EIE312 COMMUNICATIONS PRINCIPLES

Outline:

Principles of communications:

1. An elementary account of the types of transmission (Analogue signal transmission and digital signal transmission). Block diagram of a communication system.
2. Brief Historical development on communications:
   a. Telegraph
   b. Telephony
   c. Radio
   d. Satellite
   e. Data
   f. Optical and mobile communications
   g. Facsimile
3. The frequency Spectrum
4. Signals and vectors, orthogonal functions.
5. Fourier series, Fourier integral, signal spectrum, convolution, power and energy correlation.
6. Modulation, reasons for modulation, types of modulation.
7. Amplitude modulation systems:
   a. Comparison of amplitude modulation systems.
   b. Methods of generating and detecting AM, DBS and SSB signals.
   c. Vestigial sideband
   d. Frequency mixing and multiplexing, frequency division multiplexing
   e. Applications of AM systems.
8. Frequency modulation systems:

a. Instantaneous frequency, frequency deviation, modulation index, Bessel coefficients, significant sideband criteria

b. Bandwidth of a sinusoidally modulated FM signal, power of an FM signal, direct and indirect FM generation,

c. Various methods of FM demodulation, discriminator, phase-lock loop, limiter, pre-emphasis and de-emphasis, Stereophonic FM broadcasting

9. Noise waveforms and characteristics. Thermal noise, shot noise, noise figure and noise temperature. Cascade network, experimental determination of noise figure. Effects of noise on AM and FM systems.

10. Block diagram of a superheterodyne AM radio receiver, AM broadcast mixer, local oscillator design, intermodulation interference, adjacent channel interference, ganging, tracking error, intermediate frequency, automatic gain control (AGC), delay AGC, diode detector, volume control.

11. FM broadcast band and specification, Image frequency, block diagram of a FM radio receiver, limiter and ratio detectors, automatic frequency control, squelch circuit, FM mono and FM stereo receivers.

12. AM broadcast band and specification.

13. TV broadcast band and specification. Signal format, transmitter and receiver block diagrams of black and white TV and colour TV.

WHAT IS COMMUNICATION?

Communication is the transfer of information from a source to a receiver through a medium. Communication can be unidirectional or bidirectional.

In bidirectional communication, one is able to get a feedback from the distant end in other to know what to say next. Examples of bidirectional communication include chatting on the internet, communicating through the mobile phone or Landline, telex or telegraph, peer to peer computer network, etc.

In unidirectional communication information is transferred from the source to the receiver without the provision of receiving a feedback from the distant end. Examples of unidirectional communication include using public address systems to address a crowd, Radio/Television systems to transmit information to the public, etc.

Outlines of a Telecommunication System

In practice, telecommunications involve the conversion of messages which may be in the form of words, text or coded symbols into electrical voltage or current which varies with time and is used to carry information from one point to another. Such electrical quantities are termed as signals. These signals are then transmitted over a communication system to the receiver where they are converted back to the original form. Figure1 shows the outline of a telecommunications system.

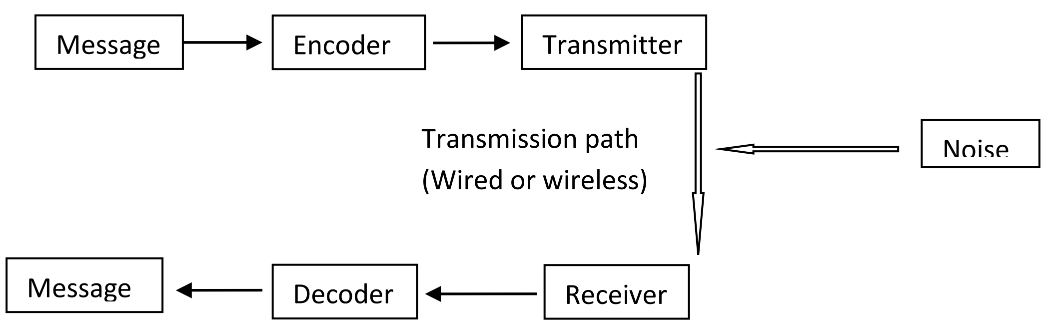


Fig.1. Outlines of a Telecommunication System

The encoder converts the message into electrical signals using appropriate transducers and feeds them to the transmitter. The signals are processed (e.g, amplified) at the transmitter and then transmitted to the receiver through a channel or transmission path which is either wireless (free space) or wired (cables). The word transmission media (medium, singular) is used to describe what is transporting telecommunication signals. There are however five basic types of medium: wire pair, coaxial cable, fiber optics, waveguides and radio. Wired communication is usually referred to as Line communication whereas wireless communication is usually referred to as Radio communication.

At the receiving end, the receiver is designed to select the desired signal, process it (demodulate it, etc,) and then deliver to the decoder an electrical signal which in all aspects resembles the signal produced by the encoder. The decoder then converts the signal into the form of the original message.

Assignment: Write out your own summary of the history of communications and submit before the next class.

Brief Historical Developments on Telecommunications

The history of telecommunications began with the use of smoke signals and drums in Africa, the Americas and parts of Asia. In the 1790s, the first fixed semaphore systems emerged in Europe; however it was not until the 1830s that electrical telecommunication systems started to appear.

This article details the history of telecommunication and the individuals who helped make telecommunication systems what they are today. The history of telecommunication is an important part of the larger history of communication.

Information transmission evolved in a series of dramatic sub-revolutions and with most of them the innovations arose in many places. Sometimes, as with the telegraph, the advent was rather sudden, other times as with Computer software, it was confused and drawn out.

Present day telecommunications arose over the last two centuries because of four great trends. These were:

1. The invention of electric signalling technology (telegraph, telephone, radio, etc.).
2. Scientific and mathematical understanding of these signalling technologies.
3. The advent of micro circuits (or micro-chips) which made the equipment small fast, reliable and very cheap.
4. The software concepts which makes possible complex algorithms.

It is interesting to observe that telecommunications from telegraphs up to television was based on science known in the 1800s. Only with data networking and the internet did communication make use of technology of the mid 1900s, namely computer software and micro circuitry.

We will now take a look at the revolutions of information transmission in the order they occurred: telegraph, telephone, radio, television, cable TV, mobility and the internet.

The Telegraph (Beacon and Optical Telegraphy)

Early telecommunications included smoke signals and drums. Talking drums were used by natives in Africa, New Guinea and South America, and smoke signals in North America and China.

Contrary to what one might think, these systems were often used to do more than merely announce the presence of a camp.

In 1792, a French engineer, Claude Chappe built the first visual telegraphy (or semaphore) system between Lille and Paris. A Semaphore line is a system of long-distance communication based on towers with moving arms (e.g Flag *semaphore system*; Railway semaphore signals for railway, etc.). It was built to transmit messages by optical signal over a distance of hundreds of kilometres without physical transport of written letters. The semaphore telegraph line usually comprised several stations each furnished with a signal mast with six cable-operated arms. The stations were equipped with telescopes that operators used to copy coded messages and forward them to the next station.

This was followed by a line from Strasbourg to Paris. In 1794, a Swedish engineer, Abraham Edelcrantz built a quite different semaphore telegraph system from Stockholm to Drottningholm. As opposed to Chappe's system which involved pulleys rotating beams of wood, Edelcrantz's system relied only upon shutters and was therefore faster. However semaphore as a communication system suffered from the need for skilled operators and expensive towers often at intervals of only ten to thirty kilometres (six to nineteen miles). As a result, the last commercial line was abandoned in 1880.

The Telegraph (Electrical Telegraphy)

A very early experiment in electrical telegraphy was an 'electrochemical' telegraph created by the German physician, anatomist and inventor Samuel Thomas von Sömmering in 1809, based on an earlier, less robust design of 1804 by Catalan polymath and scientist Francisco Salva Campillo. Both their designs employed multiple wires (up to 35) in order to visually represent almost all Latin letters and numerals. Thus, messages could be conveyed electrically up to a few kilometers (in von Sömmering's design), with each of the telegraph receiver's wires immersed in a separate glass tube of acid. An electrical current was sequentially applied by the sender through the various wires representing each digit of a message; at the recipient's end the currents electrolysed the acid in the tubes in sequence, releasing streams of hydrogen bubbles next to each associated letter or numeral. The telegraph receiver's operator would visually observe the bubbles and could then record the transmitted message, albeit at a very low baud rate. The principal disadvantage to the system was its prohibitive cost, due to having to manufacture and string-up the multiple wire circuits it employed, as opposed to the single wire (with ground return) used by later telegraphs.

The first commercial electrical telegraph was constructed in England by Sir Charles Wheatstone and Sir William Fothergill Cooke. It used the deflection of needles to represent messages and started operating over twenty-one kilometres of the Great Western Railway on 9 April 1839. Both Wheatstone and Cooke viewed their device as "an improvement to the [existing] electromagnetic telegraph" not as a new device.

On the other side of the Atlantic Ocean, Samuel Morse independently developed a version of the electrical telegraph that he unsuccessfully demonstrated on 2 September 1837. Soon after he was joined by Alfred Vail who developed the register — a telegraph terminal that integrated a logging device for recording messages to paper tape. This was demonstrated successfully over three miles (five kilometres) on 6 January 1838 and eventually over forty miles (sixty-four kilometres) between Washington, DC and Baltimore on 24th May 1844. The patented invention proved lucrative and by 1851 telegraph lines in the United States spanned over 20,000 miles (32,000 kilometres).

The first successful transatlantic telegraph cable was completed on 27th July 1866, allowing transatlantic telecommunication for the first time. Earlier transatlantic cables installed in 1857 and 1858 only operated for a few days or weeks before they failed. The international use of the telegraph has sometimes been dubbed the "Victorian Internet".

The telegraph is hardly present today but its chief significance is in how it led to other technologies and ways of social organisations.

Telephone

As with the telegraph, the scientific effect that underlay the telephone was simply the fact that electricity propagated down a wire. There was no theory of charged particle flow and no one had ideas about whether brute electricity could carry what we call a waveform today.

The electric telephone was invented in the 1870s, based on earlier work with harmonic (multi-signal) telegraphs. The first commercial telephone services were set up in 1878 and 1879 on both sides of the Atlantic in the cities of New Haven and London. Alexander Graham Bell, a Scot, held the master patent for the telephone that was needed for such services in both countries. His first demonstration of the telephone which comprises a microphone, electricity carrying a waveform and a reproducer was in Boston in 1876. He was not trying to demonstrate voice transmission but was trying to send several telegraph signals down the same wire line by interrupting different frequency tones and he noticed that something resembling a voice could be transmitted.

The first demonstration was only between two rooms after which demonstrated telephone calls over several kilometres at Brantford, Canada. The first telephone exchange was in Hamilton Canada in 1878. As early as 1878, Alexander Graham Bell was the first to conceive of a central office where wires could be connected together as desired to establish direct communication between any two places in the city which he claimed will be the outcome of the introduction of the telephone to the public.

The concept of message based charging was also developed with the commercialisation of the telephone., in which users paid for each call and was not forced to pay all at once for installation of expensive lines.

Telephone was fundamentally different from the telegraph in the following ways:

1. It carried voice instead of letters
2. It served almost everyone directly
3. Connections were set up and taken down in a switched way.

In 1880, Bell and co-inventor Charles Sumner Tainter conducted the world's first wireless telephone call via modulated lightbeams projected by photophones. The scientific principles of their invention would not be utilized for several decades, when they were first deployed in military and fiber-optic communications.

The telephone technology grew quickly from this point, with inter-city lines being built and telephone exchanges in every major city of the United States by the mid-1880s. Despite this, transatlantic voice communication remained impossible for customers until January 7, 1927 when a connection was established using radio. However no cable connection existed until TAT-1 was inaugurated on September 25, 1956 providing 36 telephone circuits.

In summary, the scientific basis of the telephone was electricity and the inventions that made it practical were simple: just an earphone, a microphone and the switching concept.

Radio

James Clerk Maxwell, the Scottish physicist, was very interested in Michael Faraday's work on electromagnetism. Faraday explained that electric and magnetic effects result from lines of force that surround conductors and magnets. Maxwell drew an analogy between the behaviour of the lines of force and the flow of a liquid, deriving equations that represent electric and magnetic effects. In 1855 he produced a paper which built on Faraday's ideas, and in 1861 developed a model for a hypothetical medium, that consisted of a fluid which could carry electric and magnetic effects. He also considered what would happen if the fluid became elastic and a charge was applied to it. This would set up a disturbance in the fluid, which would produce waves that would travel through the medium. The German physicists Friedrich Kohlrausch and Wilhelm Weber calculated that these waves would travel at the speed of light. Maxwell finally published this work in his 'Treatise on Electricity and Magnetism' in 1873.

In 1888 German physicist Heinrich Hertz made the sensational discovery of radio waves, a form of electromagnetic radiation with wavelengths too long for our eyes to see, confirming Maxwell's ideas. He devised a transmitting oscillator, which radiated radio waves, and detected them using a metal loop with a gap at one side. When the loop was placed within the transmitter's electromagnetic field, sparks were produced across the gap. This proved that electromagnetic waves could be sent out into space, and be remotely detected. These waves were known as 'Hertzian Waves' and Hertz managed to detect them across the length of his laboratory.

Italian born Guglielmo Marconi was fascinated by Hertz's discovery, and realised that if radio waves could be transmitted and detected over long distances, wireless telegraphy could be developed. He started experimenting in 1894 and set up rough aerials on opposite sides of the family garden. He managed to receive signals over a distance of 100 metres, and by the end of 1895 had extended the distance to over a mile. He approached the Italian Ministry of Posts and Telegraphs, informing them of his experiments. The Ministry was not interested and so his cousin, Henry Jameson-Davis arranged an interview with Nyilliam Preece, who was Engineer-in-Chief to the British Post Office.

Fig. Guglielmo Marconi and his family in 1933

He came to England in February 1896 and gave demonstrations in London at the General Post Office Building. His transmissions were detected 1.5 miles away, and on 2nd September at Salisbury plain the range was increased to 8 miles. In 1897 he obtained a patent for wireless telegraphy, and established the Wireless Telegraph and Signal Company at Chelmsford. The world's first radio factory was opened there in 1898. On 11th May 1897 tests were carried out to establish that contacts were possible over water. A transmitter was set up at Lavernock Point, near Penarth and the transmissions were received on the other side of the Bristol Channel at the Island of Holm, a distance of 3.5 miles. The Daily Express was the first newspaper to obtain news by wireless telegraphy in August 1898, and in December of that year communication was set up between Queen Victoria's Royal yacht, off Cowes and Osborne House. The Queen received regular bulletins on the Prince of Wales' health, by radio, from the yacht, where he was convalescing.

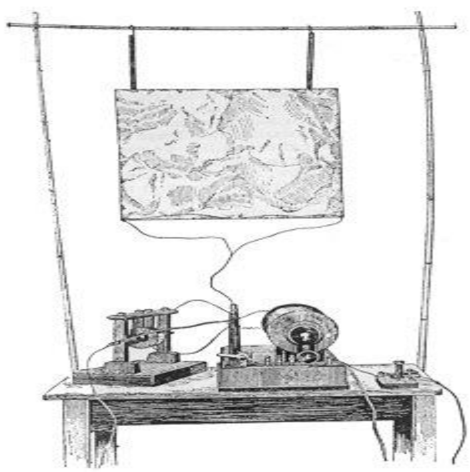
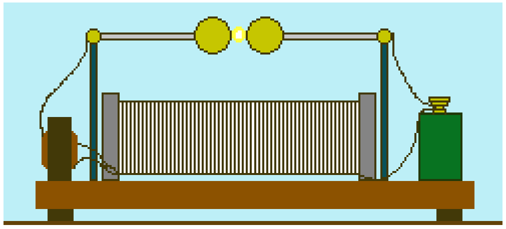


Fig. An early Marconi transmitter

Also in December of that year, wireless communication was set up between the East Goodwin light ship and the South Foeland lighthouse. On 3rd March 1899 Marconi obtained a lot of publicity when the first life was saved by wireless telegraphy, which was used to save a ship in distress in the North Sea. By the summer cross channel communication had been established and the first ocean newspaper published bulletins sent by wireless.



About this time Marconi began to develop tuned circuits for wireless transmission, so that a wireless can be tuned to a particular frequency, to remove all other transmissions except the one of interest. He patented this on 26th April 1900, under the name of 'Tuned Syntonic Telegraphy'. On Thursday 12th December 1901, Marconi and his associates succeeded in transmitting a signal across the Atlantic Ocean. He sailed to Newfoundland with G.S. Kemp and P.W. Paget, and received a transmission from Poldhu, Cornwall. The transmission was received at Signal Hill using a kite aerial. The British government and admiralty were greatly impressed and many people wanted to invest in the new technology.

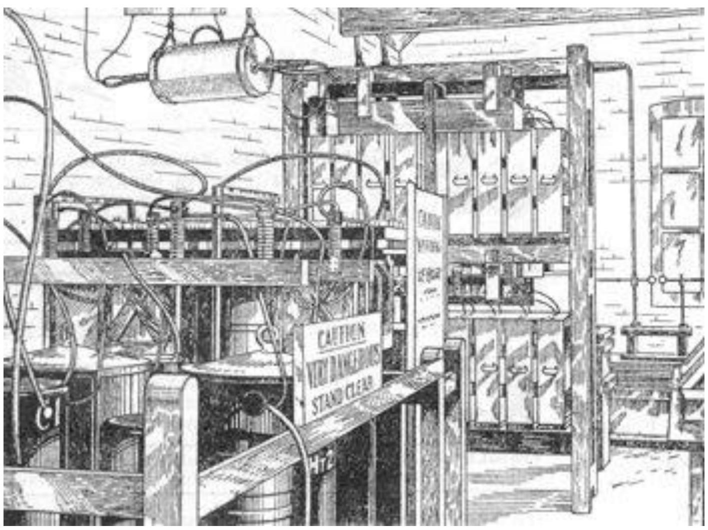


Fig. The transmitter at Poldhu

Demand grew and large numbers of ships carried the new apparatus, which saved many lives at sea. One of the most famous occasions was when the Titanic sank. Signals transmitted by its Marconi wireless summoned help and saved many lives.

Receivers at this time were mainly crystal sets, which were extremely insensitive and unselective. They were connected to a pair of headphones and required a long aerial.

In addition to Marconi, two of his contemporaries Nikola Tesla and Nathan Stufflefield took out patents for wireless radio transmitters. Nikola Tesla is now credited with being the first person to patent radio technology; the Supreme Court overturned Marconi's patent in 1943 in favor of Tesla.

Television

Television was not invented by a single inventor, instead many people working together and alone over the years, contributed to the evolution of television.

Broadcasting Pioneers: The Many Innovators Behind Television History
At the dawn of television history there were two distinct paths of technology experimented with by researchers.

Early inventors attempted to either build a mechanical television system based on the technology of Paul Nipkow's rotating disks; or they attempted to build an electronic television system using a cathode ray tube developed independently in 1907 by English inventor A.A. Campbell-Swinton and Russian scientist Boris Rosing.

Electronic television systems worked better and eventual replaced mechanical systems.

Mechanical TV systems
German, Paul Nipkow developed a rotating-disc technology to transmit pictures over wire in 1884 called the Nipkow disk. Paul Nipkow was the first person to discover television's scanning principle, in which the light intensities of small portions of an image are successively analyzed and transmitted.

John Logie Baird switched from mechanical television and became a pioneer of colour television using cathode-ray tubes. In the 1920's, John Logie Baird patented the idea of using arrays of transparent rods to transmit images for television. Baird's 30 line images were the first demonstrations of television by reflected light rather than back-lit silhouettes. John Logie Baird based his technology on Paul Nipkow's scanning disc idea and later developments in electronics. It was a system which scanned images using a rotating disk with holes arranged in a spiral pattern and were demonstrated by John Logie Baird in England and Charles Francis Jenkins in the United States in the 1920s.

Charles Jenkins invented a mechanical television system called radiovision and claimed to have transmitted the earliest moving silhouette images on June 14, 1923.

Cathode Ray Tube - Electronic Television History
Electronic television is based on the development of the cathode ray tube, which is the picture tube found in some modern TV sets. German scientist, Karl Braun invented the cathode ray tube oscilloscope (CRT) in 1897.

Russian inventor, Vladimir Zworykin invented an improved cathode-ray tube called the kinescope in 1929. The kinescope tube was sorely needed for television. Zworykin was one of the first to demonstrate a television system with all the features of modern picture tubes.

Electronic television was first successfully demonstrated in San Francisco on Sept. 7, 1927. The system was designed by Philo Taylor Farnsworth, a 21-year-old inventor who had lived in a house without electricity until he was 14. While still in high school, Farnsworth had begun to conceive of a system that could capture moving images in a form that could be coded onto radio waves and then transformed back into a picture on a screen. Boris Rosing in Russia had conducted some crude experiments in transmitting images 16 years before Farnsworth's first success.

However, Farnsworth's invention, which scanned images with a beam of electrons, is the direct ancestor of modern television. The first image he transmitted on it was a simple line. Soon he aimed his primitive camera at a dollar sign because an investor had asked, "When are we going to see some dollars in this thing, Farnsworth?"

Early television was quite primitive. All the action at that first televised baseball game had to be captured by a single camera, and the limitations of early cameras forced actors in dramas to work under impossibly hot lights, wearing black lipstick and green makeup (the cameras had trouble with the color white). The early newscasts on CBS were "chalk talks," with a newsman moving a pointer across a map of Europe, then consumed by war. The poor quality of the picture made it difficult to make out the newsman, let alone the map. World War II slowed the development of television, as companies like RCA turned their attention to military production. Television's progress was further slowed by a struggle over wavelength allocations with the new FM radio and a battle over government regulation.

It was in June of 1956, that the TV remote controller first entered the American home. The first TV remote control called "Lazy Bones," was developed in 1950 by Zenith Electronics Corporation (then known as Zenith Radio Corporation). The very first prototype for a plasma display monitor was invented in 1964 by Donald Bitzer, Gene Slottow, and Robert Willson. After mid-century the spread of coaxial cable and microwave radio relay allowed television networks to spread across even large countries.

By the 1980s politicians and government leaders were familiar enough with the workings of television to be able to exploit the medium to their own ends. This seemed particularly apparent during the presidency of Ronald Reagan, himself formerly the host of a television show (*General Electric Theater*, 1954Ð62). Reagan's skilled advisors were masters of the art of arranging flags and releasing balloons to place him in the most attractive settings. They also knew how to craft and release messages to maximize positive coverage on television newscasts. The Persian Gulf War in 1991 provided further proof of the power of television, with pictures of U.S. bombs falling on the Iraqi capital broadcast live in the United States. Both Iraqi and U.S. leaders admitted to monitoring CNN to help keep up with news of the war. However, the U.S. Defense Department, armed with lessons learned in Vietnam, succeeded in keeping most reporters well away from the action and the bloodshed. Instead, pictures were provided to television by the military of "smart" bombs deftly hitting their targets.

In the 1980s, home videocassette recorders became widely available. Viewers gained the ability to record and replay programs and, more significantly, to rent and watch movies at times of their own choosing in their own homes. Video games also became popular during this decade, particularly with the young, and the television, formally just the site of passive entertainment, became an intricate, moving, computerized game board. The number of cable networks grew throughout the 1980s and then exploded in the 1990s as improved cable technology and direct-broadcast satellite television multiplied the channels available to viewers. The number of broadcast networks increased also, with the success of the Fox network and then the arrival of the UPN and WB networks.

In 1997 the federal government gave each U.S. television broadcaster an additional channel on which to introduce high definition television, or HDTV. Initial transmissions of this high-resolution form of television, in which images appear much sharper and clearer, began in 1998. Standard television sets cannot pick up HDTV and will presumably have to be replaced or modified by 2006, when traditional, low-definition television broadcasts are scheduled to end and broadcasters are scheduled to return their original, non-HDTV channel to the government. The HDTV format approved in the United States calls for television signals to be transmitted digitally. This will allow for further convergence between computers, the Internet, and television.

Web TV

Web TV was rolled out in 1996. In 1998 it was already possible to view video on the World Wide Web and to see and search television broadcasts on a computer. As computers become more powerful, they should be able to handle video as easily as they now handle text. The television schedule may eventually be replaced by a system in which viewers are able to watch digitally stored and distributed programs or segments of programs whenever they want. Such technological changes, including the spread of new cable networks, have been arriving slower in most other countries than in the United States. Indeed, according to one survey, it was only in the 1990s that the spread of television transmitters, television sets, and electricity made it possible for half of the individuals in the world to watch television. However, television's attraction globally is strong. Those human beings who have a television set watch it, by one estimate, for an average of two-and-a-half hours a day.

Mobility

There is nothing new about mobile Radio. Radio at its beginning was a wireless radio service. By 1922, there were experiments with fire and police mobile radio. By the mid 1930s, mobile radio broadcast receivers were appearing in cars.

Mobile telephony for the general public slowly grew and became less expensive and clumsy to use. Eventually a service emerged that would be recognisable today, except that the handheld telephone was considerably larger. These service used analogue frequency modulation worked at either 400 or 800MHz and were cellular meaning that many transmitters covered small patches of land and the user was handed off from one to the other as he moved around.

In the US, the system was called Advanced mobile phone system (AMPS) and it was one of the last contributions of the old Bell System. Very similar systems referred to as the first Generation of mobile systems appeared in the rest of the world at the same time.

The explosion of the public interest in mobile telephony came with the digital second generation systems. The cutting edge was in Europe and the system was the Global system for Mobile (GSM) that we still use today. First test were carried out in 1986 and GSM began to be installed in the mid 1990s. Public use of mobile telephony grew rapidly and dramatically with the introduction of GSM for several reasons:

1. In part, digital telephony was inherently more efficient.
2. It adapted easily to other services such as e-mail.
3. High frequency radio and digital circuitry were dropping rapidly in price and size in any case.
4. It offered for the first time a more subtle advantage called roaming because it was unified over many countries and now over much of the world. Its software procedures also made identification of visiting telephones easy thus people could roam over much of the earth and still use the same mobile phone.

All these factors combined to make a very attractive product and public use of the mobile phone doubled in the first years of the second generation technologies. Today in some countries such as Nigeria, there are more mobile users than fixed users.

Today, mobile phone based technologies have progressed to third generations having higher bandwidth (so that it can handle music with video) and a combining of mobility with internet access. The third generation is thus a fusion of a data base (The Internet) with communication to it.

Fourth generation technologies are already appearing in the market. They offer even higher bandwidths than 3rd Generation technologies thus they are excellently for live video streaming applications.

Communications Satellites: Making the Global Village Possible

In 500 years, when humankind looks back at the dawn of space travel, Apollo's landing on the Moon in 1969 may be the only event remembered. At the same time, however, Lyndon B. Johnson, himself an avid promoter of the space program, felt that reconnaissance satellites alone justified every penny spent on space. Weather forecasting has undergone a revolution because of the availability of pictures from geostationary meteorological satellites--pictures we see every day on television. All of these are important aspects of the space age, but satellite communications has probably had more effect than any of the rest on the average person. Satellite communications is also the only truly commercial space technology- -generating billions of dollars annually in sales of products and services.

The Billion Dollar Technology

In fall of 1945 an RAF electronics officer and member of the British Interplanetary Society, Arthur C. Clarke, wrote a short article in *Wireless World* that described the use of manned satellites in 24-hour orbits high above the world's land masses to distribute television programs. His article apparently had little lasting effect in spite of Clarke's repeating the story in his 1951/52 *The Exploration of Space*. Perhaps the first person to carefully evaluate the various technical options in satellite communications *and* evaluate the financial prospects was John R. Pierce of AT&T's Bell Telephone Laboratories who, in a 1954 speech and 1955 article, elaborated the utility of a communications "mirror" in space, a medium-orbit "repeater" and a 24-hour-orbit "repeater." In comparing the communications capacity of a satellite, which he estimated at 1,000 simultaneous telephone calls, and the communications capacity of the first trans-atlantic telephone cable (TAT-1), which could carry 36 simultaneous telephone calls at a cost of 30-50 million dollars, Pierce wondered if a satellite would be worth a billion dollars.

After the 1957 launch of Sputnik I, many considered the benefits, profits, and prestige associated with satellite communications. Because of Congressional fears of "duplication," NASA confined itself to experiments with "mirrors" or "passive" communications satellites (ECHO), while the Department of Defense was responsible for "repeater" or "active" satellites which amplify the received signal at the satellite--providing much higher quality communications. In 1960 AT&T filed with the Federal Communications Commission (FCC) for permission to launch an experimental communications satellite with a view to rapidly implementing an operational system. The U.S. government reacted with surprise-- there was no policy in place to help execute the many decisions related to the AT&T proposal. By the middle of 1961, NASA had awarded a competitive contract to RCA to build a medium-orbit (4,000 miles high) active communication satellite (RELAY); AT&T was building its own medium-orbit satellite (TELSTAR) which NASA would launch on a cost-reimbursable basis; and NASA had awarded a sole- source contract to Hughes Aircraft Company to build a 24-hour (20,000 mile high) satellite (SYNCOM). The military program, ADVENT, was cancelled a year later due to complexity of the spacecraft, delay in launcher availability, and cost over-runs.

By 1964, two TELSTARs, two RELAYs, and two SYNCOMs had operated successfully in space. This timing was fortunate because the Communications Satellite Corporation (COMSAT), formed as a result of the Communications Satellite Act of 1962, was in the process of contracting for their first satellite. COMSAT's initial capitalization of 200 million dollars was considered sufficient to build a system of dozens of medium-orbit satellites. For a variety of reasons, including costs, COMSAT ultimately chose to reject the joint AT&T/RCA offer of a medium-orbit satellite incorporating the best of TELSTAR and RELAY. They chose the 24-hour-orbit (geosynchronous) satellite offered by Hughes Aircraft Company for their first two systems and a TRW geosynchronous satellite for their third system. On April 6, 1965 COMSAT's first satellite, EARLY BIRD, was launched from Cape Canaveral. Global satellite communications had begun.

The Global Village: International Communications

Some glimpses of the Global Village had already been provided during experiments with TELSTAR, RELAY, and SYNCOM. These had included televising parts of the 1964 Tokyo Olympics. Although COMSAT and the initial launch vehicles and satellites were American, other countries had been involved from the beginning. AT&T had initially negotiated with its European telephone cable "partners" to build earth stations for TELSTAR experimentation. NASA had expanded these negotiations to include RELAY and SYNCOM experimentation. By the time

EARLY BIRD was launched, communications earth stations already existed in the United Kingdom, France, Germany, Italy, Brazil, and Japan. Further negotiations in 1963 and 1964 resulted in a new international organization, which would ultimately assume ownership of the satellites and responsibility for management of the global system. On August 20, 1964, agreements were signed which created the International Telecommunications Satellite Organization (INTELSAT).

By the end of 1965, EARLY BIRD had provided 150 telephone "half- circuits" and 80 hours of television service. The INTELSAT II series was a slightly more capable and longer-lived version of EARLY BIRD. Much of the early use of the COMSAT/INTELSAT system was to provide circuits for the NASA Communications Network (NASCOM). The INTELSAT III series was the first to provide Indian Ocean coverage to complete the global network. This coverage was completed just days before one half billion people watched APOLLO 11 land on the moon on July 20, 1969.

From a few hundred telephone circuits and a handful of members in 1965, INTELSAT has grown to a present-day system with more members than the United Nations and the capability of providing hundreds of thousands of telephone circuits. Cost to carriers per circuit has gone from almost $100,000 to a few thousand dollars. Cost to consumers has gone from over $10 per minute to less than $1 per minute. If the effects of inflation are included, this is a tremendous decrease! INTELSAT provides services to the entire globe, not just the industrialized nations.

Hello Guam: Domestic Communications

In 1965, ABC proposed a domestic satellite system to distribute television signals. The proposal sank into temporary oblivion, but in 1972 TELESAT CANADA launched the first domestic communications satellite, ANIK, to serve the vast Canadian continental area. RCA promptly leased circuits on the Canadian satellite until they could launch their own satellite. The first U.S. domestic communications satellite was Western Union's WESTAR I, launched on April 13, 1974. In December of the following year RCA launched their RCA SATCOM F- 1. In early 1976 AT&T and COMSAT launched the first of the COMSTAR series. These satellites were used for voice and data, but very quickly television became a major user. By the end of 1976 there were 120 transponders available over the U.S., each capable of providing 1500 telephone channels or one TV channel. Very quickly the "movie channels" and "super stations" were available to most Americans. The dramatic growth in cable TV would not have been possible without an inexpensive method of distributing video.

The ensuing two decades have seen some changes: Western Union is no more; Hughes is now a satellite operator as well as a manufacturer; AT&T is still a satellite operator, but no longer in partnership with COMSAT; GTE, originally teaming with Hughes in the early 1960s to build and operate a global system is now a major domestic satellite operator. Television still dominates domestic satellite communications, but data has grown tremendously with the advent of very small aperture terminals (VSATs). Small antennas, whether TV-Receive Only (TVRO) or VSAT are common-place sight all over the country.

New Technology

The first major geosynchronous satellite project was the Defense Department's ADVENT communications satellite. It was three-axis stabilized rather than spinning. It had an antenna that directed its radio energy at the earth. It was rather sophisticated and heavy. At 500-1000 pounds it could only be launched by the ATLAS- CENTAUR launch vehicle. ADVENT never flew, primarily because the CENTAUR stage was not fully reliable until 1968, but also because of problems with the satellite. When the program was cancelled in 1962 it was seen as the death knell for geosynchronous satellites, three-axis stabilization, the ATLAS-CENTAUR, and complex communications satellites generally. Geosynchronous satellites became a reality in 1963, and became the only choice in 1965. The other ADVENT characteristics also became commonplace in the years to follow.

In the early 1960s, converted intercontinental ballistic missiles (ICBMs) and intermediate range ballistic missiles (IRBMs) were used as launch vehicles. These all had a common problem: they were designed to deliver an object to the earth's surface, not to place an object in orbit. Upper stages had to be designed to provide a delta-Vee (velocity change) at apogee to circularize the orbit. The DELTA launch vehicles, which placed all of the early communications satellites in orbit, were THOR IRBMs that used the VANGUARD upper stage to provide this delta-Vee. It was recognized that the DELTA was relatively small and a project to develop CENTAUR, a high-energy upper stage for the ATLAS ICBM, was begun. ATLAS-CENTAUR became reliable in 1968 and the fourth generation of INTELSAT satellites used this launch vehicle. The fifth generation used ATLAS-CENTAUR and a new launch-vehicle, the European ARIANE. Since that time other entries, including the Russian PROTON launch vehicle and the Chinese LONG MARCH have entered the market. All are capable of launching satellites almost thirty times the weight of EARLY BIRD.

In the mid-1970s several satellites were built using three-axis stabilization. They were more complex than the spinners, but they provided more despun surface to mount antennas and they made it possible to deploy very large solar arrays. The greater the mass and power, the greater the advantage of three-axis stabilization appears to be. Perhaps the surest indication of the success of this form of stabilization was the switch of Hughes, closely identified with spinning satellites, to this form of stabilization in the early 1990s. The latest products from the manufacturers of SYNCOM look quite similar to the discredited ADVENT design of the late 1950s.

Much of the technology for communications satellites existed in 1960, but would be improved with time. The basic communications component of the satellite was the traveling-wave-tube (TWT). These had been invented in England by Rudoph Kompfner, but they had been perfected at Bell Labs by Kompfner and J. R. Pierce. All three early satellites used TWTs built by a Bell Labs alumnus. These early tubes had power outputs as low as 1 watt. Higher- power (50-300 watts) TWTs are available today for standard satellite services and for direct-broadcast applications. An even more important improvement was the use of high-gain antennas. Focusing the energy from a 1-watt transmitter on the surface of the earth is equivalent to having a 100-watt transmitter radiating in all directions. Focusing this energy on the Eastern U.S. is like having a 1000-watt transmitter radiating in all directions. The principal effect of this increase in actual and effective power is that earth stations are no longer 100-foot dish reflectors with cryogenically-cooled maser amplifiers costing as much as $10 million (1960 dollars) to build. Antennas for normal satellite services are typically 15-foot dish reflectors costing $30,000 (1990 dollars). Direct-broadcast antennas will be only a foot in diameter and cost a few hundred dollars.

Mobile Services

In February of 1976 COMSAT launched a new kind of satellite, MARISAT, to provide mobile services to the United States Navy and other maritime customers. In the early 1980s the Europeans launched the MARECS series to provide the same services. In 1979 the UN International Maritime Organization sponsored the establishment of the International Maritime Satellite Organization (INMARSAT) in a manner similar to INTELSAT. INMARSAT initially leased the MARISAT and MARECS satellite transponders, but in October of 1990 it launched the first of its own satellites, INMARSAT II F-1. The third generation, INMARSAT III, has already been launched.

An aeronautical satellite was proposed in the mid-1970s. A contract was awarded to General Electric to build the satellite, but it was cancelled--INMARSAT now provides this service. Although INMARSAT was initially conceived as a method of providing telephone service and traffic-monitoring services on ships at sea, it has provided much more. The journalist with a briefcase phone has been ubiquitous for some time, but the Gulf War brought this technology to the public eye.

The United States and Canada discussed a North American Mobile Satellite for some time. Later the first MSAT satellite, in which AMSC (U.S.) and TMI (Canada) cooperate, would be launched providing mobile telephone service via satellite to all of North America.

Competition

In 1965, when EARLY BIRD was launched, the satellite provided almost 10 times the capacity of the submarine telephone cables for almost 1/10th the price. This price-differential was maintained until the laying of TAT-8 in the late 1980s. TAT-8 was the first fiber-optic cable laid across the Atlantic. Satellites are still competitive with cable for point-to-point communications, but the future advantage may lie with fiber-optic cable. Satellites still maintain two advantages over cable: they are more reliable and they can be used point-to-multi-point (broadcasting).

Cellular telphone systems have risen as challenges to all other types of telephony. It is possible to place a cellular system in a developing country at a very reasonable price. Long-distance calls require some other technology, but this can be either satellites or fiber-optic cable.

The LEO Systems

Cellular telephony has brought us a new technological "system"-- the personal communications system (PCS). In the fully developed PCS, the individual would carry his telephone with him. This telephone could be used for voice or data and would be usable anywhere. Several companies have committed themselves to providing a version of this system using satellites in low earth orbits

(LEO). These orbits are significantly lower than the TELSTAR/RELAY orbits of the early 1960s. The early "low-orbit" satellites were in elliptical orbits that took them through the lower van Allen radiation belt. The new systems will be in orbits at about 500 miles, below the belt.

The most ambitious of these LEO systems is Iridium, sponsored by Motorola. Iridium plans to launch 66 satellite into polar orbit at altitudes of about 400 miles. Each of six orbital planes, separated by 30 degrees around the equator, will contain eleven satellites. Iridium originally planned to have 77 satellites-- hence its name. Element 66 has the less pleasant name Dysprosium. Iridium expects to be providing communications services to hand- held telephones in 1998. The total cost of the Iridium system is well in excess of three billion dollars.

In addition to the "Big LEOS" such as Iridium and Globalstar, there are several "little LEOs." These companies plan to offer more limited services, typically data and radio determination. Typical of these is ORBCOM which has already launched an experimental satellite and expects to offer limited service in the very near future.

A Selective Communications Satellite Chronology

- 1945 Arthur C. Clarke Article: "Extra-Terrestrial Relays"
- 1955 John R. Pierce Article: "Orbital Radio Relays"
- 1956 First Trans-Atlantic Telephone Cable: TAT-1
- 1957 Sputnik: Russia launches the first earth satellite.
- 1960 1st Successful DELTA Launch Vehicle
- 1960 AT&T applies to FCC for experimental satellite communications license
- 1961 Formal start of TELSTAR, RELAY, and SYNCOM Programs
- 1962 TELSTAR and RELAY launched
- 1962 Communications Satellite Act (U.S.)
- 1963 SYNCOM launched
- 1964 INTELSAT formed
- 1965 COMSAT's EARLY BIRD: 1st commercial communications satellite
- 1969 INTELSAT-III series provides global coverage
- 1972 ANIK: 1st Domestic Communications Satellite (Canada)
- 1974 WESTAR: 1st U.S. Domestic Communications Satellite
- 1975 INTELSAT-IVA: 1st use of dual-polarization
- 1975 RCA SATCOM: 1st operational body-stabilized comm. satellite
- 1976 MARISAT: 1st mobile communications satellite
- 1976 PALAPA: 3rd country (Indonesia) to launch domestic comm. satellite
- 1979 INMARSAT formed.
- 1988 TAT-8: 1st Fiber-Optic Trans-Atlantic telephone cable

Facsimile (FAX)?

It refers to the transmission of photographs, drawings, maps, and written or printed words by electric signals. Facsimile (Fax) is a method of encoding data, transmitting it over the telephone lines or radio broadcast, and receiving hard (text) copy, line drawings, or photographs. Light waves reflected from an image are converted into electric signals, transmitted by wire or radio to a distant receiver, and reconstituted on paper or film into a copy of the original.

Facsimile is used by news services to send news and photos to newspapers and television stations, by banks, airlines, and railroads to transmit the content of documents, and by many other businesses as an aid in data handling and record keeping.

Facsimile systems involve optical scanning, signal encoding, modulation, signal transmission, demodulation, decoding, and copy making.

Scanning

Scanning is done in a manner similar to that used in television. An original, a photo for example, is illuminated and systematically examined in small adjacent areas called pixels (picture elements). Light reflected from each pixel is converted into electric current by an electronic device, a photocell, photodiode, or charge-coupled device (CCD).

A single such device may be used to cover one pixel after another in a row, row after row from top to bottom until the entire image has been translated into electric impulses. This is rectilinear scanning. Scanning may also be done a row at a time by a battery of devices; this is array scanning.

We owe development of fax to a Scottish inventor, Alexander Bain, who was granted a patent for his creation back in 1843. Bain's original concept is still the basis for modern facsimile machines.

Fax was invented in 1842 by Alexander Bain, a Scottish clockmaker, who used clock mechanisms to transfer an image from one sheet of electrically conductive paper to another. He invented a machine capable of receiving signals from a telegraph wire and translating them into images on paper.

Bain's fax transmitter was designed to scan a two-dimensional surface (Bain proposed metal type as the surface) by means of a stylus mounted on a pendulum.

Analog Telephone Facsimile - Digital Facsimile

Between 1920 and 1923 the American Telephone & Telegraph Company (AT&T) worked on telephone facsimile technology, and in 1924 the telephotography machine was used to send pictures from political conventions in Cleveland, Ohio, and Chicago to New York City for publication in newspapers.

Computer Networks and the Internet

On September 11, 1940, George Stibitz was able to transmit problems using teletype to his Complex Number Calculator in New York and receive the computed results back at Dartmouth

College in New Hampshire. This configuration of a centralized computer or mainframe with remote dumb terminals remained popular throughout the 1950s. However it was not until the 1960s that researchers started to investigate packet switching — a technology that would allow chunks of data to be sent to different computers without first passing through a centralized mainframe. A four-node network emerged on December 5, 1969 between the University of California, Los Angeles, the Stanford Research Institute, the University of Utah and the University of California, Santa Barbara. This network would become ARPANET, which by 1981 would consist of 213 nodes. In June 1973, the first non-US node was added to the network belonging to Norway's NORSAR project. This was shortly followed by a node in London.

ARPANET's development centred around the Request for Comment process and on April 7, 1969, RFC 1 was published. This process is important because ARPANET would eventually merge with other networks to form the Internet and many of the protocols the Internet relies upon today were specified through this process. In September 1981, RFC 791 introduced the Internet Protocol v4 (IPv4) and RFC 793 introduced the Transmission Control Protocol (TCP) — thus creating the TCP/IP protocol that much of the Internet relies upon today. A more relaxed transport protocol that, unlike TCP, did not guarantee the orderly delivery of packets called the User Datagram Protocol (UDP) was submitted on 28 August 1980 as RFC 768. An e-mail protocol, SMTP, was introduced in August 1982 by RFC 821 and http://1.0 a protocol that would make the hyperlinked Internet possible was introduced on May 1996 by RFC 1945.

However not all important developments were made through the Request for Comment process. Two popular link protocols for local area networks (LANs) also appeared in the 1970s. A patent for the Token Ring protocol was filed by Olof Söderblom on October 29, 1974 and a paper on the Ethernet protocol was published by Robert Metcalfe and David Boggs in the July 1976 issue of *Communications of the ACM*.

Internet access became widespread late in the 19^{th} century, using the old telephone and television networks.

ELECTROMAGNETIC SPECTRUM

The entire range of electromagnetic radiation frequencies is called the electromagnetic spectrum. The frequency range suitable for radio transmission (the radio spectrum) extends from 10 kilo Hertz to 300,000 mega Hertz. The electromagnetic spectrum is divided into a number of bands, as shown in Table 1.

Below the radio spectrum, but overlapping it, is the audio frequency band, extending from 20 to 20,000 hertz. Above the radio spectrum are heat and infrared, the visible spectrum (light in its various colours), ultraviolet, Xrays, gamma rays, and cosmic rays. Waves shorter than 30 centimeters are usually called microwaves.

Table1. Electromagnetic spectrum.

Band	**Abbreviation**	**Range of frequency**	**Range of wavelength**	**Area of Use**
Audio frequency	AF	20 to 20,000 Hz	15,000,000 to 15,000 m	
Radio frequency	RF	10 kHz to 300,000 MHz	30,000m to 0.1 cm	Used for transmission of data, via modulation as in Television, Mobile phones, wireless networking, Amateur radio broadcasting etc.
Very low frequency	VLF	10kHz to 30 kHz	30,000 to 10,000 m	Used for digital radio communication. Efficient during magnetic storm and for transmission of signals below surface of the sea
Low frequency	LF	30kHz to 300 kHz	10,000 to 1,000 m	Most suitable within groundwave distance of the transmitter and useful for radio direction finding and time dissemination.
Medium frequency	MF	300kHz to 3,000 kHz	1,000 to 100 m	Good for skywave reception, used in standard broadcast band for commercial stations
High frequency	HF	3MHz to 30 MHz	100 to 10 m	Used for ship to ship and ship to shore communication
Very high frequency	VHF	30MHz to 300 MHz	10 to 1 m	Used for direct wave plus ground reflected wave. This band is used much for communication
Ultra high frequency	UHF	300MHz to 3,000 MHz	100 to 10 cm	Can be used for ground waves and ground reflected waves. Also Used for ship to ship and ship to shore communication with sharp directive antennas
Super high frequency	SHF	3,000MHz to 30,000 MHz	10 to 1 cm	Used for Marine Navigation Radar
Extremely high frequency	EHF	30,000MHz to 300,000 MHz	1 to 0.1 cm	Used for direct and ground reflected waves

Heat and infrared*		10^6 to 3.l9X10^8 MHz	0.03 to 7.6X10^{-5} cm	
Visible spectrum*		3.9X10^8 to 7.9X10^8 MHz	7.6 X 10^{-5} to 3.8 X 10^{-5} cm	
Ultraviolet*		7.9X10^8 to 2.3X10^{10} MHz	3.8X10^{-5} to 1.3X10^{-6} cm	
X-rays*		2..0X109 to 3.0X10^{13} MHz	1.5X10^{-5} to 1.0X10^{-9} cm	
Gamma rays*		2.3X10^{12} to 3.0X10^{14} MHz	1.3X10^{-8} to 1.0X10^{-10} cm	
Cosmic rays*		>4.8X10^{15} MHz	<6.2X10^{-12} cm	

* Values approximate.

Frequency is an important consideration in radio wave propagation. The following summary indicates the principal effects associated with the various frequency bands, starting with the lowest and progressing to the highest usable radio frequency.

Very Low Frequency (VLF, 10 to 30 kHz) (λ=30,000 to 10,000 m):

The VLF signals propagate between the bounds of the ionosphere and the earth and are thus guided around the curvature of the earth to great distances with low attenuation and excellent stability. Diffraction is maximum in this band. Because of the long wavelength, large antennas are needed, and even these are inefficient, permitting radiation of relatively small amounts of power. Magnetic storms have little effect upon transmission because of the efficiency of the "earth-ionosphere waveguide." During such storms, VLF signals may constitute the only source of radio communication over great distances. However, interference from atmospheric noise may be troublesome. Signals may be received from below the surface of the sea.

Low Frequency (LF, 30 to 300 kHz)(λ=10,000 to 1,000 m):

As frequency is increased to the LF band and diffraction decreases, there is greater attenuation with distance, and range for a given power output falls off rapidly. However, this is partly offset by more efficient transmitting antennas. LF signals are most stable within groundwave distance of the transmitter. A wider bandwidth permits pulsed signals at 100 kHz. This allows separation of the stable groundwave pulse from the variable skywave pulse up to 1,500 km, and up to 2,000 km for overwater paths. This band is also useful for radio direction finding and time dissemination.

Medium Frequency (MF, 300 to 3,000 kHz)(λ= 1,000 to 100 m):

Groundwaves provide dependable service, but the range for a given power is reduced greatly. This range varies from about 643.737Km at the lower portion of the band to about 24.1401Km at the upper end for a transmitted signal of 1kilowatt. These values are influenced, however, by the power of the transmitter, the directivity and efficiency of the antenna, and the nature of the terrain over which signals travel. Elevating the antenna to obtain direct waves may improve the transmission. At the lower frequencies of the band, skywaves are available both day and night. As the frequency is increased, ionospheric absorption increases to a maximum at about 1,400 kHz. At higher frequencies the absorption decreases, permitting increased use of skywaves. Since the ionosphere changes with the hour, season, and sunspot cycle, the reliability of skywave signals is variable. By careful selection of frequency, ranges of as much as 12874.752Km with 1 kilowatt of transmitted power are possible, using multihop signals. However, the frequency selection is critical. If it is too high, the signals penetrate the ionosphere and are lost in space. If it is too low, signals are too weak. In general, skywave reception is equally good by day or night, but lower frequencies are needed at night. The standard broadcast band for commercial stations (535 to 1,605 kHz) is in the MF band.

High Frequency (HF, 3 to 30 MHz)(λ= 100 to 10 m):

As with higher medium frequencies, the groundwave range of HF signals is limited to a few miles, but the elevation of the antenna may increase the direct-wave distance of transmission. Also, the height of the antenna does have an important effect upon skywave transmission because the antenna has an "image" within the conducting earth. The distance between antenna and image is related to the height of the antenna, and this distance is as critical as the distance between elements of an antenna system. Maximum usable frequencies fall generally within the HF band. By day this may be 10 to 30 MHz, but during the night it may drop to 8 to 10 MHz. The HF band is widely used for ship-to-ship and ship-to-shore communication.

Very High Frequency (VHF, 30 to 300 MHz)(λ=10 to 1 m):

Communication is limited primarily to the direct wave, or the direct wave plus a ground-reflected wave. Elevating the antenna to increase the distance at which direct waves can be used results in increased distance of reception, even though some wave interference between direct and ground-reflected waves is present. Diffraction is much less than with lower frequencies, but it is most evident when signals cross sharp mountain peaks or ridges. Under suitable conditions, reflections from the ionosphere are sufficiently strong to be useful, but generally they are unavailable. There is relatively little interference from atmospheric noise in this band. Reasonably efficient directional antennas are possible with VHF. The VHF band is much used for communication.

Ultra High Frequency (UHF, 300 to 3,000 MHz)(λ=100 to 10 cm):

Skywaves are not used in the UHF band because the ionosphere is not sufficiently dense to reflect the waves, which pass through it into space. Groundwaves and ground-reflected waves are used, although there is some wave interference. Diffraction is negligible, but the radio horizon extends about 15% beyond the visible horizon, due principally to refraction. Reception of UHF signals is virtually free from fading and interference by atmospheric noise.

Sharply directive antennas can be produced for transmission in this band, which is widely used for ship-to-ship and ship-to-shore communication.

Super High Frequency (SHF, 3,000 to 30,000 MHz)(λ=10 to 1 cm):

In the SHF band, also known as the microwave or as the centimeter wave band, there are no skywaves, transmission being entirely by direct and ground-reflected waves. Diffraction and interference by atmospheric noise are virtually non-existent. Highly efficient, sharply directive antennas can be produced. Thus, transmission in this band is similar to that of UHF, but with the effects of shorter waves being greater. Reflection by clouds, water droplets, dust particles, etc., increases, causing greater scattering, increased wave interference, and fading. The SHF band is used for marine navigational radar.

Extremely High Frequency (EHF, 30,000 to 300,000MHz)(λ=1 to 0.1 cm):

The effects of shorter waves are more pronounced in the EHF band, transmission being free from wave interference, diffraction, fading, and interference by atmospheric noise. Only direct and ground-reflected waves are available. Scattering and absorption in the atmosphere are pronounced and may produce an upper limit to the frequency useful in radio communication.

TYPES OF SIGNAL TRANSMISSION

The electrical signals produced by encoders are of two types, namely analogue signals and digital signals. These two types of signals results in two types of signal transmission:

1. **Analogue signal transmission** which is used in a communication that involves the transmission of analogue signals from the transmitter to the receiver. Analogue signals continuously vary with time. They are sinusoidal in nature and usually have harmonics. They represent the variations of physical quantities such as sound, pressure, temperature, etc. and are represented by voltage waveforms that have different amplitudes at different instants of time. Examples of analogue signal transmissions are voice transmission through a telephone line, Radio and TV broadcast to the general public. Sometimes analogue signals are first converted into digital signals before being transmitted.

2. **Digital signal transmission** which is used in a communication that involves the transmission of digital signals from the transmitter to the receiver. Digital signals are not continuous. They are made up of pulses which occur at discrete intervals of time. The pulses may occur singly at a definite period of time or as a coded group. These signals play a very important role in the transmission and reception of coded messages. Examples of digital signals are

    a. **Telegraph signal** which is generated by a telegraph and teleprinter which are the most common instruments being used to transmit written text in the form of coded signals.

    b. **Radar signal** which is generated by a radar (a device being used to find out the location of distant objects in terms of location and bearing by transmitting a short period signal and beaming it to the location of the target. The reflected signal is picked up by the radar

    c. **Data signals** which are generated by several devices and are required to transmit data from one place to another. The data to be transmitted are converted into electrical pulses before transmission is done.

SIGNAL SPECTRUM

Plotting the amplitude of a signal at various instants of time is used to represent the signal in the ***Time domain*.** Plotting the amplitudes of the different frequency components is termed the ***frequency domain*** representation. This plot gives the spectral component amplitudes of the signal against frequency.

Analogue signals when analysed are found to comprise of certain fundamental frequencies and their harmonics. They occupy only a small portion of the frequency spectrum which is termed as the ***Discrete spectrum.***

The analysis of digital signals on the other hand gives an infinite number of frequencies. Such a spectrum is termed as ***Continuous spectrum.***

SIGNAL ANALYSIS

Signals are single valued functions of time (t) and are of complex nature. No matter how complex a signal wave form may be, it comprises of one or more sine and / or cosine functions. Assume that we have a square wave given by the expression 1.1

$$f(x) = \begin{Bmatrix} 1(0 < t < \pi) \\ -1(\pi > t < 2\pi) \end{Bmatrix} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1.1$$

The signal is represented by the figure 1.

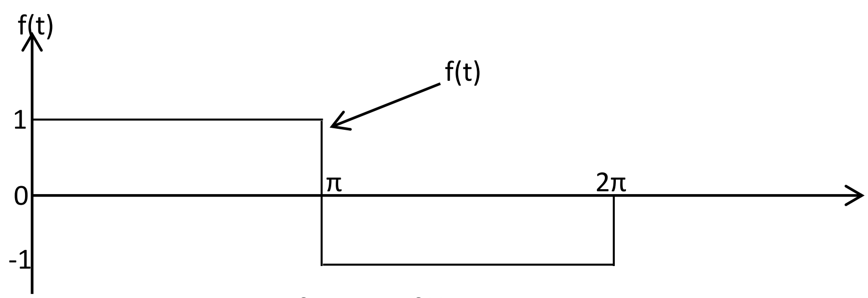


Fig. 1a A square wave function f(t).

Let us try to see how a sine function of the same time period can be used to represent this square wave form.

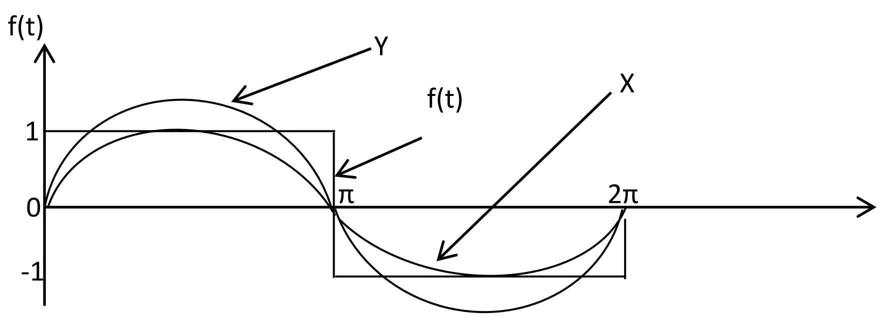


Fig. 2a. A square wave function f(t) and two sine wave functions.

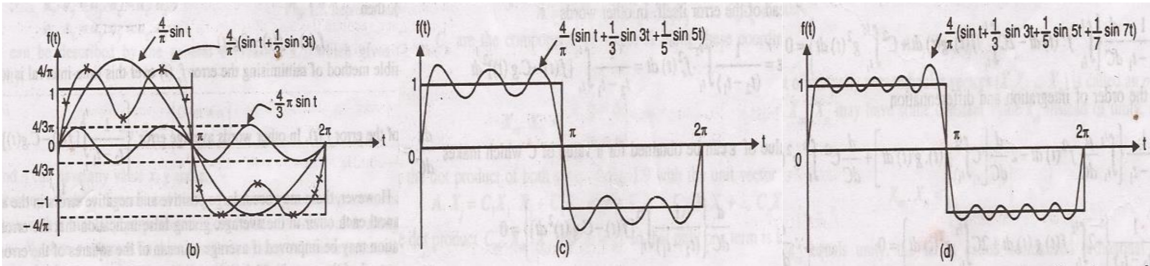


Fig2(b), (c), (d): A square wave function approximated by a sine wave function.

In Fig2a, we introduced a sine function marked X, having the same peak magnitude as the square wave f(t), hence its magnitude is equal to the square wave only at the peak point indicating that it is a very poor representation or approximation of the square wave.

If the magnitude of the sine wave X, is increased as shown by the sine wave Y, its magnitude becomes equal to the sine wave magnitude at two points. This provides an approximation slightly better than the first curve even though it is still a very poor approximation.

In figure 2b, another sine wave component is added to improve the approximation. This component has a frequency thrice the first component. It is easily seen that this provides a better approximation. The approximated wave approaches more closely to the square wave when more sine wave components are added. As shown in Figure 2c-2d.

The graphical method of approximating one function to another gives a clear understanding but is difficult to use in practice, hence it is always necessary to use analytical methods of approximating the square wave function with a sine wave function.

Consider two signals f(t) and g(t). Assume that f(t) is to be approximated in terms of g(t) over the interval (t_1-t_2). This approximation may be written as;

$f(t) \cong C.g(t)$ for ($t_1<t<t_2$)…………………………………………………..2

Where C is a constant and has a value such that error between the actual function and the approximated function is minimum over the time-interval considered. If the error function if denoted as $f_e(t) = f(t) - C.g(t)$ … .3

One possible method of minimizing the error $f_e(t)$ over this time-interval is to minimise the average value of the error $f_e(t)$. In other words average error $\left\{\frac{1}{t_2-t_1}\int_{t_1}^{t_2}[f(t) - C.g(t)]dt\right\}$ should be kept minimum. However, there may occur large positive and negative errors in the approximation which cancel each other in the average giving false indication that the error is minimum.

The situation may be improved if average or mean of the squares of the error denoted by ε is minimised, instead of the error itself. In other words, ε= $\left\{\frac{1}{t_2-t_1}\int_{t_1}^{t_2} f_e^2(t)dt\right\} = \left\{\frac{1}{t_2-t_1}\int_{t_1}^{t_2}[f(t)-C.g(t)]^2dt\right\}$ 4

Minimum value of ε can be obtained for a value of C which makes $\frac{d\varepsilon}{dC}=0$

Or $\frac{d}{dC}\left\{\frac{1}{t_2-t_1}\int_{t_1}^{t_2}[f(t)-C.g(t)]^2dt\right\}=0$

Therefore, $\frac{1}{t_2-t_1}\frac{d}{dC}\left\{\int_{t_1}^{t_2} f^2(t)dt - 2C\int_{t_1}^{t_2} f(t).g(t)dt + C^2\int_{t_1}^{t_2} g^2(t)\,dt\right\}=0$

Interchanging the order of integration and differentiation

$$\frac{1}{t_2-t_1}\left\{\int_{t_1}^{t_2}\frac{d}{dC}f^2(t)dt - 2\frac{d}{dC}\left[C\int_{t_1}^{t_2}f(t).g(t)dt\right] + \frac{d}{dC}C^2\int_{t_1}^{t_2} g^2(t)\,dt\right\}=0$$

But $\frac{d}{dC}f^2(t)dt=0$,

Therefore the expression becomes,

$$\frac{1}{t_2-t_1}\left\{-2\int_{t_1}^{t_2} f(t).g(t)dt + 2C\int_{t_1}^{t_2} g^2(t)\,dt\right\}=0$$

Therefore C= $\frac{\int_{t_1}^{t_2} f(t).g(t)dt}{\int_{t_1}^{t_2} g^2(t)dt}$ 5

Equation 5 gives the value of C for obtaining the best approximation.

ORTHOGONAL FUNCTIONS

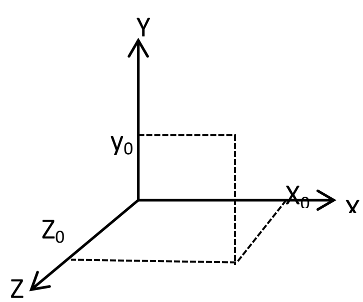


Fig. 3. Representation of a vector in 3 coordinates

The concept of orthogonality can be understood by considering the example of Vector **A** represented by Fig 3.

If $\mathbf{a}_x$, $\mathbf{a}_y$ and $\mathbf{a}_z$ are the unit vectors along x, y and z axes, then vector components along the three axes are $x_0 . \mathbf{a}_x$, $y_0 . \mathbf{a}_y$ and $z_0 . \mathbf{a}_z$ respectively so that

$A = x_0 . \mathbf{a}_x + y_0 . \mathbf{a}_y + z_0 . \mathbf{a}_z$……………………………………………………………6

Since the three vector are mutually perpendicular, the dot product will be

$$a_x . a_y = a_y . a_z = a_z . a_x = 0$$

and

$$a_x . a_x = a_y . a_y = a_z . a_z = 1$$

This can be described by the general expression of equation 7 which gives the condition for orthogonality.

$$a_m . a_n = \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases}$$ ……………………………………………………………………7

where m and n can have any value x, y and z.

The concept of three dimensional vector representation may be extended to n-dimensional representation with n-mutually perpendicular coordinates. If unit vectors along these coordinates are X_1, X_2 ......X_n, then a vector A in this coordinate system can be represented as:

$$A = C_1X_1 + C_2X_2 + C_3X_3 + ... C_nX_n +$$ …………………………………………………………8

where C_1, C_2 .... C_n are the components of vector A along these coordinates.

Equation 6 may be rewritten as

$$X_m . X_n = \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases}$$ ……………………………………………………………………9

Taking the dot product of both sides of Eq. 1.9 with the unit vector X_j, we obtain

$$A. X_j = C_1X_1 . X_j + C_2X_2 . C_3X_3 . X_j + ... C_j . X_j X_j + ... C_nX_nX_j$$

But the dot product $C_m . X_m . X_j = 0\ (m \neq j)$ so that only one term is left on the right side of this expression

$$AX_j = C_jX_jX_j = C_j$$ ……………………………………………………………………10

The set of mutually perpendicular vectors $(X_1X_2 ... X_n)$ is called an *orthogonal vector space*.

The product $\mathbf{X}_m.\mathbf{X}_n$ may have some constant value $\mathbf{k}_m$ instead of unity so that Eq. 9 may be rewritten as

$$X_m . X_n = \begin{cases} 0 & m \neq n \\ k_m & m = n \end{cases}$$ ……………………………………………………………………11.

Where $\mathbf{k}_m$ equals unity, the set is called normalized orthogonal set. Eq. 10 may be modified as

$$A . X_j = C_jX_j . X_j = C_jK_j$$

or

$$C_j = A . \frac{X_j}{K_j}$$

To summarise, we may say that orthogonal vector space comprises mutually perpendicular vector components. If a vector is orthogonal to another vector, it has no component along the other. Similarly, if a function is orthogonal to another function, it does not contain any component or the form of the other function. *A function cannot be approximated to another function orthogonal to it.* If we try to approximate a function with its orthogonal function, the error will be larger than the original function. This is similar to orthogonal vector components a_x and a_y. There can be no component of a_x in a_y and vice-versa.

When two signal are orthogonal

$$C_j = 0$$

Equation 1.5 then becomes

$$C=\frac{\int_{t_1}^{t_2} f(t)\,g(t)\,dt}{\int_{t_1}^{t_2} g^2(t) f\,dt}=0$$

or

$$\int_{t_1}^{t_2} f(t)\,g(t)=0 \qquad \text{... (1.13)}$$

Example 1.1. *Show that the functions sin nwt and cos mwt are orthogonal over the interval* $\left(t_0, t_0+\frac{2\pi}{\omega_0}\right)$ *where n and m are any integers.*

Solution. Let $f(t) = \sin n\,\omega_0\,t$

and $\qquad g(t) = \cos m\,\omega_0\,t$

$$C=\int_{t_1}^{t_2} f(t).\,g(t)\,dt$$

$$=\int_{t_0}^{t_0+\frac{2\pi}{\omega_0}} [\sin n\,\omega_0\,t \cos m\,\omega_0 t.\,dt]$$

$$=\int_{t_0}^{t_0+\frac{2\pi}{\omega_0}} \frac{1}{2}(2\sin n\,\omega_0\,t\cos m\,\omega_0 t)\,dt$$

$$=\frac{1}{2}\int_{t_0}^{t_0+\frac{2\pi}{\omega_0}} [\sin(n-m)\,\omega_0 t+\sin(n+m)\,\omega_0 t]\,dt$$

$$=-\frac{1}{2\omega_0}\left[\frac{1}{(n-m)}\cos(n-m)\,\omega_0 t+\frac{1}{(n+m)}\cos(n+m)\,\omega_0 t\right]_{t_0}^{t_0+\frac{2\pi}{\omega_0}}$$

$$=0$$

Since $C = 0$, the two functions are orthogonal.

Similarly, it can be shown that $\sin n\,\omega t$ and $\sin m\,\omega t$ and also $\cos n\,\omega t$ and $\cos m\,\omega t$ are orthogonal.

Example 1.2. *Determine the magnitude of the curve marked Y in Fig. 1.6 to ensure that mean square error is minimum when approximated to the square wave function.*

Solution. Let the square and sine wave functions be represented as

$$f(t)=\begin{cases} 1 & (0<t<\pi) \\ -1 & (\pi<t<2\pi)\end{cases}$$

$$g(t) = \sin t \text{ respectively}$$

$f(t)$ will be approximated with $g(t)$ over the period (0 to 2π)

$$\therefore \qquad f(t) \simeq C\sin t$$

Optimum value of C to minimise the mean square error in the approximation is given as

$$C=\frac{\int_0^{2\pi} f(t)\sin t\,dt}{\int_0^{2\pi}\sin^2 t\,dt}$$

$$=\frac{\int_0^{\pi}\sin t\,dt-\int_{\pi}^{2\pi}\sin t\,dt}{\int_0^{2\pi}\left(\frac{1-\cos 2t}{2}\right)dt}$$

$$=\frac{-[\cos t]_0^{\pi}+[\cos t]_{\pi}^{2\pi}}{\left[1-\frac{1}{2}\sin 2t\right]_0^{2\pi}}$$

$$=\frac{4}{\pi} \quad \textbf{Ans.}$$

ORTHOGONALITY IN COMPLEX FUNCTIONS

The previous discussion of orthogonality was limited to functions of real variables. If $f(t)$ and $g(t)$ are complex functions of real variable t, then $f(t)$ may be approximated over the interval (t_1, t_2) as

$$f(t) \cong Cg(t)$$

It can be shown that the optimum value of C to ensure minimum mean square error is given as

$$C = \frac{\int_{t_1}^{t_2} f(t)\, g^*(t)\, dt}{\int_{t_1}^{t_2} g(t)\, g^*(t)\, dt} \quad \text{... (1.14)}$$

where $g^*(t)$ is conjugate of $g(t)$.

The condition for orthogonality of the two functions is obtained by equating Eq. 1.14 to zero.

i.e.,

$$\int_{t_1}^{t_2} f(t)\, g^*(t)\, dt = \int_{t_1}^{t_2} f^*(t)\, g(t) = 0$$

where $f^*(t)$ is conjugate of $f(t)$.

where $f^*(t)$ is conjugate of $f(t)$.

For a mutually orthogonal set of complex functions over the interval (t_1, t_2), condition for orthogonality is given by Eq. 1.15

$$\int_{t_1}^{t_2} f_m(t) f_n^*(t)\, dt = \begin{Bmatrix} 0 & m \neq n \\ 1 & m = n \end{Bmatrix} \quad \text{... [1.15 (a)]}$$

If $g(t)$ is a complete set of function, then function $f(t)$ can be expressed as

$$f(t) = C_1 g_1(t) + C_2 g_2(t) + \dots + C_n g_n(t) \quad \text{... [1.15 (b)]}$$

where

$$C_n = \frac{1}{K_n} \int_{t_1}^{t_2} f(t)\, g_n^*(t)\, dt$$

It should be remembered that in case of real functions $g^*(t) = g(t)$.

APPROXIMATING A FUNCTION BY A SET OF MUTUALLY ORTHOGONAL FUNCTIONS

Assume a set of function,

$g_1(t), g_2(t) \dots g_n(t)$ orthogonal to one another over an interval of t_1 to t_2.

We know that

$$g_j(t)\, g_k(t) = \begin{bmatrix} 0 & j \neq k \\ K_j & j = k \end{bmatrix}$$

When

$$j = k = n,\ g_n^2(t) = K_n$$

If a function $f(t)$ is to be approximated over an interval (t_1 to t_2) by the above set of functions, then

$$f(t) \cong C_1 g_1(t) + C_2 g_2(t) + \dots C_n g_n(t)$$

$$= \sum_1^n C_n\, g_n(t) \quad \text{... (1.16)}$$

The mean square error ε is given as

$$\varepsilon = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} [f(t) - \sum_1^n C_n g_n(t)]^2 dt \quad \text{... (1.17)}$$

Since, the mean square error ε is function of constants $C_1, C_2 \dots C_n$, minimum value of e will be obtained when

$$\frac{\delta\varepsilon}{\delta C_1} = \frac{\delta\varepsilon}{\delta C_2} = \dots\dots\dots \frac{\delta\varepsilon}{\delta C_n} = 0$$

It can be easily shown that minimum value of mean square error will be obtained when

$$C_j = \frac{\int_{t_1}^{t_2} f(t) g_j(t)\, dt}{\int_{t_1}^{t_2} g_j^2(t)\, dt} \quad \text{... (1.18)}$$

$$= \frac{1}{K_j}\int_{t_1}^{t_2} f(t)\, g_j(t)\, dt \qquad \text{... (1.19)}$$

In order to obtain the best approximation with minimum mean square error, the coefficients $C_1, C_2 \ldots C_n$ should be chosen as given in Eq. 1.19.

EVALUATION OF MEAN SQUARE ERROR

The mean square error "e" given is determined by the use of equation 1.17.

$$\varepsilon = \frac{1}{(t_2 - t_1)}\int_{t_1}^{t_2}\left[f(t) - \sum_1^n C_n g_n(t)\right]^2 dt$$

$$\varepsilon = \frac{1}{(t_2 - t_1)}\left[\int_{t_1}^{t_2} f^2(t)\, dt + \sum_1^n \int_{t_1}^{t_2} C_n^2 g_n^2(t)\, dt - 2\sum_1^n \int_{t_1}^{t_2} C_n f(t) g_n(t)\, dt\right] \qquad \text{... (1.20)}$$

But from Eq. 1.19: $\int_{t_1}^{t_2} f(t) g_n(t)\, dt = C_n \cdot K_n$

Also $\int_{t_1}^{t_2} g_n^2(t) dt = K_n$. Substituting these values in Eq. 1.20.

$$\varepsilon = \frac{1}{(t_2 - t_1)}\left[\int_{t_1}^{t_2} f^2(t)\, dt + \sum_1^n C_n^2 K_n - 2\sum_1^n C_n^2 K_n\right]$$

$$= \frac{1}{(t_2 - t_1)}\left[\int_{t_1}^{t_2} f^2(t)\, dt - \sum_1^n C_n^2 K_n\right] \qquad \text{... (1.21)}$$

$$= \frac{1}{(t_2 - t_1)}\left[\int_{t_1}^{t_2} f^2(t)\, dt - (C_1^2 K_1 + c_2^2 K_2 + \ldots . C_n^2 K_n)\right] \qquad \text{... (1.22)}$$

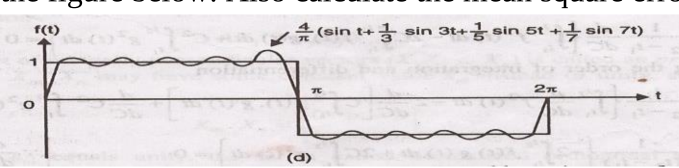
Equation. 1.22 shows that the mean square error can be decreased by increasing n in the approximation. When n is increased to infinity, the error becomes zero. Under this condition

$$\int_{t_1}^{t_2} f^2(t)\, dt = \sum_{n=1}^{n=\infty} C_n^2 . K_n \qquad \text{... (1.23)}$$

and $\qquad f(t) = C_1 g_1(t) + C_2 g_2(t) \ldots + C_n g_n(t)$: $(n \to \infty)$ $\qquad$ ... (1.24)

The series is said to converge in the mean. Eq. 1.24 is known as the Generalised *Fourier Series representation* of the function $f(t)$.

Assignment: Determine the values of constants C1, C2, …….C7 in the approximated waveform of the figure below. Also calculate the mean square error.



(d)

Solution. The approximated function may be written as

$$f(t) = C_1 \sin t + C_2 \sin 2t + C_3 \sin 3t + \ldots .\ C_7 \sin 7t$$

The value of constant C_n is determined by the use of Eq. 1.18.

(*a*)
$$C_n = \frac{\int_0^{2\pi} f(t) \sin nt\, dt}{\int_0^{2\pi} \sin^2 nt\, dt}$$

$$= \frac{\left(\int_0^{\pi} \sin nt\, dt - \int_{\pi}^{2\pi} \sin nt\, dt\right)}{\int_0^{2\pi} \frac{1 - \cos nt}{2}\, dt}$$

$$= \frac{-\frac{1}{n}[\cos nt]_0^{\pi} - \frac{1}{n}[\cos nt]_{\pi}^{2\pi}}{\frac{1}{2}\left[t - \frac{\sin nt}{n}\right]_0^{2\pi}}$$

$$= \frac{-\frac{1}{n}[\cos nt]_0^{\pi} - \frac{1}{n}[\cos nt]_{\pi}^{2\pi}}{\frac{1}{2}\left[t - \frac{\sin nt}{n}\right]_0^{2\pi}}$$

$$f(t) = \frac{4}{\pi}\sin t + \frac{4}{3\pi}\sin t + \frac{4}{5\pi}\sin 5t + \frac{4}{7\pi}\sin 7t$$
$$= \frac{4}{\pi}\left(\sin t + \frac{1}{3}\sin 3t + \frac{1}{5}\sin 5t + \frac{1}{7}\sin 7t \right)$$

(*b*) The mean square error ε is given by Eq. 1.17

$$\varepsilon = \frac{1}{(t_2 - t_1)}\left[\int_{t_1}^{t_2} f^2(t)\, dt - \sum\nolimits_1^n C_n^2 K_n \right] dt$$

Now (*i*)
$$t_2 - t_1 = 2\pi$$
$$f(t) = \begin{Bmatrix} 1 & (0 < t < \pi) \\ -1 & (\pi < t < 2\pi) \end{Bmatrix}$$

(*ii*)
$$\int_0^{2\pi} f^2(t)\, dt = 2\pi$$

(*iii*)
$$C_n = \begin{Bmatrix} \dfrac{4}{n\pi} & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{Bmatrix}$$

(*iv*)
$$K_n = \int_0^{2\pi} \sin^2 nt\, dt = \pi$$

$\therefore$
$$\varepsilon_1 = \frac{1}{2\pi}\left[2\pi - \left(\frac{4}{\pi}\right)^2 \times \pi \right] = 0.19$$

$$\varepsilon_3 = \frac{1}{2\pi}\left(2\pi - \left(\frac{4}{\pi}\right)^2 \times \pi - \left(\frac{4}{3\pi}\right)^2 \times \pi = 0.1 \right.$$

$$\varepsilon_5 = \frac{1}{2\pi}\left[2\pi - \left(\frac{4}{\pi}\right)^2 \times \pi - \left(\frac{4}{3\pi}\right)^2 \times \pi - \left(\frac{4}{5\pi}\right)^2 \times \pi \right]$$
$$= 0.0675$$

$$\varepsilon_7 = \frac{1}{2\pi}\left[2\pi - \left(\frac{4}{\pi}\right)^2 \times \pi - \left(\frac{4}{3\pi}\right)^2 \times \pi - \left(\frac{4}{5\pi}\right)^2 \times \pi - \left(\frac{4}{7\pi}\right)^2 \times \pi \right]$$
$$= 0.051$$

It can be summarised from the example above that as the number of terms is increased in the approximation, the mean square error is reduced. Representation of a function over an interval by a linear set of mutually orthogonal functions is termed the *Fourier Series representation*. Since there exists a large number of sets of orthogonal functions, a function may be represented in terms of different sets of orthogonal functions. This is analogous to the representation of a vector using different coordinate systems.

FOURIER SERIES

The trigonometric Fourier series is given by

$$= a_0 + \sum_{n=1}^{\infty} (a_n \cos n\omega_0 t + b_n \sin n\omega_0\, t)\left(t_0 < t < t_0 + \frac{2\pi}{\omega_0} \right) \qquad \ldots\ldots 1.27$$

The values of constants a_n and b_n are given by

$$a_n = \frac{\int_{t_0}^{\left(t_0 + \frac{2\pi}{\omega_0}\right)} f(t)\cos n\omega_0 t\, dt}{\int_{t_0}^{\left(t_0 + \frac{2\pi}{\omega_0}\right)} \cos^2 n\omega_0 t\, dt} \qquad b_n = \frac{\int_{t_0}^{\left(t_0 + \frac{2\pi}{\omega_0}\right)} f(t)\sin n\omega_0\, t\, dt}{\int_{t_0}^{\left(t_0 + \frac{2\pi}{\omega_0}\right)} \sin^2 n\omega_0\, t\, dt},$$

Let $2\frac{\pi}{w_0} = T$. When n=0, then

$$a_0 = \frac{1}{T}\int_{t_0}^{t_0+T} f(t)\, dt \qquad \ldots (1.28a)$$
$$b_0 = 0$$

Also
$$\int_{t_0}^{t_0+T} \cos^2 n\omega_0\, t\, dt = \int_{t_0}^{t_0+T} \sin^2 n\omega_0\, t\, dt = \frac{T}{2} \qquad \ldots (1.29)$$

$\therefore$
$$a_n = \frac{2}{T}\int_0^{(t_0+T)} f(t)\cos n\omega_0\, t\, dt \qquad n = 1, 2, 3, \ldots.. \qquad \ldots (1.30)$$

and $b_n = \frac{2}{T}\int_0^{(t_0+T)} f(t)\sin n\omega_0\, t\, dt \qquad n = 1, 2, 3, \ldots\ldots$ ... (1.31)

The constant gives the average for DC value of the function over the interval $(t_0, t_0 + T)$.

Compact Trigonometric Series

The trigonometric Fourier series in Eq. 1.27 contains sine and cosine terms of the same frequency. We can combine the two terms in a single term of the same frequency using the trigonometric identity.

$$a_n \cos n\omega_0 t + b_n \sin n\omega_0 t = C_n \cos(n\omega_0 t + Q_n) \qquad \ldots(1.31a)$$

Where $C_n = \sqrt{a_n^2 + b_n^2}$ .... (1.31b)

$$\theta_n = \tan^{-1}\left(\frac{-b_n}{a_n}\right) \qquad \ldots (1.31c)$$

For consistency we denote dc term a_0 by c_0 that is

$$C_0 = a_0 \qquad \ldots(1.31d)$$

Using the identity (1.31a). The Fourier series in equation 1.27 can be expressed as compact form of the trigonometric Fourier series as

$$f(t) = C_0 + \sum_{n=1}^{\infty} C_n \cos(n\omega_0 t + Q_n) \qquad t_0 < t < t_0 + \frac{2\pi}{\omega_0}$$

Where the coefficients C_n and Q_n are computed from a_n and b_n using equation 1.31.
Equation 1.28a shows that a_o or Co is the average value of f(t) averaged over one period. This is determined by inspection of f(t).

Exponential Fourier series

The exponential Fourier series over the interval (t_o, t_o+T) where T=2π/w_o is given by

$$f(t) = F_0 + F_1 e^{j\omega_0 t} + F_2 e^{j2\omega_0 t} + \ldots\ldots F_n e^{jn\omega_0 t} \ldots\ldots$$
$$+ F_{-1} e^{-j\omega_0 t} + F_{-2} e^{-j2\omega_0 t} + \ldots\ldots F_{-n} e^{-jn\omega_0 t} + \ldots\ldots$$
$$= \sum_{n=-\infty}^{n=\infty} F_n e^{jn\omega_0 t} (t_0 < t < t_0 + T) \qquad \ldots (1.33)$$

The coefficient F_n is analogous to C_n and its value is obtained by evaluating our earlier equation 1.14 which is repeated below.

$$C = \frac{\int_{t_1}^{t_2} f(t)\, g^*(t)\, dt}{\int_{t_1}^{t_2} g(t)\, g^*(t)\, dt} \qquad \ldots (1.14)$$

Therefore

$$F_n = \frac{\int_{t_0}^{(t_0+T)} f(t).\, e^{jn\omega_0 t} dt}{\int_{t_0}^{(t_0+T)} e^{jn\omega_0 t} e^{jn\omega_0 t} dt} = \frac{1}{T}\int_{t_0}^{(t_0+T)} f(t)\, e^{-jn\omega_0 t} dt \qquad \ldots (1.34)$$

The coefficients F_n are closely related to the coefficients of the trigonometric Fourier series. In fact, the exponential and trigonometric Fourier series are not different series but two different methods of expressing the same series. The relation between the coefficients of the two series is given below.

$a_0 = F_0$, $a_n = F_n + F_{-n}$, $b_n = j\,(F_n - F_{-n})$, $F_n = \frac{1}{2}(a_n - j\,b_n)$

REPRESENTING A PERIODIC FUNCTION BY THE FOURIER SERIES OVER THE ENTIRE INTERVAL ($-\infty, < t, < \infty$).

So far the discussion has been limited to the representation of a function by the Fourier series over the interval (t_o, t_o+T) where T=2π/w_o. The function $f(t_0)$ and the corresponding Fourier series may not be equal outside this interval. However, if the function is periodic, it may be represented by the Fourier series over the entire interval ($-\infty, < t, < \infty$).

Consider the function $f(t) = \sum_{n=-\infty}^{n=\infty} F_n\, e^{jn\omega_0 t}\,(t_0 < t < t_0 + T)$ ...[1.37(*a*)]

The equation is true over the interval ($t_0 < t < T$) and outside this interval, this representation may not be true.

However, $e^{jn\omega_0 t} = \cos n\omega_0 t + j \sin n\omega_0 t$ and is, therefore, periodic with period $T = \omega_0 / 2\pi$.

Therefore, the representation of Eq. 1.37 (*a*) is true over the entire interval $(-\infty < t < \infty)$.

i.e.,

$$f(t) = \sum_{n=-\infty}^{n=\infty} F_n\, e^{jn\omega_0 t} \qquad (-\infty < t < \infty)$$

where

$$F_n = \frac{1}{T}\int_{t_0}^{t_0+T} f(t)\, e^{-jn\omega_0 t}\, dt \qquad \text{... [1.37(}b\text{)]}$$

Assignment

Represent the output of a full wave rectifier shown in the figure below in terms of exponential Fourier series.

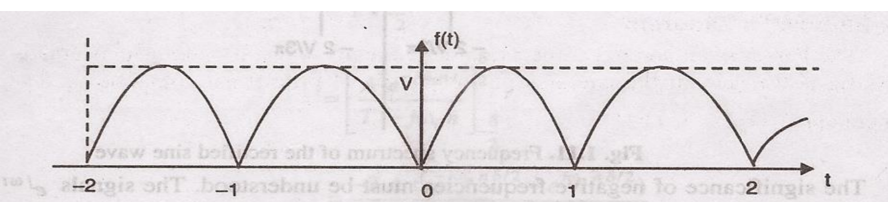


Fig. 1.10. Waveform of the full wave rectifier output.

Solution. Wave form of the full wave rectifier is shown in Fig. 1.10. The peak amplitude of the rectified output is V.

This wave form can be represented by the Fourier series as

$$f(t) = \sum_{n=-\infty}^{n=\infty} F_n \, e^{jn\omega_0 t}$$

For the wave form given, $T = 1$ and $\omega_0 = 2\pi$

$$F_n = V\int_0^1 \sin \pi t . e^{-j2\pi nt} dt$$

$$= \frac{-2V}{(4n^2-1)\pi}$$

$$\therefore \qquad f(t) = \frac{2V}{\pi}\sum_{n=-\infty}^{n=\infty} \frac{1}{4n^2-1} e^{j2\pi nt}$$

$$\text{Hence } f(t) = \frac{2V}{\pi} - \frac{2Ve^{j2\pi t}}{3\pi} - \frac{2V}{15\pi}e^{j4\pi t} - \frac{2Ve^{j6\pi t}}{35\pi} + \ldots\ldots$$

$$-\frac{2V}{3\pi}e^{-j2\pi t} - \frac{2V}{15\pi}e^{-j4\pi t} - \frac{2V}{35\pi}e^{-j6\pi t} + \ldots\ldots$$

1.5.7. Complex Fourier Spectrum. Expansion of a periodic function in terms of Fourier series is equivalent to resolving the function in terms of different frequency components. A periodic function $f(t)$ with a period T has frequency components with angular velocities $\pm\omega, \pm 2\omega, \pm 3\omega \ldots\ldots \pm n\omega$ when $\omega = 2\pi / T$. . This spectrum of frequencies defines a function. Every function has its own frequency spectrum, which can be determined by using Fourier series. Alternatively, a function may be found for a given frequency spectrum. A function may be defined as a function of time in terms of relative magnitudes possessed by various frequency components of the spectrum. This is termed as the *time domain* representation. Frequency spectrum comprises discrete frequencies ω, 2ω, 3ω,.....etc., and is not a continuous spectrum. A function may be *represented in frequency domain* with a series of equally spaced vertical lines ; their heights giving the magnitude of each frequency component. Frequency spectrum of the rectified sine wave of Fig. 1.10 is given below in Fig. 1.11.

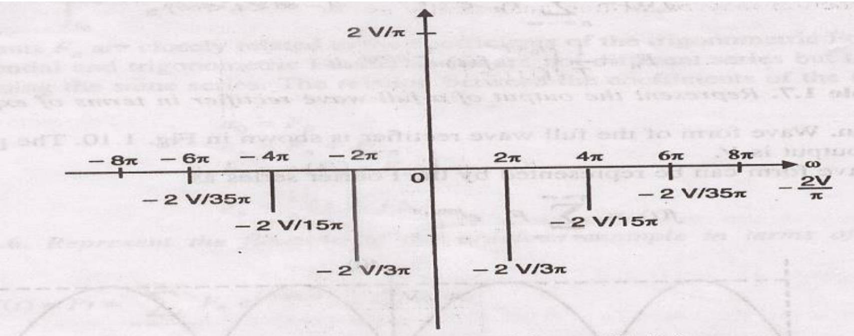


Fig. 1.11. Frequency spectrum of the rectified sine wave

The significance of negative frequencies must be understood. The signals $e^{j\omega t}$ and $e^{-j\omega t}$ have the same angular velocity ω but rotate in opposite directions. When added, they yield a real function of time.

$$\frac{e^{j\omega_0 t} + e^{-j\omega_0 t}}{2} = \cos \omega_0 t.$$

In general, the amplitudes F_n of different frequency components are complex and we need two representations, one for the *magnitude spectrum* and the other for the *phase spectrum*. In practice, however, the amplitudes of different frequency components may be either real or imaginary and it becomes possible to represent the function by one spectrum only. This can be seen from Fig. 1.11. The spectrum is symmetrical about vertical axis passing through the origin. This is true for the spectrum of every periodic function with a real F_n. If F_n is complex, then phase spectrum is antisymmetrical for F_{-n}.

Fig 1.12 shows a rectangular pulse Train in the time domain while Fig 1.13 shows the frequency domain representation.

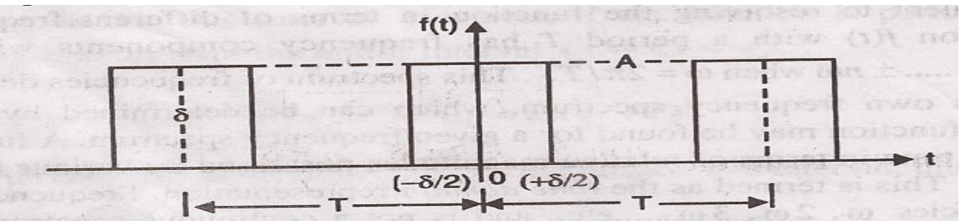


Fig.1.12. A rectangular pulse train.

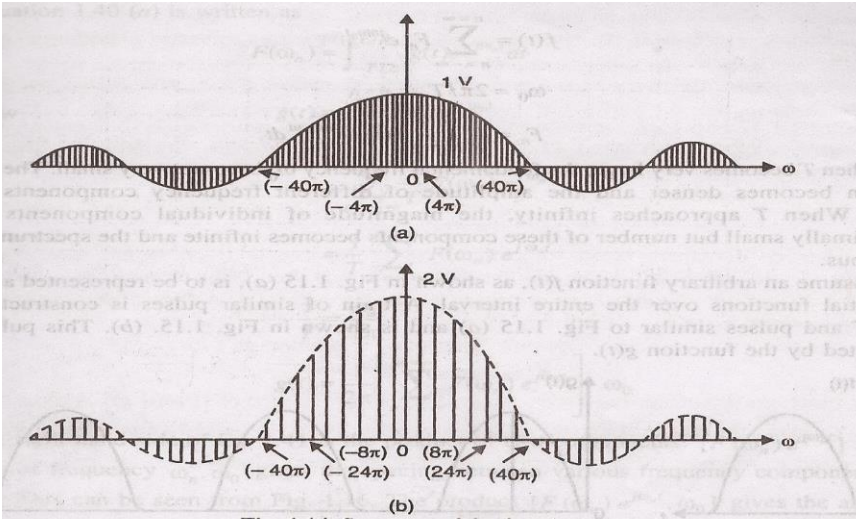


Fig 1.14: Spectrum of the function in the frequency domain.

The frequency spectrum for Fig 1.14a exists at frequencies 0, $\pm 4\pi, \pm 8\pi \ \pm 12\pi$ and the fundamental frequency component has a magnitude of 1V.

For Fig 1.14b the frequency spectrum exists at 0, $\pm 8\pi, \pm 16\pi \ \pm 24\pi \ \pm 32\pi$ and the fundamental frequency component has a magnitude of 2V.

A comparison of the two plots of Fig. 1.14 shows that amplitude F_n of different components is dependant on the amplitude A of the rectangular wave and the ratio δ/T. An increase in any of these results in the increase in the amplitude F_n.

The fundamental frequency w_0 varies inversely as pulse period T is increased. Increased T results in reduced fundamental frequency w_0 and as a consequence, harmonic components are more closely spaced. Amplitude F_n is also reduced. This can be clearly visualized from the figure.

As period T becomes larger and larger, the fundamental frequency w_0 becomes smaller and smaller and the frequency spectrum becomes denser but the shape of the spectrum is not affected. The shape or envelope of the spectrum depends entirely upon the shape of the pulse to which this spectrum belongs. When the period T becomes infinite, the function $f(t)$ comprises one non–repetitive function over the entire $(-\infty < t < \infty)$ interval and the spectrum becomes continuous.

THE FOURIER TRANSFORM

We have already shown that a non-repeatitive arbitrary function can be represented in terms of exponential or trigonometric series over a finite interval. For a periodic function, the representation can be extended over the entire interval. In this section we shall show that a non-periodic function can be expressed as a continuous sum (or integral) of exponential components. This is in contrast to periodic signals, which are represented by a discrete sum of exponential signals.

Consider the rectangular pulse train of Fig 1.12. Let f(t) represent the first cycle of this periodic function. Assume that the pulse period T is increased to become infinite. This periodic function will now have only one cycle in the interval (-∞, < t, < ∞) and is still represented by f(t). The function can be represented by an exponential Fourier series;

$$\therefore \qquad f(t) = \sum_{n=-\infty}^{n=\infty} F_n \, e^{jn\omega_0 t}$$

where $\qquad \omega_0 = 2\pi / T$

and $$F_n = \frac{1}{T} \int_{-T/2}^{T/2} f(t) \, e^{-j\omega_0 nt} \, dt.$$

When T becomes very large, the fundamental frequency ω_0 becomes very small. The frequenc spectrum becomes denser and the amplitude of different frequency components become smaller. When T approaches infinity, the magnitude of individual components become infinitesimally small but number of these components becomes infinite and the spectrum become continuous.

Assume an arbitrary function $f(t)$, as shown in Fig. 1.15 (a), is to be represented as a sum exponential functions over the entire interval. A train of similar pulses is constructed with period T and pulses similar to Fig. 1.15 (a) and is shown in Fig. 1.15. (b). This pulse train represented by the function $g(t)$.

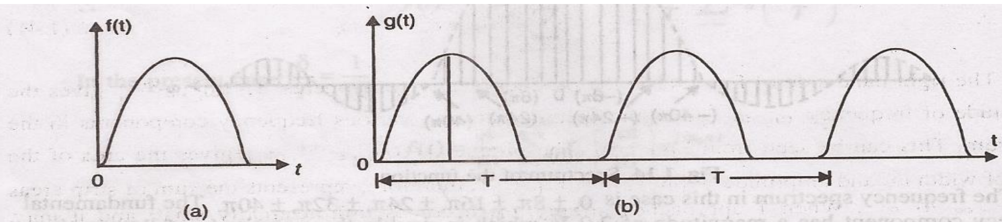


Fig. 1.15 (a) An arbitrary function $f(t)$. (b) A pulse train built with pulses.

The function $g(t)$ is periodic and can be represented by an exponential Fourier series. No assume that the period T is made infinite, then pulses in this periodic function will repeat after a infinite interval and the function $g(t)$ will become identical to $f(t)$.

Thus $\qquad \underset{T\to\infty}{\mathrm{LT}} \; g(t)\, t = f(t)$

The Fourier series representing $g(t)$ over the entire interval $(-\infty < t < \infty)$ will also represent $f(t)$

Now $$g(t) = \sum_{n=-\infty}^{n=\infty} F_n \, e^{jn\omega_0 t} \qquad \text{... [1.39 (a}$$

and $$F_n = \frac{1}{T} \int_{-T/2}^{T/2} g(t) \, e^{-jn\omega_0 t} \, dt \qquad \text{... [1.39 (b}$$

Let $n\omega_0 = \omega_n$. Since F_n is a function of $n\omega_0$, or ω_n, it may be denoted as $F_n(\omega_n)$. Eq. 1.3 (b) may, therefore be, rewritten as

$$F_n(\omega_n) = \frac{1}{T}\int_{-T/2}^{T/2} g(t)\, e^{-j\omega_n t}\, dt$$

or $$T.\,F_n(\omega_n) = \int_{-T/2}^{T/2} g(t)\, e^{-j\omega_n t}\, dt \qquad \text{...[1.40 (a)}$$

Assume that $T.\,F_n(\omega_n)$ is denoted simply as $F_n(\omega_n)$

i.e., $$T.\,F_n(\omega_n) = F(\omega_n) \quad \text{or} \quad F_n(\omega_n) = \frac{1}{T}F(\omega_n)$$

Equation 1.40 (a) is written as

$$F(\omega_n) = \int_{-T/2}^{T/2} g(t)e^{-j\omega_n t}\, dt \qquad \text{... [1.40 (b)]}$$

Now $$g(t) = \sum_{n=-\infty}^{n=\infty} F_n\, e^{jn\,\omega_0 t}$$

$$= \sum_{n=-\infty}^{n=\infty} \frac{1}{T} F(\omega_n)\, e^{j\,\omega_n t}$$

$$= \frac{1}{T}\sum_{n=-\infty}^{n=\infty} F(\omega_n)\, e^{j\,\omega_n t}$$

But $$\frac{1}{T} = \frac{\omega_0}{2\pi}$$

$\therefore$ $$g(t) = \frac{1}{2\pi}\left[\sum_{n=-\infty}^{n=\infty} F(\omega_n)\, e^{j\omega_n t}\right].\ \omega_0 \qquad \text{... (1.41)}$$

The right hand side of Eq. 1.41 is the product of two components. $[F(\omega_n)\, e^{j\omega_n t}]$ gives the amplitude of frequency ω_n. ω_0 gives the spacing between various frequency components in the spectrum. This can be seen from Fig. 1.16. The product $[F(\omega_n)\, e^{j\omega_n t}.\,\omega_0]$ gives the area of the strip of width ω_0 and amplitude $F(\omega_n)\, e^{j\omega_n t}$. Eq. 1.41, therefore, represents the sum of strip areas *from* $-\infty$ *to* ∞. *When T approaches infinity,* ω_0 *becomes infinitesimally small and it may,* therefore, be represented as $d\omega$. Equation 1.41. representing the sum of area of strips of width $d\omega$ may be written as the integral under the same limits.

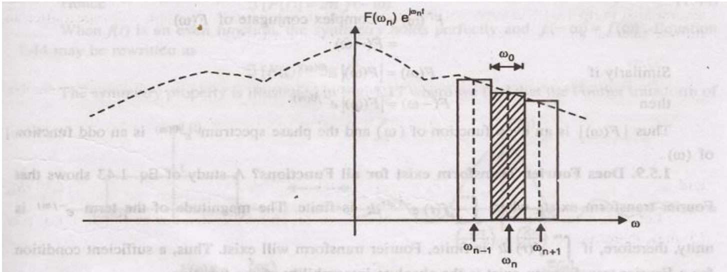


Fig. 1.6. Graphical Rpresentaretion of Eq. 1.42.

Thus $$\underset{T\to\infty}{\text{Lt}}\; g(t) = f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\, e^{j\omega t}\, d\omega \qquad (1.42)$$

and $$F(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-j\omega t}\, dt \qquad \text{... (1.43)}$$

ω_0 has been changed to ω in the above two equations, but it does not make any difference, since both ω_n and ω are used to represent any desired frequency in the spectrum.

Equations 1.42 and 1.43. show that a non-periodic function $f(t)$ can be represented by an exponential Fourier series over the entire interval $(-\infty < t < \infty)$ as a sum of exponential functions with frequencies lying in the entire interval.

$F(\omega)$ represents the frequency spectrum of the function $f(t)$ and is called the *spectrum density function.*

Equations 1.42 and 1.43 are called the *Fourier transform pair*. Eq. 1.42 is called the *direct Fourier transform* of $f(t)$ and Eq. 1.43 is termed the *inverse Fourier transform* of $F(\omega)$. These transforms are written symbolically as

$$\Im\,[f(t)] = F(\omega)$$

and $$\Im^{-1}\,[F(\omega)] = f(t)$$

The Fourier transform is an important tool to resolve a given signal into exponential components. A plot of $F(\omega)$ gives amplitudes of various frequency components and may be referred as a *Frequency domain representation* of the function $f(t)$. A plot of $f(t)$ gives the *time-domain representation* of the function.

$F(\omega)$ is generally a complex quantity and may be written as :

$$F(\omega) = |F(\omega)|\; e^{j\theta(w)}$$

Two plots are therefore, required to represent $F(\omega)$. A plot of $|F(\omega)|$ gives the magnitude, while $e^{j\theta(\omega)}$ gives the phase. In many cases, however, $F(\omega)$ is real or imaginary and only one plot is required to represent it. It will be shown that for a real function $f(t)$, $F^*(\omega) = F(-\omega)$.

Now $$F(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-j\omega t}\, dt$$

In a similar way $$F(-\omega) = \int_{-\infty}^{\infty} f(t)\, e^{j\omega t}\, dt$$

Thus if $f(t)$ is a real function, we have

$$F^*(\omega) = \text{Complex conjugate of } F(\omega) = F(-\omega)$$

Similarly if $$F(\omega) = |F(\omega)|\, e^{j\theta(\omega)}$$

then $$F(-\omega) = |F(\omega)|\, e^{-j\theta(\omega)}$$

Thus $|F(\omega)|$ is an even function of (ω) and the phase spectrum $e^{j\theta(\omega)}$ is an odd function of (ω).

1.5.9. Does Fourier Transform exist for all Functions? A study of Eq. 1.43 shows that Fourier transform exists when $\int_{-\infty}^{\infty} f(t)\, e^{-j\omega t} dt$ is finite. The magnitude of the term $e^{-j\omega t}$ is unity, therefore, if $\int_{-\infty}^{\infty} f(t)\, dt$ is finite, Fourier transform will exist. Thus, a sufficient condition for a Fourier transform to exist is the absolute integrability of the function.

The condition of absolute integrability is not always necessary and some functions not absolutely integratable possess Fourier transforms. Some functions like sin ωt, cos ωt, $u(t)$ etc., are not absolutely integratable but they have Fourier transforms in a limit.

1.5.10. Properties of the Fourier Transfrom. The Fourier transform of a function is the way of expressing the function in terms of its exponential components of various frequencies. Thus, we may express a function either in time-domain (plot of magnitude against time) or in frequency–domain (plot of magnitude against frequency).

In this section, we shall study the effect of an operation performed in one domain over the representation in the other domain. For example, if a function represented in time-domain is differentiated or integrated, how is the frequency-domain representation affected ?

The equations defining the two domains show some symmetry. This symmetry can be expected even when an operation is performed. Thus the effect on the time-domain due to differentiation in the frequency-domain should be similar to the effect on the frequency-domain due to differentiation in the time-domain. Assume that we show the correspondence between two domains by a double arrow, then $f(t) \leftrightarrow F(\omega)$ shows that $F(\omega)$ is the direct Fourier transform of $f(t)$ and inverse Fourier transform of $F(\omega)$ is $f(t)$.

1. **Symmetry.** If $f(t) \leftrightarrow F(\omega)$

then $F(t) \leftrightarrow 2\pi f(-\omega)$

The symmetry property is illustrated in Fig 1.17 where we find that the Fourier transform of a gate function is a sampling function and the Fourier transform of a sampling function is a gate function. If the function f(t) is not an even function, the symmetry is not perfect.

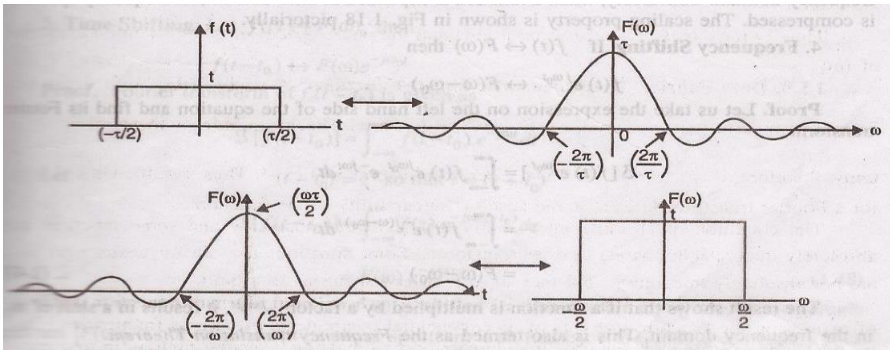


Fig. 1.17. Symmetry between a gate function and its Fourier transform.

2. **Linearity.** Assume that $f_1(t) \leftrightarrow f_1(\omega)$

and $f_2(t) \leftrightarrow f_2(\omega)$

then $C_1 f_1(t) + C_2 f_2(t) \leftrightarrow C_1 F_1(\omega) + C_2 F_2(\omega)$

where C_1 and C_2 are constants. The linearity is valid for n functions

$$C_1 f_1(t) + C_2 f_2(t) + \ldots . C_n f_n(t) \leftrightarrow C_1 F_1(\omega) + C_2 F_2(\omega) + \ldots . C_n F_n(\omega)$$

3. **Scaling.** Assume $f(t) \leftrightarrow F(\omega)$ and C is a real constant, then

$$f(Ct) \leftrightarrow \frac{1}{(C)} F\left(\frac{\omega}{a}\right)$$

4. **Frequency Shifting.** If $f(t) \leftrightarrow F(\omega)$ then

$$f(t)\, e^{j\omega_0 t} \leftrightarrow F(\omega - \omega_0)$$

Proof. Let us take the expression on the left hand side of the equation and find its Fourier transform.

$$\Im\,[f(t)\, e^{j\omega_0 t}] = \int_{-\infty}^{\infty} f(t)\, e^{j\omega_0 t}\, e^{-j\omega t}\, dt$$

$$= \int_{-\infty}^{\infty} f(t)\, e^{-j(\omega-\omega_0)t}\, dt$$

$$= F(\omega - \omega_0) \qquad \ldots (1.47)$$

The result shows that if a function is multiplied by a factor $e^{j\omega_0 t}$, it results in a shift of ω_0 in the frequency domain. This is also termed as the *Frequency translation Theorem.*

Consider the process of amplitude modulation where a carrier signal $f(t)$ is multiplied by a modulating sinusoidal signal. The resultant wave is given by

$$f(t)\cos\omega_0 t = f(t)\left[\frac{e^{j\omega_0 t} + e^{-j\omega_0 t}}{2}\right]$$

$$= \frac{1}{2}\left[f(t)\, e^{j\omega_0 t} + f(t)\, e^{-j\omega_0 t}\right]$$

Applying frequency translation theorem, we obtain

$$f(t)\cos\omega_0 t = \frac{1}{2}[F(\omega + \omega_0) + F(\omega - \omega_0)]$$

It can be shown similarly that

$$f(t)\,.\,\sin\omega_0 t = \frac{1}{2}[F(\omega + \omega_0) - F(\omega - \omega_0)]$$

5. **Time Shifting.** Let $f(t) \leftrightarrow F(\omega)$, then

$$f(t-t_0) \leftrightarrow F(\omega)e^{-j\omega_0 t}$$

Proof. Fourier transform of $f(t-t_0)$ is given as

$$\Im[f(t-t_0)] = \int_{-\infty}^{\infty} f(t-t_0)\, e^{-j\omega t}\, dt$$

Let $(t-t_0) = z$ so that $t = (z + t_0)$

$$\therefore \quad \Im[f(t-t_0)] = \int_{-\infty}^{\infty} f(z)\, e^{-j\omega(z+t_0)}\, dz$$

$$= F(\omega)\, e^{-j\omega t_0} \qquad \text{... (1.48)}$$

Thus, it is seen that if a function $f(t)$ is shifted in time-domain by t_0 sec, its magnitude spectrum $|F(\omega)|$ is not affected but its phase-spectrum is shifted by $-\omega t_0$. In other words, a shift of t_0 in the time-domain is equivalent to multiplying by $e^{-j\omega t_0}$ in the frequency-domain.

6. **Differentiation / Integration with Respect to Time.** If $f(t) \leftrightarrow F(\omega)$ then $\frac{df}{dt} \leftrightarrow (j\omega)\, F(\omega)$

and $\int_{-\infty}^{t} f(\tau)\, d\tau \leftrightarrow \frac{1}{j\omega} F(\omega)$ provided that when $t = 0$, $F = (0) = 0$ or $\int_{-\infty}^{\infty} f(t)\, dt = 0$

The process of differentiation of a function is the same as multiplying the Fourier transform by jω. Integration of the function results in dividing the Fourier transform by jω.

7. **Differentiation in Frequency Domain.** Assume $f(t) \leftrightarrow F(\omega)$, then

$$\frac{dF}{d\omega} \leftrightarrow -jtf(t)$$

The result may be extended to the nth derivative of $F(\omega)$

$$\therefore \quad \frac{d^n F}{d\omega^n} = (-jt)^n \cdot f(t)$$

Thus, we find that differentiation in the frequency-domain is equivalent to multiplication by jt in the time-domain.

8. **The Convolution Theorem.** (1) The convolution of two functions in time-domain is equivalent to multiplication of their spectra in the frequency-domain.

(2) The multiplication of two functions in the time-domain is equivalent to convolution of their spectra in the frequency-domain.

According to the first convolution theorem, if we are given two functions $f_1(t)$ and $f_2(t)$, with their Fourier transforms $F_1(\omega)$ and $F_2(\omega)$ respectively, then

$$f(t) = f_1(t)^* f_2(t) = \int_{-\infty}^{\infty} f_1(\tau) f_2(t-\tau)\, d\tau \leftrightarrow F_1(\omega)F_2(\omega) \qquad \text{... [1.51 (a)]}$$

where sign * denotes the convolution integral.

Table 1.3 Important properties of the Fourier Transforms.

Operation	*Time Domain* $f(t)$	*Frequency Domain* $f(\omega)$
Scaling	$f(a.\,t)$	$\frac{1}{\lvert a \rvert} F\left(\frac{\omega}{a}\right)$
Frequecny shifting	$f(t)\, e^{j\,\omega_0 t}$	$F(\omega - \omega_0)$
Time shifting	$f(t - t_0)$	$F(\omega)\, e^{-j\,\omega\, t_0}$
Time derivative	$\frac{d^n f}{dt^n}$	$j\omega^n F(\omega)$
Frequency derivative	$(-jt)n\, f(t)$	$\frac{d^n F}{d\omega^n}$
Time Integration	$\int_{-\infty}^{t} f(t)\, dt$	$\frac{1}{j\omega} F(\omega)$
Time convolution	$f_1(t) * f_2(t)$	$F_1(\omega) F_2(\omega)$
Frequency convolution	$f_1(t)\, f_2(t)$	$\frac{1}{2\pi} F_1(\omega) * F_2(\omega)$

Example 1.11. *Find the Fourier transform of a gate function G (t) shown in Fig. 1.21 (a) and defined as*

$$G(t) = \begin{cases} 1 & |t| < \tau/2 \\ 0 & |t| > \tau/2 \end{cases}$$

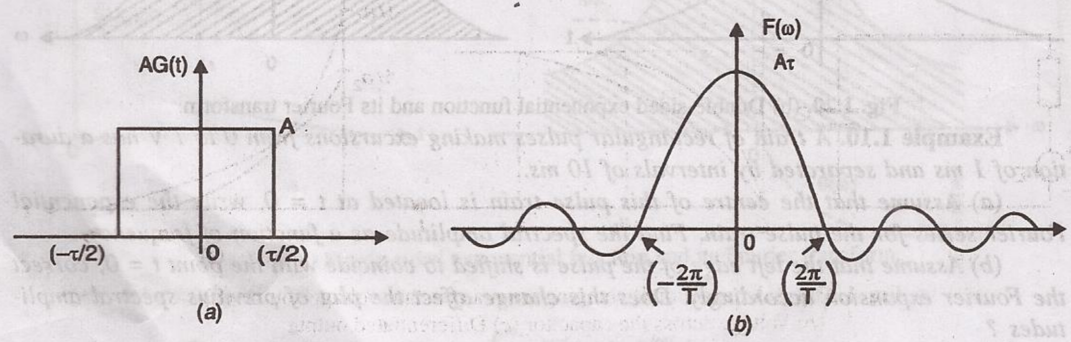


Fig. 1.21 Gate function *Ag* (*t*) and its frequency spectrum.

Solution. The Fourier transform of this gate function is given as

$$F(\omega) = \int_{-\tau/2}^{\tau/2} A\, e^{-j\omega\tau}\, dt$$

$$= \frac{A}{j\omega}\left(e^{j\omega\tau/2} - e^{-j\omega\tau/2}\right)$$

$$= \frac{A\tau}{\frac{\omega\tau}{2}} \times \left(\frac{\frac{e^{j\omega\tau}}{2} - \frac{e^{-j\omega\tau}}{2}}{2j}\right) \qquad \frac{\left[e^{j\omega\tau/2} - e^{-j\omega\tau/2}\right]}{2j}$$

$$= A\tau = \frac{\sin\omega\tau/2}{\omega\tau/2}$$

$$= A\tau\, S_a(\omega\,\tau/2) \qquad \dots (1.54)$$

Fourier transform of this gate function is shown in Fig. 1.21 (*b*).

1.5.11. Impulse Functions. Figure 1.22 shows an unit impulse function. The unit impulse function denoted as $\delta(t)$ is not a true function in the mathematical sense, because it is not defined for all values of *t*. The use of impulse function is, however, very common in science and engineering to represent ideally, point charges, point masses, point sources, surface charges etc.

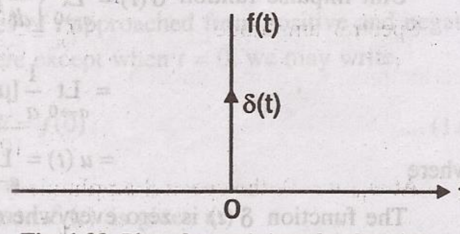


Fig. 1.22. Plot of an unit impulse function.

If we wish to write a mathematical expression for the impulse function, we face a difficulty, because the function starts from zero at $t = 0$ and reaches its maximum magnitude while the t is still zero. This is an impractical condition. Thus, we look at the function from another angle. Consider the gate function described in Fig. 1.21 and mathematically expressed in Example 1.10. Assume, we reduce the width of this gate function. As the width τ approaches zero, the gate function approaches the impulse function. Let us study some more details about the unit impulse function.

Assume a constant voltage source having an open circuit voltage of *E* volts and a capacitor *C* connected across this generator, as shown in Fig. 1.23 (*a*). Circuit wave form is shown in 1.23(*b*).

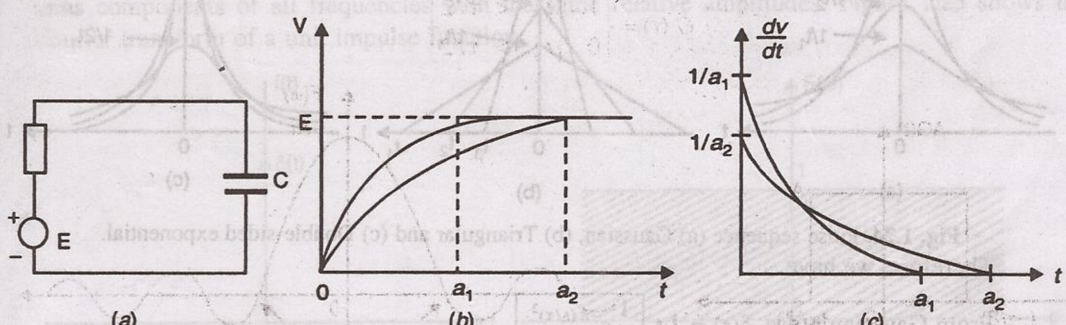


Fig. 1.23. (*a*) A capacitor C charging from a constant voltage source (E) (*b*) Voltage across the capacitor (*c*) Differentiated output.

When the circuit is switched on at $t = 0$, the voltage V across the capacitor is zero and it begins to build up at a rate that depends upon the internal resistance R of the generator. After some time, the capacitor charges to a voltage that equals the generated voltage E. There is no current flow in the circuit and hence no voltage drop across R. Capacitor voltage $V = E$ at this instant. The time taken by the capacitor to charge to this voltage is dependent upon the resistance R. Less the resistance R, the less the time required to charge the capacitor. Thus, when $R = R_1$, the charging time required is a_1 and when $R = R_2$, charging time is a_2. It is obvious that $R_2 > R_1$.

Now assume that we connect a differentiating circuit to differentiate the capacitor voltage. The differentiated output is as shown in Fig. 1.23 (*c*). It is seen that if the internal resistance R of the voltage source is decreased, the height of the differentiated output is increased and its width is reduced. Thus when R approaches zero (ideal case), the differentiated output approaches an infinite height and zero width. Thus, it becomes an impulse function represented by Fig. 1.22. The area of the pulse, however, remains same.

An impulse function may be defined as the derivative of a unit step function $u(t)$. When $E = 1$ V, the step function becomes unit step function and impulse function is called *unit impulse function*. Area of an unit impulse function is, therefore, unity.

$$\text{Unit impulse funtion } \delta(t) = \underset{a \to 0}{\text{Lt}} \left\{ \frac{d}{dt} [u_a(t)] \right\}$$

$$= \underset{a \to 0}{\text{Lt}} \frac{1}{a} [\mu(t) - \mu(t-a)] \quad \text{... (1.55)}$$

where

$$= u(t) = \underset{a \to 0}{\text{Lt}} [u_a(t)]$$

The function $\delta(t)$ is zero everywhere except when $t = a$. Since the area of a unit impulse function is unity, it may be written as

$$\int_{-\infty}^{\infty} \delta(t)\, dt = 1$$

and $\quad \delta(t) = 0 \,;\; t \neq 0 \quad$ ... (1.56)

A unit impulse function may also be obtained from the pulse sequences shown in Fig. 1.24 as τ is made to approach zero.

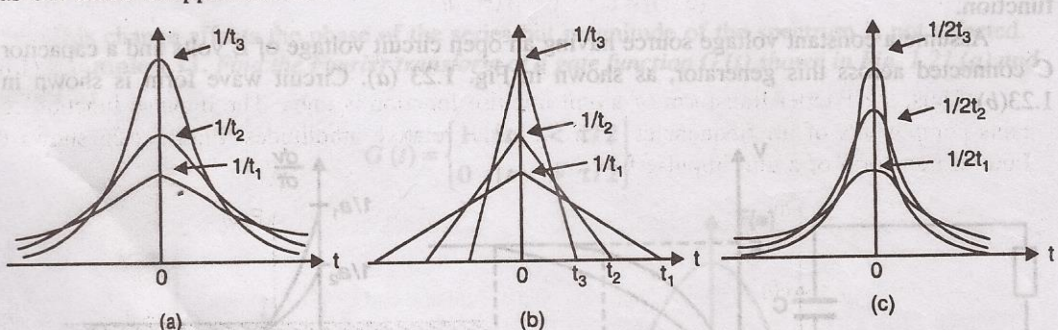


Fig. 1.24. Pulse sequence (a) Gaussian, (b) Triangular and (c) Double-sided exponential.

Therefore, we have

$$\text{From Gaussian pulse } \delta(t) = \underset{\tau \to 0}{\text{Lt}} \left[\frac{1}{\tau} e^{-\pi (t/\tau)^2} \right]$$

$$\text{From Triangular pulse } \delta(t) = \left\{ \underset{\tau \to 0}{\text{Lt}} \left[\frac{1}{\tau} \left(1 - \frac{|t|}{\tau} \right) \right] \; |t| < \tau \right\}$$

From Exponential pulse $\delta(t) = \underset{\tau \to 0}{\mathrm{Lt}} \left[\frac{1}{2\tau} \frac{e^{-|t|}}{\tau} \right]$.

Lastly, we consider the sampling function shown in Example 1.13. It can be shown that the integral

$$\int_{-\infty}^{\infty} \frac{k}{\pi} S_a(kt)\, dt = 1$$

As k is increased, the amplitude of this function increases and the function oscillates faster but decays rapidly from the origin. As k approaches infinity, the function exists only at the origin, the amplitude approaches infinity but the net area still remains unity. Thus this function becomes the impulse function.

$$\therefore \qquad \delta(t) = \underset{k \to \infty}{\mathrm{Lt}} \left[\frac{k}{\pi} S_a(kt) \right] \qquad \text{...(1.58)}$$

Since the area of a unit impulse function is concentrated at the origin, we may say

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 = \int_{0^-}^{0^+} \delta(t).dt \qquad \text{...(1.59)}$$

where 0^+ and 0^- denote arbitrarily small values of t approached from positive and negative side respectively. Also since the $\delta(t) = 0$ everywhere except when $t = 0$, we may write,

$$\int_{-\infty}^{\infty} f(t)\, \delta(t)\, dt = f(0) \int_{-\infty}^{\infty} \delta(t)\, dt = f(0) \qquad \text{... (1.60)}$$

Fourier Transform of an Impulse Function

The Fourier transform of a unit impulse function $\delta(t)$ is given as

$$\Im[\delta(t)] = \int_{-\infty}^{\infty} \delta(t)\, e^{-j\omega t}\, dt$$

But from Eq. 1.60, we have the integral

$$\int_{-\infty}^{\infty} \delta(t)\, f(t)\, dt = f(0) \int_{-\infty}^{\infty} \delta(t)\, dt$$

$$\therefore \qquad \Im[\delta(t)] = e^{-j\omega t} \int_{-\infty}^{\infty} \delta(t)\; dt$$

$$= 1 \qquad [\because \text{ At } t = 0, e^{-j\omega t} = 1] \qquad \text{... (1.61)}$$

Thus, the Fourier transform of a unit impulse function is unity. The impulse function contains components of all frequencies with the same relative amplitudes. Figure 1.25 shows the Fourier transform of a unit impulse function.

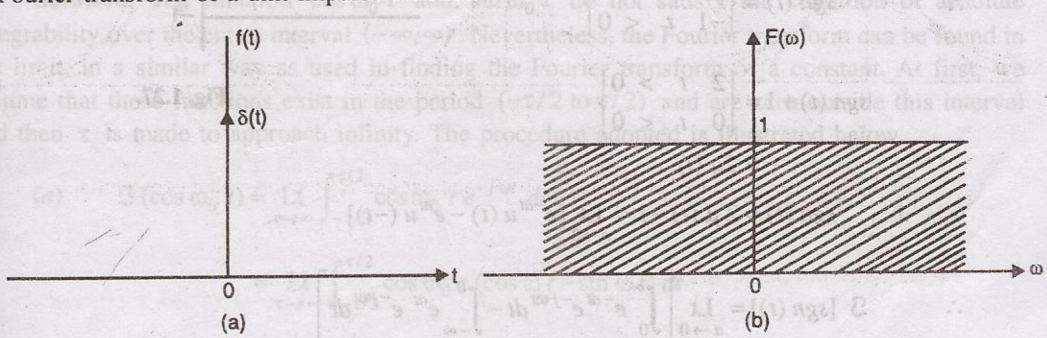


Fig. 1.25. Fourier transform of a unit impulse function.

Example 1.12. *Find the Fourier transform of the function shown in Fig. 1.26.*

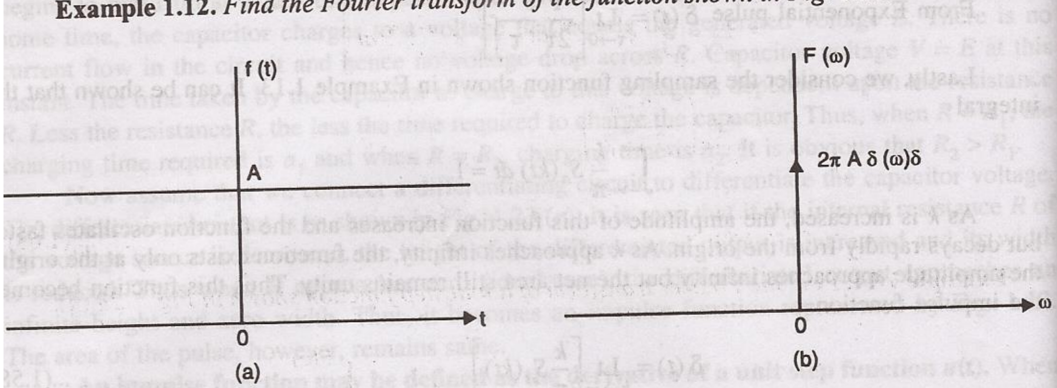


Fig.1.26. Plot of a constant and its Fourier transform.

Solution. As can be seen that the function shown in Fig. 1.26 *(a)* has a value *A* that remains same for all values of *t*.

$$\therefore \qquad f(t) = A$$

This function may be considered as a rectangular pulse with infinite pulse-width. Therefore, Fourier transform of this function is same as the gate function, when pulse-width approaches infinity.

$$\Im\, f(t) = \Im\,[A] = \underset{\delta\to\infty}{\mathrm{Lt}} \left[A\delta\, S_a \frac{\omega\delta}{2} \right]$$

$$= 2\pi A \underset{\delta\to\infty}{\mathrm{Lt}} \left[\frac{\delta}{2\pi} S_a \frac{\omega\delta}{2} \right]$$

But $\qquad \underset{\delta\to\infty}{\mathrm{Lt}} \left[\frac{\delta}{2\pi} S_a \frac{\omega\delta}{2} \right] = \delta(\omega)$ [See Eq. 1.58]

$$\therefore \qquad \Im\,[A] = 2\pi A \delta\,(\omega) \qquad \text{... (1.62)}$$

Eq. 1.62 shows that the Fourier transform of a constant *A* is an impulse function.

Example 1.13. *Find the Fourier transform of the function shown in Fig.* 1.27.

Solution. This function has a magnitude equal to 1 when *t* is greater than zero and –1 when $t < 0$. It is known as *signum* function.

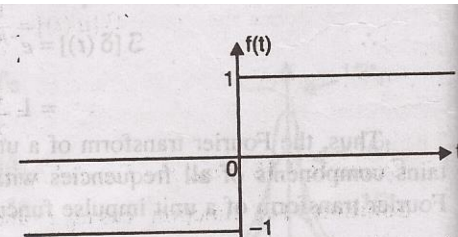


Fig. 1.27.

$$\therefore \qquad sgn\,(t) = \begin{Bmatrix} 1 & t > 0 \\ -1 & t < 0 \end{Bmatrix}$$

$$sgn\,(t) + 1 = \begin{Bmatrix} 2 & t > 0 \\ 0 & t < 0 \end{Bmatrix}$$

$$= 2u\,(t)$$

$$sgn\,(t) = 2\,\mu\,(t) - 1 = \underset{a\to 0}{\mathrm{Lt}}\, [e^{-at} u\,(t) - e^{at} u\,(-t)]$$

$$\therefore \qquad \Im\,[sgn\,(t)] = \underset{a\to 0}{\mathrm{Lt}} \left[\int_0^\infty e^{-at} e^{-j\omega t} dt - \int_{-\infty}^0 e^{at}\, e^{-j\omega t} dt \right]$$

$$= \underset{a \to 0}{\text{Lt}} \left[\int_0^{\infty} \overline{e}^{\,(a+j\omega t)}\, dt - \int_{-\infty}^{0} e^{\,(a-j\,\omega t)}\, dt \right]$$

$$= \underset{a \to 0}{\text{Lt}} \left(\frac{-2j\omega}{a^2+\omega^2} \right)$$

$$= \frac{2}{j\omega} \qquad \text{... (1.63)}$$

Example 1.14. *Find the Fourier transform of a unit step function shown in Fig. 1.28 (a).*

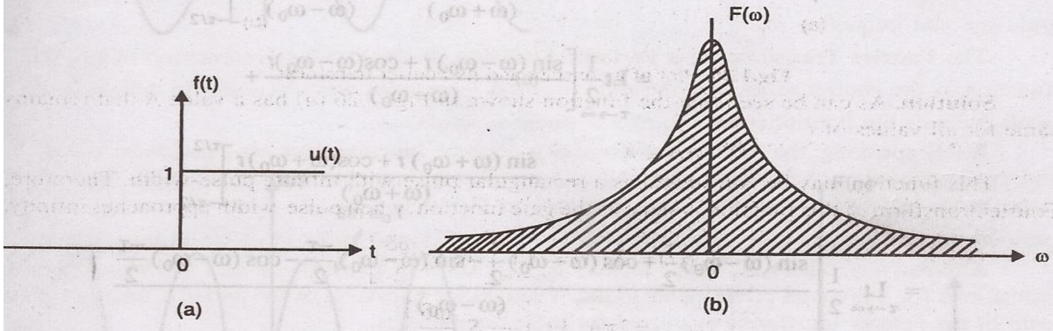


Fig. 1.28. (*a*) A unit step function (*b*) Fourier transform.

Solution. In the previous example, it was stated that $sgn\ (t) = 2u(t) - 1$

$$\therefore \qquad u\,(t) = \frac{1}{2}\ [sgn(t) + 1]$$

$$\Im\,[u\,(t)] = \frac{1}{2}\{\Im\,[sgn\ (t) + 1]\}$$

$$= \frac{1}{2}\Im\,[sgn\ (t) + \frac{1}{2}\ \Im\,[1] = \frac{1}{j\omega} + \pi\delta\,(\omega) \qquad \text{... (1.64)}$$

Table 1.4 gives transforms of some common functions.

	Function	*Fourier transform* $F\,(\omega)$
1.	$\mu\,(t)\, e^{-at}$	$\dfrac{1}{a+j\omega}$
2.	$e^{-a\lvert t \rvert}$	$\dfrac{2a}{a^2+\omega^2}$
3.	$\delta\,(t)$	1
4.	A	$2\pi A\delta\,(\omega)$
5.	$\mu\,(t)$	$\dfrac{1}{j\omega} + \pi\delta\,(\omega)$
6.	$\cos\omega_0\, t$	$\pi\,[\delta\,(\omega-\omega_0) + \delta\,(\omega+\omega_0)]$
7.	$\sin\omega_0\, t$	$j\pi\,[\delta\,(\omega-\omega_0) - \delta\,(\omega-\omega_0)]$
8.	$e^{j\,\omega_0\, t}$	$2\pi\delta\,(\omega-\omega_0)$
9.	$\delta_T(t)$	$\omega_0 \sum_{n=-\infty}^{n=\infty} \delta\,(\omega - n\omega_0)$
10.	$\lvert\, t \,\rvert$	$\dfrac{-2}{\omega^2}$
11.	$\mu\,(t)\, te^{-at}$	$\dfrac{1}{(a+j\omega)^2}$

12.	$\mu(t)\cos\omega_0 t$	$\frac{\pi}{2}[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)]+\frac{j\omega}{\omega_0^2-\omega^2}$
13.	$\mu(t)\sin\omega_0 t$	$-j\frac{\pi}{2}[\delta(\omega-\omega_0)-\delta(\omega+\omega_0)]+\frac{\omega_0}{\omega_0^2-\omega^2}$
14.	$\mu(t)e^{-at}\sin\omega_0 t$	$\frac{\omega_0}{(a+j\omega)^2+\omega_0^2}$
15.	$G_\tau(t)$	$\tau S_a\left(\frac{\omega\tau}{2}\right)$
16.	$\begin{cases}1-\frac{\lvert t\rvert}{\tau} & \lvert t\rvert<\tau \\ 0 & \lvert t\rvert>\tau\end{cases}$	$\tau\left[S_a\left(\frac{\omega\tau}{2}\right)\right]^2$

1.6. Parseval's Theorem

The Parseval's theorem can be explained in many ways. It is defined as energy of *a* components Eg is the sum of the energies of the individual orthogonal components. The energy of *a* component $C_n x_n(t)$ is $C_n^2 E_n$,

or
$$Eg = C_1^2 E_1 + C_2^2 E_2 + C^3 E_3 + \ldots\ldots$$
$$= \sum_n C_n^2 E_n \qquad \ldots (1.73)$$

This important result is called *Parseval's theorem*. In terms of area (the area under the squared value of a signal) the signal energy is analogous to the square of the length of a vector in the vector-signal analogy. In vector space it is seen that, the square of the length of a vector is equal to the sum of the squares of the lengths of its orthogonal components. This statement is the same as Parseval's theorem equation (1.73) as applied to the signal.

In terms of a periodic signal $f(t)$ is a power signal, and every term in its Fourier series is also a power signal. The power P_f of $f(t)$ is equal to the power of its Fourier series. The power of the Fourier series is equal to the sum of the powers of its Fourier Components. This is same as *Perseval's theorem*. Thus for the trigonometric Fourier Series

$$f(t) = C_o + \sum_{n=1}^{\infty} C_n \cos(n\omega_0 t + \theta_n)$$

The power of $f(t)$ is given by

$$P_f = C_0^2 + \frac{1}{2}\sum_{n=1}^{\infty} C_n^2 \qquad \ldots. (1.74)$$

For the exponential Fourier series

$$f(t) = D_o + \sum_{n=-\infty}^{\infty} D_n e^{jn\omega_0 t} \qquad (\text{where } n \neq 0)$$

The power of the exponential Series is given by

$$P_f = \sum_{n=-\infty}^{\infty} |D_n|^2 \qquad \ldots. (1.75)$$

For a real $f(t)$, $|D_{-n}| = |D_n|$ therefore

$$P_f = D_0^2 + 2\sum_{n=1}^{\infty} |D_n|^2 \quad \text{... (1.76)}$$

The Parseval's theorem of the form eq (1.73) applies to the energy signals, and of the form eq. (1.74) applies to periodic signals represented by the trigonometric Fourier series and eq. (1.75) applies to periodic signals represented by the exponential Fourier series.

Another form of Parseval's theorem can be related to Fourier transform $F(\omega)$ as given below

$$E_f = \int_{-\infty}^{\infty} f(t)\, f'(t)\, dt = \int_{-\infty}^{\infty} f(t) \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} F'(\omega)\, e^{-j\omega t} d\omega \right] dt$$

where $F(\omega)$ is the signal spectrum and E_f is the signal energy. Also $f'(t)$ is the conjugate of $f(t)$. Now interchanging the order of integration

$$E_f = \frac{1}{2\pi} \int_{-\infty}^{\infty} F'(\omega) \left[\int_{-\infty}^{\infty} f(t)\, e^{-j\omega t} dt \right] d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)\, F'(\omega)\, d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega \quad \text{... (1.77)}$$

This relation is the same as statement of *Parseval's theorem.* This relation allows us to determine the signal energy from either the frequency-domain specification $F(\omega)$ or the time-domain specification $f(t)$ of the same signal.

Example 1.21. *Verify Parseval's Theorem for the signal*

$$f(t) = e^{-\alpha t} \mu(t) \qquad \text{(where } \alpha > 0\text{)}$$

Solution. We have

$$E_f = \int_{-\infty}^{\infty} f^2(t)\, dt = \int_{0}^{\infty} e^{-2\alpha t} dt = \frac{1}{2\alpha}$$

We now determine E_f from the signal spectrum

$F(\omega)$ given by $F(\omega) = \dfrac{1}{j\omega + \alpha}$

And from analysis of Fourier transform eq. (1.77)

$$E_f = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2\, d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\omega^2 + \alpha^2}\, d\omega$$

$$= \frac{1}{2\pi\alpha} \left| \tan^{-1} \frac{\omega}{\alpha} \right|_{-\infty}^{\infty} = \frac{1}{2\alpha}$$

Thus verifies Parseval's theorem.