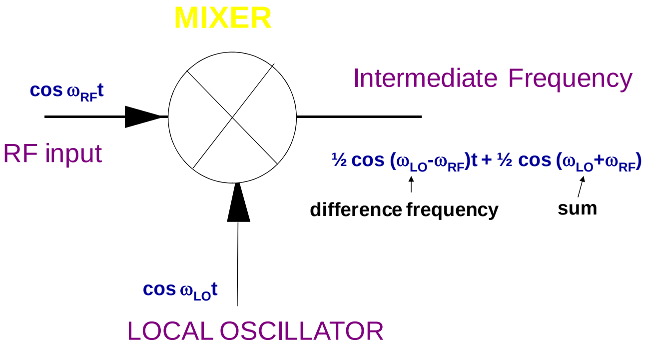


Fig. 5 Mixer symbol & signals

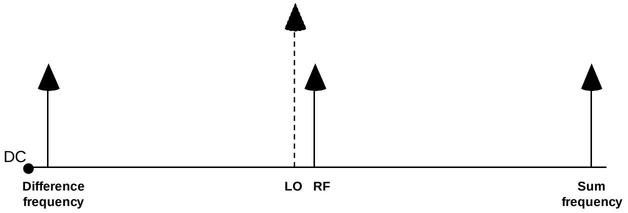
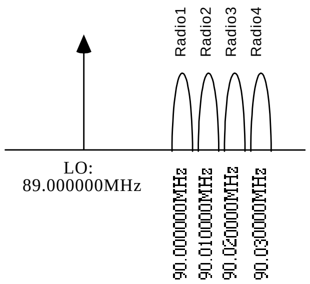


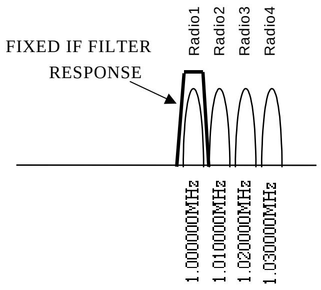
Fig. 6 Spectrum showing sum and difference signals

In the fictitious example below, the radio stations are in the 90 MHz band and separated by 10 kHz. The IF frequency is chosen to be 1 MHz. Hence, to pick up the station which is exactly at 90.000000MHz (“Radio1”) the LO is tuned to 89.000000MHz: This means that the RF signals are converted into two bands: 1 MHz and 179 MHz (difference and sum frequencies). The 179 MHz ones are unwanted but can easily be filtered out. So, the desired radio station is translated to 1.000000MHz. A bandpass filter, centred on 1.000000MHz with a 6 kHz bandwidth can separate the desired radio station from the others. Once the desired radio signal is isolated, amplification and envelope detection (for AM) will recover the speech/music.

SIGNALS IN RF BAND



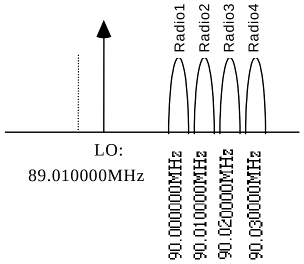
SIGNALS IN THE IF BAND



The real beauty of the superheterodyne receiver becomes apparent when we look at how its can be tuned to different radio stations (different channels): The *LO frequency* is variable, but the IF filter is kept fixed. Thus, the IF filter can be extremely narrowband (selective) and the radio stations can be received cleanly even when they are close together in frequency. The adjacent stations and noise are filtered out properly: No buzzes, squeaks, or hissing. The second example, below, shows the same radio example, with the LO increased by 10kHz so that “Radio2” falls within the IF filter bandwidth:-

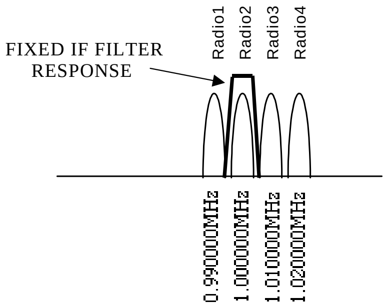
SIGNALS IN THE RF BAND:

The LO moves up by 10 kHz



IN THE IF BAND:

All the stations shift **down** by 10 kHz (IF=RF−LO)



However, another improvement is necessary to make a really good receiver. If the LO frequency is controlled by an L-C resonator its frequency is not precisely fixed and is prone to drifting with time, temperature, etc. The mechanical tuning knob has to be precisely set and may need adjusting from time to time, even if it is always left tuned to the same station. So, the LO frequency must be adjustable and yet more accurately fixed. This is achieved by phase-locking the LO to a reference oscillator which uses a quartz crystal as its resonator. This is exactly the same quartz device used in modern watches. This technique, in marketing terms, is often called "digital tuning", "PLL-tuning", or "synthesised". In passing, it is interesting to note the synergy with the early crystal sets.

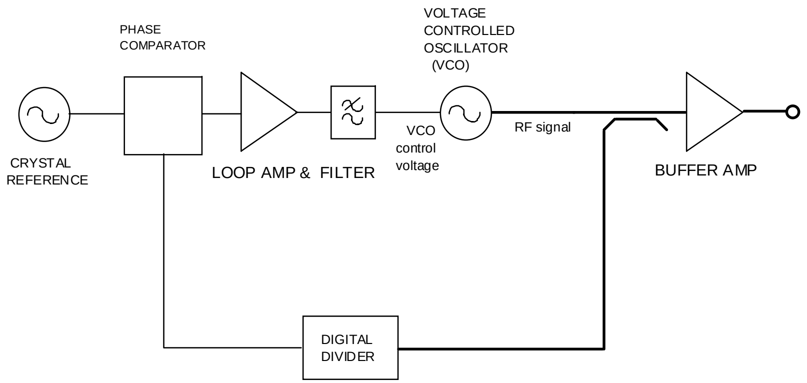


Fig. 7 Basic phase-locked loop system for creating a high stability local oscillator

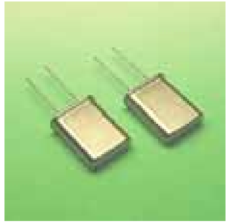


Fig. 8 Typical quartz crystals

The IMAGE band

An important, aspect of the superhet is that the mixer will downconvert another signal – referred to as the image which is the RF band *below* the LO frequency, but giving the same IF. This is why the block diagram has an RF filter before the mixer – this is to remove the image before down conversion. The RF spectra shown in the tuning example only show RF signals above the LO. However, in practice there will be either noise or other channels *below* the LO signal. In the first example, a signal at 88.000000MHz will also be converted down to 1.000000MHz since 89-88=1. This signal, on the other side of the LO to the desired RF signal, is called the image signal. It can only be filtered out at the RF frequency. After the mixer, it is at exactly the same frequency as the desired signal and cannot be removed.

The presence of the image signal means that an RF filter is required before the mixer. The major task of this is to remove the image signal. The frequency separation between the image and the RF is equal to double the IF frequency. Hence, the IF must be chosen high enough that the image and RF can be filtered out. But, if the IF is too high then designing a narrow-band IF filter may be difficult. So, there is a trade-off which must be addressed in the receiver design.

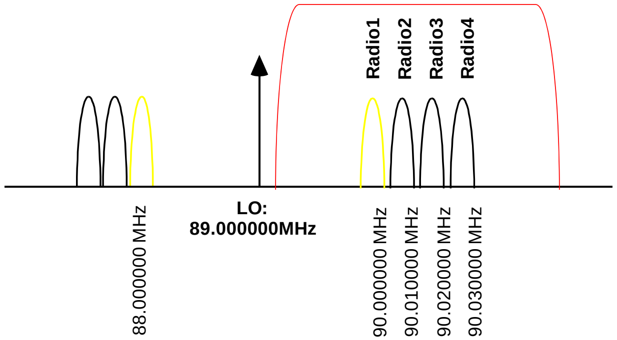


Fig. 9 Image Filtering

For some applications (particularly UHF frequencies upwards) the image and IF filter problems cannot be compromised. The solution is to use a double frequency conversion scheme. The first IF is chosen to be quite high, so that the image filtering is easy. The second IF is low to make the channel selection filter design easier. Often, three stages of frequency conversion may be necessary. In some applications, the first mixer may actually upconvert the signals.

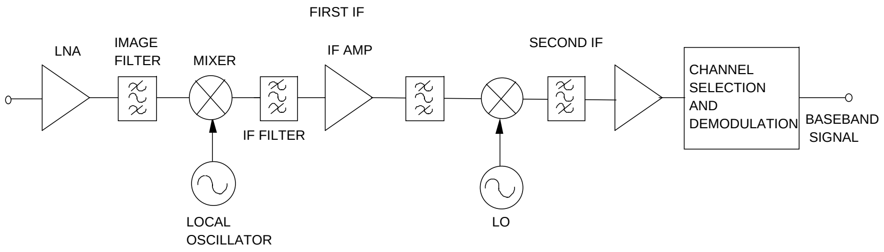


Fig. 10 Double conversion receiver

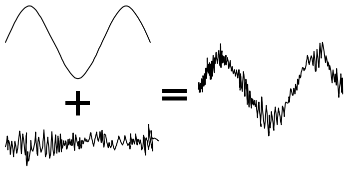
For compact silicon IC implementations it is virtually impossible to implement a receiver with all these filters, so many modern RFIC (radiofrequency integrated circuit) transceivers use what is known as direct conversion, or zero-IF architectures. These perform frequency translation directly to baseband, but have other problems and are discussed later.

FM radio

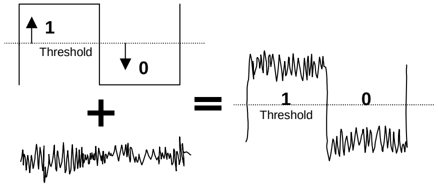
The drawback of amplitude modulation is that it is susceptible to noise and interference. 'FM' means frequency modulation and in this method the carrier *frequency* is changed very slightly in proportion to the speech signal amplitude. It’s very complex to visualise, but the end result is a system which is much more tolerant of noise and interference. Hence, the quality radio stations and stereo transmissions are on 'FM'. MW/LW refer to the carrier frequency band.

Digital radio

In radio transmission, the received signal level is often extremely close to the background thermal noise. In analogue systems, the information signal is inevitably degraded by this noise. The transmitter power can be increased to improve the received signal to noise ratio, but the expense is significant and there will be more interference with other users of the same – or adjacent – frequencies. Digital transmission is virtually immune to noise as long as the signal to noise ratio remains above some critical value. With error-correcting coding, the signal-to-noise ratio can be as low as a few dB and with a suitable receiver the digital data is still recovered successfully. Almost all modern communications systems are digital as a result. dB units are discussed in detail elsewhere: 3dB increase means double the power.



Analogue; noise adds directly



Digital; immune to noise up to a point

Fig. 11 Analogue vs. Digital communication

There are a number of digital radio broadcasting systems around the world including WorldDMB (Digital Multimedia Broadcasting), HDRadio®, and DRM (Digital Radio Mondale). The BBC (British Broadcasting Corporation) pioneered digital radio broadcasting and launched DAB (digital audio broadcasting), the world’s first service, in 1995. In 2006, DAB was extended to cover more services and became WorldDMB. WorldDMB generally uses either the VHF Band III (174.928 to 239.200 MHz) or part of the L-band (between 1452.960 to 1490.624 MHz). The channels are encoded with MPEG compression and COFDM modulation (described later) is used. Unlike the AM/FM band described so far, where each radio station has a different carrier frequency, digital radio stations can be grouped into *ensembles*, which aids channel tuning and searching. This is an important concept made possible by the digital representation of the signals; different data streams can be multiplexed together in time, as shown in Fig. 12. The quality of DAB is better than FM and especially AM, and the reception is more robust. The sound quality depends on the bit-rate, which can be varied by the broadcaster.

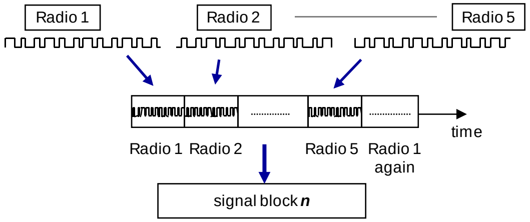


Fig. 12 Illustration of how stations are multiplexed together to form ensembles

Digital modulation

Digital data can be directly modulated onto an RF carrier using amplitude or frequency modulation, referred to as amplitude shift keying (ASK) and frequency shift keying (FSK), respectively, as illustrated in Fig. 13. In fact, FSK is believed by most people (erroneously) to be the technique still used by computer modems to communicate data down telephone lines. Although FSK was used a long time ago in the V23 modem standard, this only achieves 1.2kb/s. The limitation of this simple modulation scheme is that the drastic change in the signal (in this case its frequency) leads to a large modulated signal bandwidth which is very inefficient in terms of making use of the precious available bandwidth.

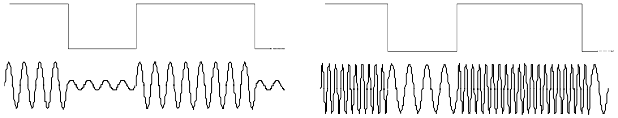

(a) (b)

Fig. 13 (a) ASK and (b) FSK transmission of data

One way of transmitting data more efficiently is to change the *phase* of the RF carrier. In what is called binary phase shift keying (BPSK) the sine wave is inverted when the data changes from a 1 to a zero, or *vice versa*, as illustrated in Fig. 14. The change is phase is quite hard to see – in fact the subtlety is the reason the spectrum does not spread much, allowing a higher data rate to be transmitted than FSK. However, in the receiver, electronic circuits that are very sensitive to the sudden phase changes can be designed to recover the data.

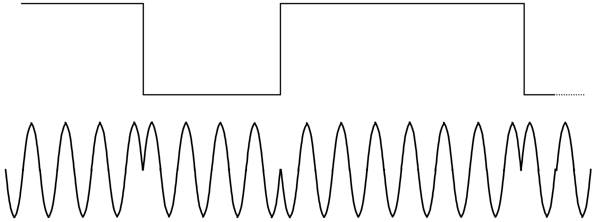


Fig. 14 PSK transmission of data

In a four-phase scheme called quaternary phase shift keying (QPSK), the signal can take four phase states; 0, 90, 180 and 270 degrees (360 degrees is indistinguishable from 0). This leads to the important concept of a **SYMBOL**. That is, the phase state has four possible values, so it can represent two bits of data, as shown in Table II.

Table II QPSK with two bits per phase state

1st Bit	2nd Bit	Carrier Phase
0	0	0
0	1	90
1	0	180
1	1	270

The sine wave is tedious to draw with four different phase states, so it is common to draw a representation of these symbols which just shows the phase angle. This is shown in Fig. 15, although in practice QPSK uses angles of 45, 135, 225 and 315 degrees.

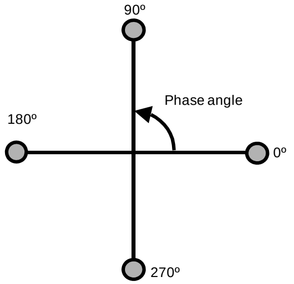


Fig. 15 Four phase states, on a phasor diagram

This multiple-phase concept can be extended; each one of 32 possible phase states could represent 5 bits of data, for example. However, in practice the phase states become too difficult to distinguish at the receiver because of noise. To get the really high data rates needed for modern digital radio and television, the concept of a symbol is extended to include the amplitude of the carrier as well. For example, 16 symbols, each representing a particular four bit sequence of data, can be created by choosing the amplitude and phase appropriately to form a square grid of amplitude/phase points. This regular grid of symbols is chosen because the points are spaced as far apart as possible apart and therefore easy to distinguish at the receiver end when noise has been added to the signal. The phasor diagram of Fig. 16 is referred to as the **constellation**, and this modulation scheme is called 16-QAM (16 symbols, quadrature amplitude modulation). The difference between BIT RATE and BAUD RATE is now simple to understand; the BAUD RATE is the rate at which SYMBOLS are transmitted. By transmitting "more complicated" symbols, more data can be transmitted in the same amount of RF bandwidth. A useful analogy might be to compare English with Chinese characters. One Chinese character, with its complex design & strokes, can represent the many English letters of a word.

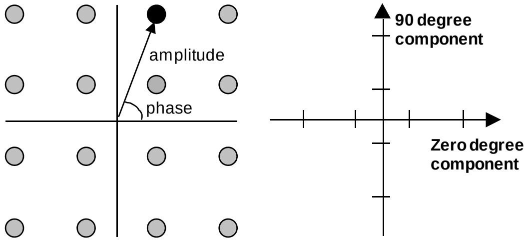


Fig. 16 16-QAM Constellation

It is called *quadrature* amplitude modulation because if you study the constellation you will see that the 16 points can be generated by adding together a separate **ZERO degree** component, with 4 amplitude levels, and a **90 degree** component, also with 4 levels. Signals 90 degrees apart are said to be in QUADRATURE. In digital communications, the zero degree component is called "*I*" (for *in*-phase) and the 90 degree component is called "*Q*". I and Q are not simply separate data signals; each permitted pair of I & Q values represents one symbol, and that symbol contains 2bits (QPSK), 4 bits (16-QAM), 8 bits (256-QAM) and so forth.

A pair of sinusoids that are in quadrature is shown in Fig. 17. The quadrature nature of these signals is important because it is the key to getting more data on the carrier without using more bandwidth. The carrier tones have the same frequency, so in the frequency domain they share the same bandwidth. However, look at point "*X*" in the time domain and the *cosine* is at a peak, whilst the *sine* is zero. Hence, if the signals are sampled at that point, the data can be read off the *cosine* signal, and the data on the *sine* signal is ignored; in other words you can send two lots of data at the same time without using any extra bandwidth. This is a simplification, but proper treatment of the subject requires considerable mathematics.

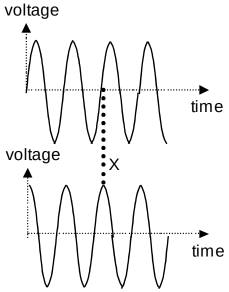


Fig. 17 Sine (top) and cosine functions; Quadrature signals

The ratio of data rate to bandwidth used is known as spectral efficiency. In the idealised case, the values for the different digital schemes described are:-

BPSK 1b/s/Hz QPSK 2b/s/Hz 16-QAM 4b/s/Hz
bits-per-second of data, per Hz of spectrum used in the transmission

---

Mobile Communications

Radio propagation has the huge advantage of being wireless – and therefore essential for applications like mobile communications. However, the radio spectrum is an extremely limited resource. Furthermore, because of the propagation characteristics, only a small part of the spectrum is suitable for mobile communications. Low frequencies (up to a few GHz) will readily diffract around buildings and other obstacles. Higher frequencies, especially above 10 GHz, will travel only along a line-of-sight path. Furthermore, higher frequencies can suffer very high atmospheric loss, limiting their use to short ranges.

That leaves only up to a few GHz for mobile communications but, of course, TV, radio, etc, all need to use some of that spectrum, as well as radar systems, radio telescopes, car alarms, police radios, etc etc. Only a tiny slice of spectrum can be allocated to mobile services, and this is why data rates are low and mobile internet is still rather limited. The cellular approach to mobile communications, where many base stations are installed each serving quite a small area, as shown in Fig. 18 is essential because:-

(a) the mobile terminal is battery powered and its transmit power and ∴ range is severely limited
(b) there are too many users for a long range system to be viable.

In the cellular system, the radio spectrum available for mobile communications can be used over and over again in different cells as long as they are not adjacent. The radio link is kept to a fairly short range by reducing the transmitter power, and a complex network of interlinked basestations is used. The basestations are linked with optical fibres under the ground, where possible, or with very high frequency line-of-sight radio links (38 GHz, for example).

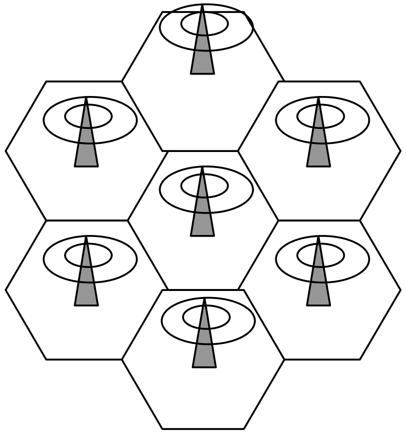


Fig. 18 Cellular System

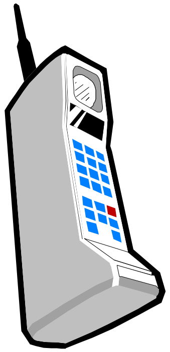


Fig. 19 Early "brick phone"

The first mobile systems were analogue and the pioneers were in Scandinavia (which is why Nokia and Ericsson have become such heavyweights in the industry) and the UK. In the UK, Racal was the leader in the technology and their work led to the TACS system being introduced in 80's. The early brick phone (Fig. 19) is an icon of this era – along with yuppies and wine bars. These first analogue systems are referred to as the First Generation systems.

In mobile systems the propagation characteristics are nasty; you rarely get line-of-sight to a basestation, you are often indoors with many obstacles between you and the basestation. With analogue systems it is hard to counteract all the ill effects this causes and the quality is not good. The early systems used FDMA (frequency division multiple access – users transmit on different frequencies) which tends to lead to a bulky mobile phone.

Digital systems were introduced in the UK in the 90's. The UK has adopted the GSM system (originally Groupe Special Mobile, a pan-European standards initiative) and this became so successful it was adopted in many countries worldwide, becoming known as Global System for Mobile communications. These first digital systems are referred to as Second Generation (2G) systems. They give excellent voice quality and GSM gives roaming capability in many countries. Different systems and countries use different frequencies, however, which is why there are dual-band and tri-band phones. The GSM system uses TDMA (time division multiple access – different users within one cell can use the same frequency but transmit in different time slots). This enabled the phones to be made smaller and lighter.

In the US, the huge size of the country is a major barrier to cellular mobile communications. No single company could install a countrywide network. So, the US has had a more piecemeal solution with different companies and standards operating in different cities. The IS-95 system using CDMA has become the most common 2G digital system.

For internet use and emailing, the 2G systems are unattractive; the data rate is too low and the system is like a dial-up service. For this reason, enhanced systems were developed which gave "always on" connection and higher data rate. In the UK these are (were?) GPRS (general packet radio system) and GSM-EDGE (enhanced data rate system). These are referred to 2.5G systems and seem to have received almost zero publicity in the UK.

The Third Generation systems (3G) were conceived from the outset as being capable of high data rates; for video and high speed internet access, not just for voice. The conventional TDMA and FDMA systems are unable to provide this for millions of users. **CDMA** (code division multiple access) is the solution. In this approach, many users transmit at the same time and in the same frequency range. However, their signals are all spread out thinly over that frequency range, and the spreading is performed by multiplication with a unique pseudorandom code for each user. In a receiver, only the signal component with the correct code will be CORRELATED with the code in the receiver. By this mechanism, the other signals are all rejected as noise-like. It sounds like it should not work at all, but in fact CDMA has proved able to accommodate many more users in a given cell than TDMA and FDMA. Huge technical hurdles had to be overcome though.

3G systems have been widely deployed around the world. As yet, the systems seem unable to offer any significant practical application over and above that offered by 2G systems, let alone 2.5G. The problem is simply that most people might be willing pay 10p for one text, which is 2kb of data, but the price they are willing to pay for a video download is much less PER kB. One advert suggested downloading jokes using 3G. Probably most sensible people are more worried about the cost of their calls, and calling cards are a booming business, while video downloading can only take off when the price is right.

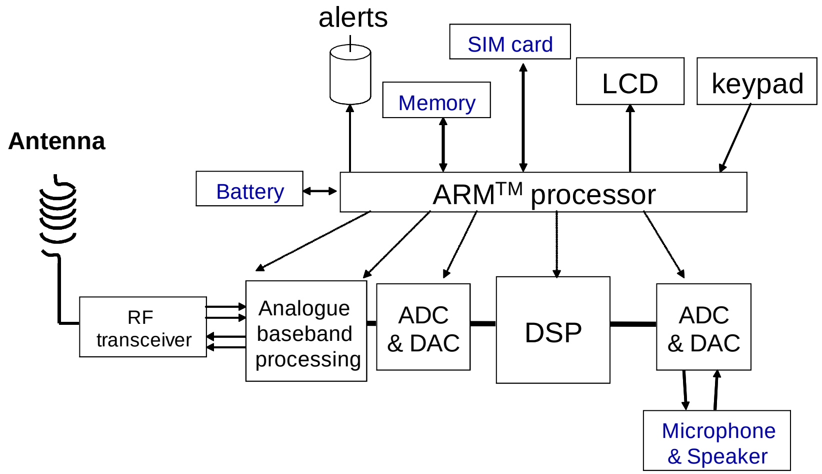


Fig. 20 A Simplified Block Diagram

Fig. 20 shows the simplified block diagram of a basic mobile phone. The components in the top half will be immediately familiar. The communications hardware is in the lower half. It might seem illogical that the signal goes from analogue to digital and then back to analogue before being modulated onto the RF carrier. The reason is that modulating a pure digital signal onto the RF carrier would lead to a spectrum with many sidelobes. To make more efficient use of spectrum the data signal has to be carefully shaped, and so a smoothed "analogue" version of the digital data is fed into the modulator. This is worthy of some explanation, as it leads to an important result.

First, note that mixers are the component most often used for modulating the RF signal. A mixer is used in the superhet receiver, described earlier, for frequency conversion. Referring to Fig. 21, it is hopefully possible to appreciate that multiplying the sinusoidal RF carrier by the digital data produces an amplitude modulated carrier.

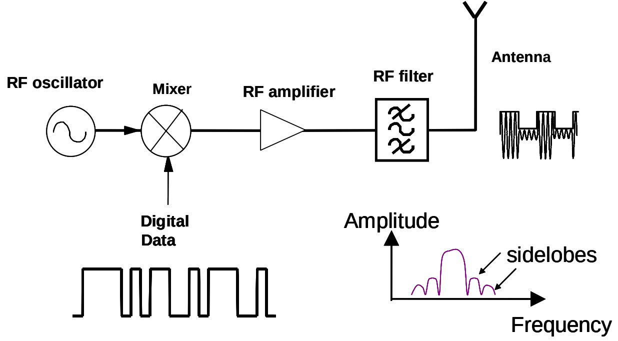


Fig. 21 Basic Digital Transmitter

From Fourier series analysis, it can be shown that modulating digital data directly onto the RF carrier leads to sidelobes, which is an inefficient use of precious spectrum. This basic transmitter can only be used in simple applications – eg a radio controlled car. To eradicate the sidelobes, the data signal must be "smoothed" to remove high frequency components:-

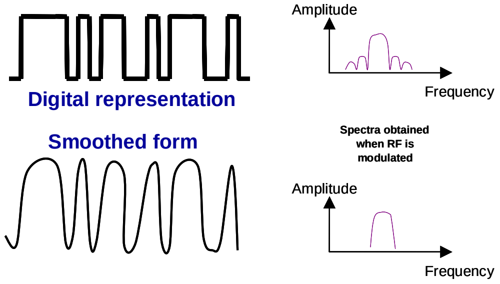


Fig. 22 Spectrum shaping by baseband filtering

Thorough study of the "smoothing" is in "Digital Communications" modules and books, but essentially it involves careful choice of *matched filters* in the transmitter **and** in the receiver. The smoothed data signal is now essentially analogue; the filtering process is usually implemented in "DSP" – i.e. in a high speed digital signal processor – a microprocessor optimised for signal processing operations. A Digital-to-Analogue Converter (DAC) is used to convert the signal to analogue before feeding the mixer:-

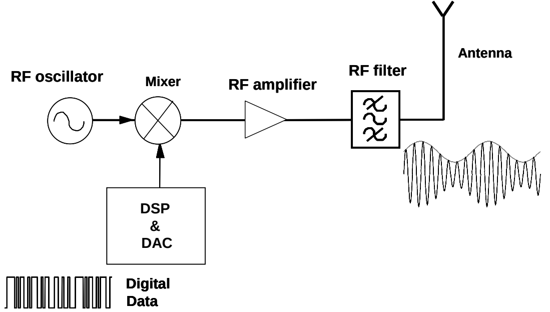


Fig. 23 Basic digital transmitter using DSP for baseband filtering (one more complication to add!)

The final complication to add is that modern spectrally efficient modulation schemes use quadrature modulation, discussed earlier. This can be conveniently implemented by using two of the mixers, operating on carrier signals that are 90 degrees out of phase, illustrated in Fig. 24. This architecture is ubiquitous nowadays in modern wireless transceivers. I and Q are not simply alternate "bits"; the incoming data is mapped onto symbols, discussed earlier, and the required I & Q values for each symbol are then determined and fed into the DACs.

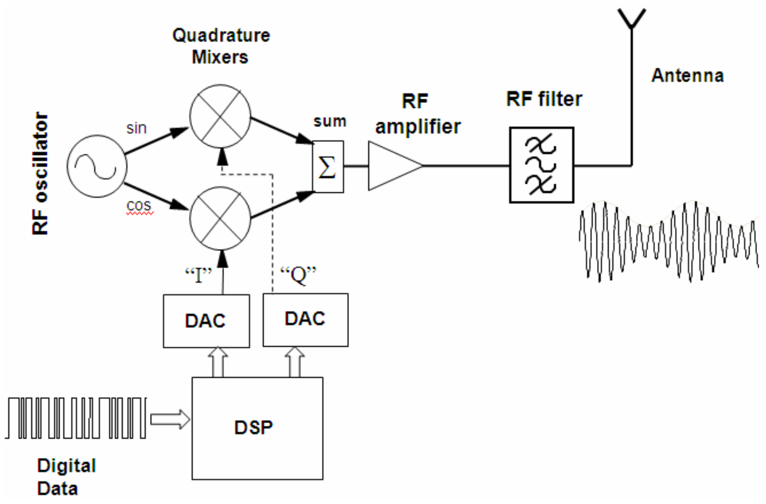


Fig. 24 Digital transmitter using the IQ modulation architecture

In this transmitter, I & Q are directly modulated onto the carrier. The reverse process can be used to make a receiver that is highly suited for RFIC integration –i.e. directly convert the received RF signal down to the baseband signals (I & Q), as illustrated in Fig. 25. This so-called direct conversion (or zero IF) receiver architecture is illustrated in Fig. 26. The wanted channel, now centred around DC, can be separated from the others with a high-order lowpass filter. Lowpass filters can be integrated onto ICs with a range of techniques (e.g. op-amps). The automatic gain control (AGC) amplifiers in the receiver are required because the analogue-to-digital converters (ADCs) cannot by themselves cope with the vast variation of signal strength encountered in a phone, for example. A pair of mixers, driven in quadrature by the LO, is required so that all the I & Q baseband information is preserved in the downconversion process. Direct conversion receivers have drawbacks of their own, but a vast R&D effort around the world has led to a range of solutions to them and they are widely employed in modern wireless products.

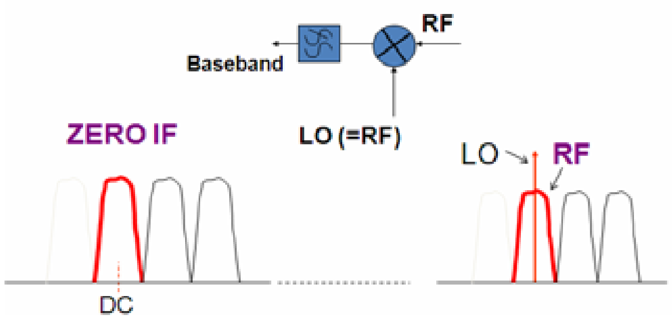


Fig. 25 Illustration of direction downconversion receiver technique

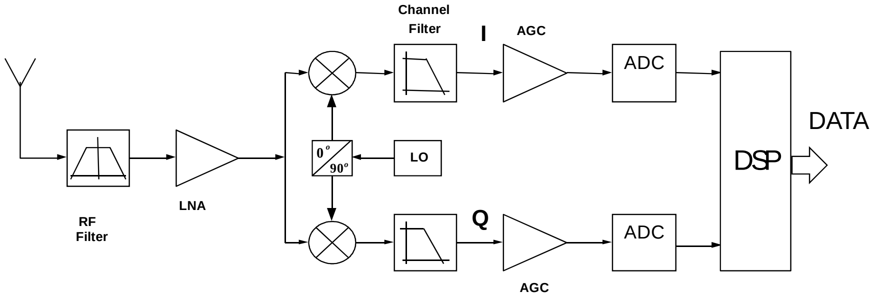


Fig. 26 Direct conversion receiver architecture

Terrestrial television

Around the world the analogue TV system has had to make way for digital systems. The analogue TV standards, NTSC (National Television System Committee), PAL (Phase Alternating Line) and SECAM (Séquentiel couleur à mémoire) are rapidly being confined to history. Digital TV is very complex though, requiring powerful video compression as well as sophisticated modulation & coding schemes. After all that clever processing, digital TV signals actually occupy a similar bandwidth to the analogue ones (e.g. around 8MHz bandwidth for PAL). A fair question is; why go to all this trouble, when the analogue system had been working very well for around 70 years?

From the TV transmitter the signal forms a travelling electromagnetic wave and the power in that wave is spread out over an increasing area as the distance from the transmitter increases. For an idealised point source transmitter, the power would be equally distributed in all directions over the area of a sphere ($4\pi r^2$), as illustrated in Fig. 27(a). This kind of transmitter radiated pattern is referred to as *omnidirectional* or *isotropic*. If the TV transmitter had an antenna like this, the vast majority of its precious RF power would be wasted - for example, radiated upwards into space. Hence, the antenna is designed to concentrate its power in the directions where it is needed – typically the radiation pattern would be doughnut shaped for a terrestrial TV transmitter, as illustrated in Fig. 27(b). In satellite communications, a dish is used to concentrate the power even more, into a narrow pencil beam as shown in (c).

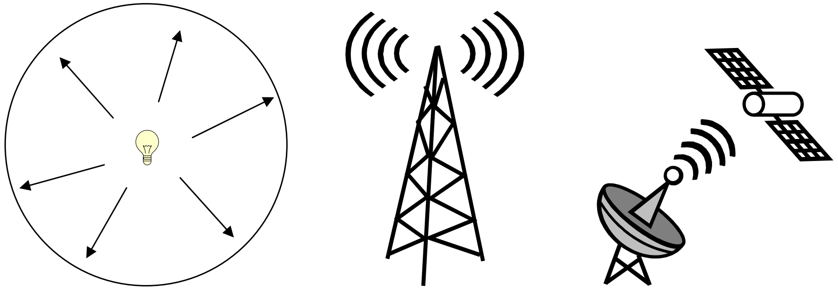


Fig. 27 (a) Idealised point source transmitter giving omnidirectional coverage (b) doughnut-shaped radiation pattern from a typical terrestrial transmitter (c) pencil beam from a satellite dish

With the power density of the wave falling off as $1/r^2$, a simple calculation shows that in theory a single transmitter antenna could comfortably cover the whole of the UK or a typical State of the US. However, the curvature of the earth means that there is a *radio horizon* which is determined by the height of the transmitter (TX) and receiver (RX) antennas and the curvature. This is given by $\sqrt{17h}$ km [Colour Television, Hutson, Shepherd, Brice], where h is the effective antenna height in metres. This is why the transmitters are such tall towers, mounted on hills. A 300m transmitter height gives a 70km radio horizon, so the Crystal Palace transmitter, for example, serves all of London. The Crystal Palace transmitter's effective isotropic radiated power (EIRP) is of the order of a megaWatt. The actual power output is less than this, but it is the effective value, which takes into account the transmitter antenna's radiation pattern, that determines the signal strength at the TV receiver antenna.

Fig. 28 shows a stylised map of the main transmitter sites in the UK for UHF TV. There are many other, smaller transmitters or repeaters required to fill in gaps in the coverage that are caused by blocking of hills and buildings. For example, in the London area alone there are about 50 low power transmitters in addition to the main Crystal Palace one. With analogue signals, receivers are very sensitive to interference; "ghosting" is a clear visual indication in analogue TV of how a receiver is affected when it receives two components of the same signal, from different transmitters. The ghosting is because of the small but noticeable difference in the time at which each signal component reaches the receiver. To avoid this interference, with analogue TV it is necessary to ensure that adjacent transmitters always use different radio carrier frequencies. The frequency allocations and the coverage must be carefully planned. In the UHF band, some diffraction occurs around buildings and even around the tops of hills because of the fairly long wavelength of the signal (500 MHz in free space gives 0.6m wavelength), but even so, repeaters are used to fill in poor coverage areas such as valleys where the main transmitter signals cannot reach. The signals from a non-adjacent transmitter with different coverage area can be on the same frequency because of the physical separation; normally the signals simply don't reach far enough to interfere. However, sometimes the atmospheric conditions can make a signal travel further than normal by refracting (bending) it past the normal horizon. Then, non-adjacent transmitters can interfere; for example, in Kent (in the South East of England) you might start receiving signals from Belgium (across the English Channel, in mainland Europe).

The need to use many different frequencies in the analogue system to avoid interference, unfortunately, leads to a very inefficient use of the precious UHF spectrum. This is the main reason why digital TV is now dominant. To show just how more efficient the digital TV system is, in the UK there were only 5 terrestrial analogue TV stations, even after many years of technical development: however, within a few years of its introduction to the UK, digital terrestrial TV offered more than 50 stations, even before the turning off of the analogue system.

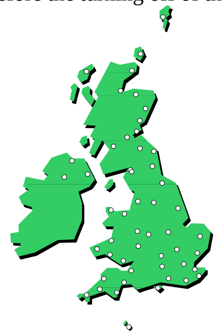


Fig. 28 UK Terrestrial TV transmitter sites

Digital terrestrial TV

On December 11, 2006, the Netherlands became the first country to switch off analogue and move entirely to digital TV broadcasting. The number one advantage of digital TV is that more channels are possible, but making this happen is a major feat. The first step is to use video data compression standards such as MPEG-2. The second key technology is to do with the highly advanced coding and modulation techniques that are used. There are five major digital terrestrial TV standards around the world at the time of writing:-

ATSC: Advanced Television Standards Committee
USA, Canada, Mexico, South Korea
The main terrestrial version of this standard uses a modulation scheme called 8-VSB (vestigial sideband), which is an advanced form of ASK, mentioned earlier, using 8-levels of amplitude and spectrum shaping to get high spectral efficiency.

DVB-T: Digital Video Broadcasting (Terrestrial)
Europe, South America, China, Africa, Asia, Australia
DVB allows for a range of different signal formats for different broadcasting systems: e.g. DVB-S2 for satellite, DVB-C for cable, DVB-H for handhelds..

ISDB: Integrated Services Digital Broadcasting
Japan

DMB-T/H : Digital Multimedia Broadcast-Terrestrial/Handheld (provisional name)
China

SBTVD: Sistema Brasileiro de Televisão Digital
Brazil

DVB-T, ISDB and others use a complicated system called OFDM (orthogonal frequency division multiplexing), the purpose of which is to combat the effects of multipath propagation. Fig. 29 illustrates how multipath propagation leads to echoes that limit the data rate. In OFDM, the digital data is split up and sent in parallel using a closely-packed block of *thousands* of RF carriers, each with a much lower individual data rate, as shown in Fig. 30. This type of signal is more resilient to the multipath fading than a single carrier carrying a high data rate. COFDM employs a coding scheme to recover lost data when any carriers are lost due to the multipath propagation.

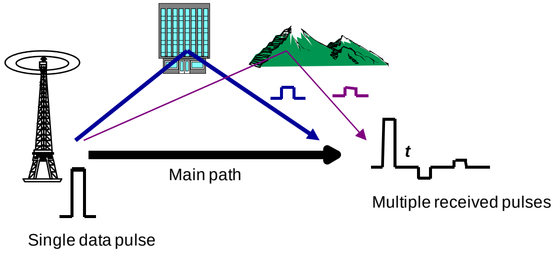


Fig. 29 Illustration of multipath propagation and its limiting effect on data rate

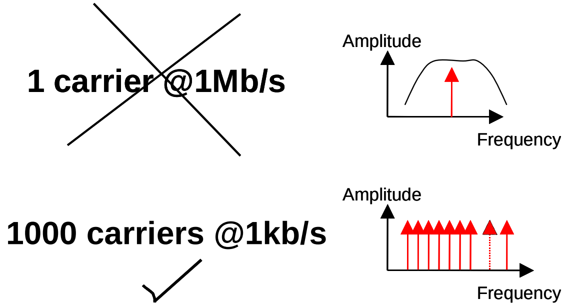


Fig. 30 The OFDM principle; the data is spread out over many sub-carriers to lower the data rate in each

Disadvantages of digital TV

Analogue systems degrade steadily and get more noisy; if, for example, a motorcycle went past your house transmitting lots of electrical noise from its ignition systems, you would typically see horizontal bars across the screen but can still see what's going on. Digital systems give very good quality when the signal-to-noise ratio is above a certain level, but data suddenly gets all corrupted when it falls below that threshold. In digital TV the picture can suddenly go all "blocky" and often freezes completely, which is very annoying because you can't see anything. The second potential disadvantage of digital TV is that artefacts of the video compression can become noticeable if the bit-rate is low. This is particularly noticeable in football matches, where the algorithms struggle with many people moving around in various directions, in front of a fixed background. Satellite and cable services seem to provide better pictures because there is more spectrum available and bit rates can be higher.

Satellite TV

Arthur C. Clarke famously proposed the use of geostationary satellites for communications. [*Peacetime Uses for V2*, Wireless World (February 1945) and *Extra-Terrestrial Relays: Can Rocket Stations Give World-wide Radio Coverage?*, Wireless World (October 1945)]. For a geostationary satellite, by setting the altitude to 35,786 km, the spacecraft naturally orbits the earth at exactly the same angular velocity as the rotation of the earth on its axis, appearing therefore to be stationary in the sky. To achieve this, the satellite must be orbiting in a plane perpendicular to the Earth's axis. Hence, in the UK the geostationary satellites will be at an elevation of 25.4 degrees (in London) to the South, as shown in Fig. 31. Around the equator, geo-satellites are directly overhead. Near the poles, geostationary satellites start to become useless as they are at a very low elevation, and easily obscured. Other orbits, like those shown in Fig. 32, can be useful for these areas and for other applications such as mobile communications where the satellite needs to be more-or-less overhead for a consistent link. Then, of course, the satellite is moving relative to the earth and will eventually disappear from view, so multiple satellites have to be used to give continuous service – the "constellation" used by communications systems such as Iridium and Globalstar™ are examples of this. The global positioning system (GPS) uses a constellation like this, and the receiver calculates the time difference between the signals received from different satellites in order to triangulate its position.

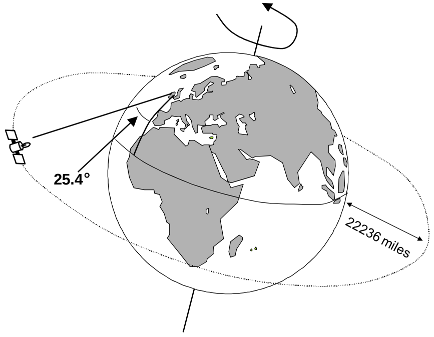


Fig. 31 TV broadcast satellite in geostationary orbit. The elevation from London is shown.

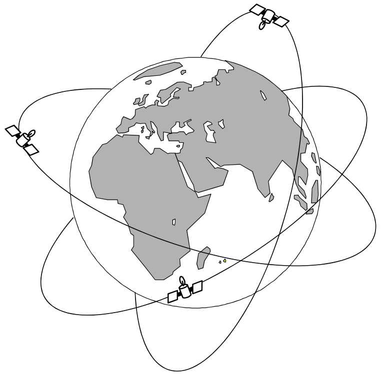


Fig. 32 An example satellite constellation

Satellite TV uses *microwave* frequencies (much higher than ordinary TV) because:-

1. There is plenty of spectrum available, allowing many channels,
2. Satellite communications requires antennas with high gain.

Antenna gain is a figure of merit which describes how effectively it concentrates power into the desired direction – i.e. towards the receiver(s). The ideal omnidirectional antenna, described earlier, spreads power in all directions and has a gain of 0 dBi (dBi=decibels relative to the isotropic case). A satellite dish achieves a narrow pencil beam of typically 2-5 degrees, and has high gain (maybe 35-40dBi). At the transmitter end of the system this is necessary because, from the satellite, the Earth looks tiny. The satellite only has solar power to rely on, so its transmitter power is severely limited. The narrow beam is necessary to concentrate the precious transmitter power on the Earth, to the appropriate countries. As illustrated in Fig. 33, at the receiver end a narrow beam is also necessary, in order to reject signals from other satellites and background noise from the sky.

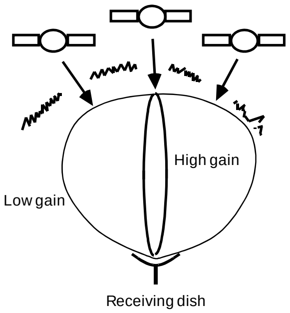


Fig. 33 Satellite dish beamwidth; low and high gain antennas compared.

Generally, a high gain antenna must be many wavelengths in size. Using microwave frequencies enables high gain to be achieved with reasonably small dishes (e.g. 60 cm for the domestic dish). This is most critical for the satellite: A 10 metre dish would cost a fortune to launch. As a result, extremely high carrier frequencies are used for satellite TV: in Europe frequencies from 10.7 to 12.75 GHz are used. To receive them on the ground, first a dish with a feed is required to collect the signal. The dish part is used to reflect the electromagnetic power and focus it on to the feed. The signals are so weak, because of the long transmission distance and low satellite transmitter power, that there must be a direct line-of-sight between the receiver dish and satellite. The feed collects the signal and sends it to a low-noise amplifier (LNA). The LNA must be mounted on the dish to amplify the weak RX signal immediately. After amplification, the signal is frequency converted, like in the superheterodyne receiver. This is necessary because microwave cable is expensive and has high loss: If you sent the microwave signal directly into the lounge, the 30 feet (approximately) of cable would cost about $1000 and the loss of the cable would be so high that the signal would disappear into the noise. By first amplifying and then converting the microwave signals down to approximately 1-2 GHz, much cheaper co-ax can be used and the loss is lower. The 1-2 GHz signal is then fed down a cable into the "set top box" (STB) where all the complicated signal processing takes place. The unit that amplifies and downconverts the signal is mounted on the dish close to the feed to avoid losses and is known as the "low noise block" (LNB). The components of a typical satellite TV installation are shown in Fig. 34. The block diagram of the LNB and a typical layout are shown in Fig. 35.

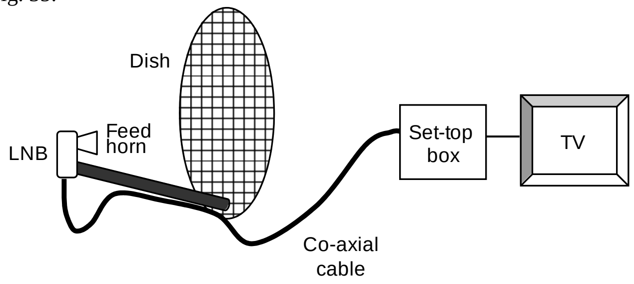


Fig. 34 Components of a satellite TV installation

(note: cables are routed with loops so that rain drops collect there and fall off, rather than flowing down the cable into the house or equipment; thanks to Dr S. Lucyszyn for this information)

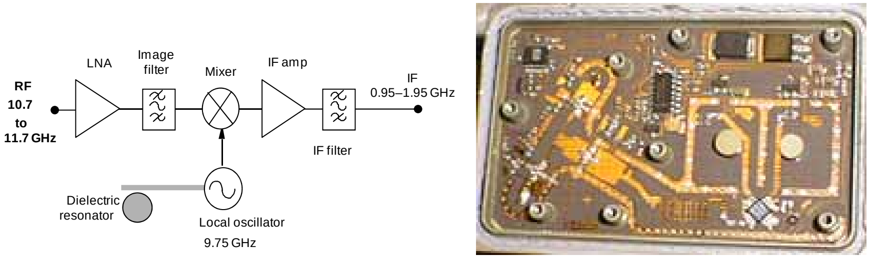


Fig. 35 (a) Block diagram of an LNB; (b) typical layout of the LNB circuitry

STB example

A satellite TV set-top box (STB) serves as a good example of the hardware required to receive digital TV. The signal frequencies and modulation schemes vary between terrestrial, cable and satellite systems, and between countries, but from a product designer's point of view that is someone else's problem; they are just blocks on a block diagram, representing different application-specific ICs for each case. Pace, based in UK, is a world leader in STB design and manufacture, with products covering all the major formats sold around the world. The Pace DS810XE High Definition Digital TV receiver has the following main specifications:-

Hardware

Main Core Processor @ 300MHz (420DMIPS)
950MHz-2150MHz QPSK/8PSK tuner and DVB-S/DVB-S2 demodulator
2Mb/s to 30Mb/s performance
22kHz LNB control and DiSEqC 1.2 communications protocol
High Definition and Standard Definition single video decoding
MPEG-2 MP@ML / MP@HL 576i25, 576p50, 720p50, 1080i25
MPEG-4 AVC/H.264 HP@L4.0 1920x1080 interlace@50Hz (25fps)
AC3 (Dolby Digital Stereo), E-AC3 (Dolby Digital Plus)
MPEG-1 layers 1, 2 and 3 (MP3)
Audio decoder, sampling rates 23, 44.1 or 48kHz
Image format 4:3 or 16:9 aspect ratios
Macrovision copy protection
64MB or 128MB DDR DRAM Memory Options
16MB Main Flash, 32KB NVRAM
IR-remote control - RC6 36/38 KHz

Software

Advanced installation and diagnostics features
Secure bootloader and upgrade capability
Multilingual graphical User Interface
DVB-subtitles according to ETS 300 743
Analogue Teletext loop through to TV
Software download via Satellite into Flash Memory

Fig. 36 shows a simplified block diagram of the STB. The heart of the design is a powerful system-on-chip, which is sufficiently power hungry to require a substantial heatsink. It is really educational to look at the block diagrams for such a chip, for example:

http://www.broadcom.com/collateral/pb/7312-PB04-R.pdf
or http://www.broadcom.com/collateral/pb/3549-PB01-R.pdf

The first functional block in the set top box chain is an RF front end that downconverts the desired signal to baseband (IQ) form. This is followed by a sophisticated demodulator chip that recovers the data stream and performs error correction. This feeds the system-on-chip MPEG decoder, which incorporates MPEG-2 video and audio decoders and a video graphics processor. This chip also includes circuitry to handle the user interface, an Ethernet or modem network connection, and a wide range of interfaces, such as USB2.0. This particular model has a separate video decoder chip for dealing with High Definition formats.

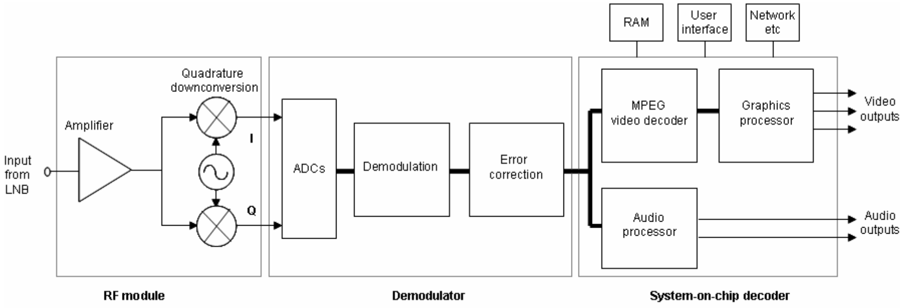


Fig. 36 Satellite HD TV STB Block diagram

Cable TV

Cable TV tends, by its nature, to be restricted to densely populated areas; certain residential areas in a town or city or individual apartment blocks. Cable TV can be a standard service from a big company like Virgin Media, serving many parts of the country, or it can be a local system just for an apartment block. CATV originally stood for Community Antenna TV, and was developed originally for areas where TV reception was poor – e.g. obstructed by mountains. In earlier systems, an operating company would distribute signals down cables in a frequency range compatible with the standard UHF antenna input of the TV. That way, many more channels could be accommodated because the interference issues relating to broadcasting were overcome. Modern systems usually require a set-top box, for TV signal reception and demodulation as well as broadband internet access. Of course, the downside of cable TV is that you have to pay a substantial subscription fee to receive the channels. Cable TV is a particularly attractive option when it is bundled with a broadband internet connection.

SMATV stands for satellite master antenna TV. Here, rather than having an apartment block festooned with many satellite dishes, the property management company has a single dish on the roof and distributes the signals by cable to every flat. Subscribers pay a fee for a receiver box, and most apartment blocks will have a clause in the lease which prevents tenants from fixing satellite dishes or even regular TV aerials to the building.

LMDS

Strictly speaking “LMDS” (local multipoint distribution system) refers to a specific standard system operating in the 26 GHz region (the exact frequency depends on the country), but the name nicely describes the concept. Rather than connecting cable to every house, which is uneconomical in rural areas and many suburbs, the “last mile” connection is made by a microwave or millimetre-wave radio distribution system. For interactivity, the normal phone network can be used for the return link – most internet traffic is “asymmetric” – that is, users download much more data than they upload. Other names for this hybrid type of approach have included WLL (wireless local loop), MMDS (microwave multipoint distribution system) and MVDS (multipoint video distribution system).

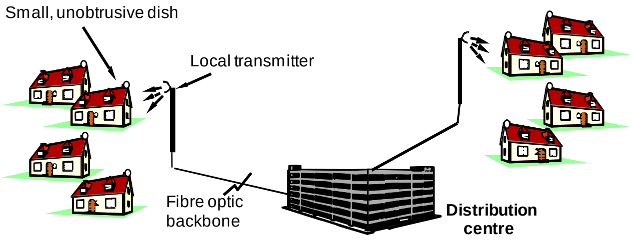


Fig. 37 The local multipoint distribution system concept