Breakthroughs in Information and Communication Technologies

Part III

V Rajaraman

In Part 1 of this series of articles, I defined what is meant by a breakthrough in ICT and listed fourteen breakthroughs in chronological order. Subsequently, nine breakthroughs were detailed in Parts 1 and 2. In this concluding part, I describe the last five breakthroughs: search engines, digitization and compression, mobile computers, cloud computing, and deep learning.

V Rajaraman is at the Indian Institute of Science, Bengaluru. Several generations of scientists and engineers in India have learnt computer science using his lucidly written textbooks on programming and computer fundamentals. His current research interests are parallel computing and history of computing.

10. Search Engines

The World Wide Web allows anyone to access a document whose URL is known. However, it would be a lot more convenient and useful, if given a set of keywords, a software could search the web and retrieve the relevant documents. For example, one may like to use a phrase – capitals of states in India – where capitals, states and India are the keywords. A software to retrieve documents from the World Wide Web that match these keywords is known as a search engine. As the required information is scattered across various websites, the search engine would navigate the web and collect all the relevant web pages that have these keywords and display them. Search engine technology has a long history. In fact, the main objective of Vannevar Bush's MEMEX was to organize and search the knowledge stored by individuals in their private knowledge repositories. As soon as the World Wide Web appeared, the designers of the web saw the importance of searching for contents in web pages scattered all over the world. The first step in this direction was to prepare a directory of websites. This, however, became impractical as the number of websites increased rapidly.

Keywords
Search engines, digitization and compression, mobile computing, cloud computing, deep learning.

The first software developed to search the contents of files stored

The first software developed to search the contents of files stored on websites was called Archie and was developed in September 1990 by a group led by Alan Emtage at the McGill University in Canada.

on websites was called Archie and was developed in September 1990 by a group led by Alan Emtage at the McGill University in Canada. Archie downloaded the directories of files on publicly accessible sites, (called the anonymous File Transfer Protocol sites), creating a searchable database of file names. The files were not indexed as their number was small. This was followed by search software called Veronica and Jughead that indexed the files and allowed keyword-based search using Gopher, a protocol that was popular in the 1990s to retrieve documents from the Internet. All these early systems were directory-based as the number of files stored on the web was small [1, 2].

A web crawler program that systematically visits websites, reads their content and creates an index that can be used by a search engine was first designed in December 1993 by Jonathon Fletcher in Scotland. He designed a search engine called JumpStation. This was not commercially exploited and remained unknown. Many search engines soon appeared that used his idea. Among these are engines named AltaVista, Excite, Infoseek, Inktomi, and Yahoo!

The search engine, Google, was developed by two Stanford University students Sergey Brin and Larry Page in 1998. This engine crawled the web and picked up the most relevant web pages using an algorithm called PageRank. In the words of Larry Page, "PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites". An index is created with the search terms and addresses of the websites where they appear. When a user types in the search bar of a search engine, a query with a set of keywords, the index is searched using these keywords. The titles of all the web pages where these keywords appear are listed in the order of their relevance.

The search engine, Google, was developed by two Stanford University students Sergey Brin and Larry Page in 1998. This engine crawled the web and picked up the most relevant web pages by using an algorithm called PageRank.

Gerald Salton defined two important metrics of a good retrieval system called 'recall' and 'precision'. Recall measures the percentage of relevant documents retrieved from the corpus that includes all the relevant documents. Precision measures the percentage of relevant documents in the set of documents retrieved. For example, if a search gives 250 documents of which only 100 are relevant, then the precision is 100/250 = 0.4. If there are actually 400 relevant documents in the corpus of documents out of which only 100 are retrieved that are relevant, then the recall is 100/400 = 0.25. It is practically impossible for a retrieval system to make both recall and precision 100%. Higher the recall of a retrieval system, lower is its precision.

Gerald Salton defined two important metrics of a good retrieval system called 'recall' and 'precision'. Recall measures the percentage of relevant documents retrieved from the corpus that includes all relevant documents. Precision measures the percentage of relevant documents in the set of documents retrieved.

Most search engines today maximise recall, to ensure that they do not miss any relevant documents with the result that the number of documents retrieved and listed is huge. However, the quality of a search engine is judged by the users based on the order of listing of documents with the most relevant ones appearing at the beginning. Many tricks considered 'trade secret' by the search engine companies are used to ensure this. Nowadays, websites are created or updated frequently. Thus, search engines have to update their indices in real-time to give the up-to-date answer to queries. Search engines are free for users. Search engine companies, however, have to spend a lot of money to continuously improve their search results. In 2003 Google realized that it could monetize search by linking search results with the advertisements of relevant companies. This resulted in computers becoming an important media for advertisers besides television and print media. Advertising revenue from its search engine made Google a rich company enabling it to invest in research and acquire other companies with good products. For example, Google acquired YouTube, a company that allowed users to create multimedia presentations, music with video, and publish them at no cost on its website. Search was extended beyond text to multimedia, and advertising revenue was earned from YouTube also.

Search engines have to update their indices in real-time to give the up-to-date answer to queries. Search engines are free for users. Search engine companies, however, have to spend a lot of money to continuously improve their search results. In 2003 Google realized that it could monetize search by linking search results with the advertisements of relevant companies. This resulted in computers becoming an important media for advertisers besides television and print media.

Currently, search engines use machine learning methods to understand the behaviour of each individual user, by observing their search preferences and cater to their specialised needs. (Machine learning may be defined as use the of artificial intelligence (AI) tools that endows a system with the ability to automatically learn

Machine learning may be defined as the use of artificial intelligence tools that endows a system with the ability to automatically learn from experience without being explicitly programmed.

from experience without being explicitly programmed). Besides simple textual queries, currently, search engines can search for images and answer spoken queries. Old and out of print books have been digitised by Google and added to their repository to improve search results. Publishers now give sample pages of their copyrighted books to Google to allow book search, besides searching websites. Search engines have enhanced the utility of the World Wide Web just like the World Wide Web enhanced the utility of the Internet. There are currently about half a dozen search engines such as Bing, Yahoo!, Ask.com, DuckDuckGo, and Baidu, besides Google. However, Google is the most popular search engine with over two-thirds of the market share.

Search engine is a breakthrough in ICT due to the following reasons:

The idea of searching for specified keywords in websites in the World Wide Web using web crawling, indexing, assigning ranks to pages, and listing pages in the order of their relevance to the search query is novel.

- The idea of searching for specified keywords on websites in the World Wide Web using web crawling, indexing, assigning ranks to pages, and listing pages in the order of their relevance to the search query is novel.
- Libraries with a large collection of books and journals are slowly transforming to a digital form. The World Wide Web along with search engines is a great convenience to researchers who can now access research material such as journals, technical reports, conference proceedings, statistical data, experimental data, and similar material from anywhere, anytime. In fact, the World Wide Web with over four billion indexed web pages as on January 2018 and Google books is a huge repository of information and it is growing rapidly. Without a search engine, finding required information on the Web would be like searching for a needle in a haystack. World Wide Web with an associated search engine may now be called an 'Information Utility'.
- Everyone with access to the Web uses a search engine to find the latest score in sporting events, the latest news, information on medicines, reviews of movies, gossip on film stars, meanings of words, synonyms and antonyms, style rules while writing an essay, route to locations in a city, and a plethora of such information.

- The publishing industry, particularly music publishers, are withering as plenty of music is available on the Web and search engines retrieve the required music instantly. Self-publishing of music as well as books has created a new industry in which the creators of intellectual material have a new medium to publish their work without depending on the publishers.

- The advertising industry has a new medium with search engines providing them data to target the correct prospects.

- Search engines coupled with maps and GPS have created a new cab hailing application that is transforming personal transport in cities.

- A number of new web-based applications such as renting spare rooms to tourists have been enabled by search engines. In fact, the entire e-commerce boom is due to the availability of search engines coupled with the World Wide Web.

The publishing industry, particularly music publishers, are withering as plenty of music is available on the Web and search engines retrieve the required music instantly. Self-publishing of music, as well as books, has created a new industry in which the creators of intellectual material have a new medium to publish their work without depending on the publishers.

11. Digitization and Compression of Multimedia

There are five types of data processed by computers. They are numbers, text, images, audio, and video. All data processed by computers are a series of binary digits (bits). Decimal numbers are converted to binary for processing. Text is a series of characters. To process text, the characters are encoded to binary in a standard format such as the American Standard for Information Interchange (ASCII) or Unicode. Early computers were used mainly to process numbers and text. With the advent of integrated circuits and the rapid reduction of the cost of digital processing, it became advantageous to convert images, audio, and video (often called multimedia) to digital form. This is called 'digitization'. The number of bits used to store multimedia data can be significantly reduced without losing the essential information contained in them using what are known as 'compression algorithms'. There are many advantages of digitization and compression among which the most important are the reduction in storage and cheaper transmission of data to remote computers. With the advent of the Internet, digital storage and transmission

The number of bits used to store multimedia data can be significantly reduced without losing the essential information contained in them using what are known as compression algorithms.

of data of all types have become essential. It is not easy to specify a particular date for this breakthrough, namely, digitization and compression of multimedia, as it was gradual and took many years. I have taken the year 1993 when video and audio compression algorithms were standardised by the International Standards Organization (ISO) as the date of this breakthrough. I describe some digitization and compression methods in what follows. [3]

Images

Handwritten documents on paper and other media such as palm leaves, old printed books, photographs, and similar objects may be classified as images. Many of these are prone to deterioration and fading. Early attempts were made in the 1950s to preserve them in electronic form using drum scanners. The first electronic drum scanner was invented in 1957 by Russell A. Kirch. In this scanner, a black and white picture was mounted on a drum. The drum was rotated, and a light source reflected light from the picture. The white spots in the picture reflected light while the black spots did not. The reflected light was sensed by an electronic photo-tube and stored as bits on a magnetic tape. The technology of scanning, as well as, photography was revolutionised when Willard S. Boyle and George E. Smith invented the 'charge-coupled device' (CCD) in 1969 for which they got the Nobel Prize in Physics in 2009.

Charge-coupled device (CCD) is an integrated circuit etched onto a silicon substrate of light sensitive elements. When light falls on the light sensitive element, it stores a charge. When the reflected light from an object falls on a CCD, the stored charge represents a picture element (pixel) of the object.

CCD is an integrated circuit etched onto a silicon substrate of light-sensitive elements. When light falls on the light sensitive element, it stores a charge. When the reflected light from an object falls on a CCD, the stored charge represents a picture element (pixel) of the object. CCD arrays are used now in flatbed scanners, barcode readers and digital cameras. CCDs were first used in the digital camera invented by Steven Sasson in 1975 at Kodak in the USA. However, the first commercial digital camera was marketed by Nikon of Japan much later in 1986.

Flatbed scanners using CCDs appeared around 1985. For using a flatbed scanner, the material to be scanned is placed on a glass plate of the scanner. A light source focuses light on an entire scan line as a thin beam. Light is reflected from bright spots and none from dark spots. The reflected light is sensed by a CCD array and stored as picture elements or pixels. Usually, a resolution of around 400 pixels per inch is adequate. The resulting data is called a bitmap as each pixel is converted into a bit. For monochrome pictures usually, 16 grey levels are used per pixel that requires 4 bits per pixel. For colour pictures, red, green and blue filters are used and each pixel is stored as 24 bits, 8 bits each for red, green and blue representing 16 million colours. In digital cameras, a set of lenses is used to focus the image of the object being photographed on a CCD array. The array stores the image as pixels.

Various file formats are used to store images. They are: bmp (bitmap), tiff (tagged image file format), and jpeg (joint photographic experts group). Each has its advantages and disadvantages that we will not discuss and one may refer to [4] in the suggested reading list. The number of bits stored when documents are scanned can be very large. For example, if a book with 1000 pages is scanned, the bitmap form requires 1.5 GB. Thus, compression algorithms are essential. The main goal of compression algorithms is to get the best image quality of the compressed image with maximum compression. There are several compression algorithms that are broadly classified as lossless algorithms and lossy algorithms. In lossless algorithms, the original image can be reconstructed exactly after decompression, whereas in a lossy compression it cannot be. Lossy compression usually gives a higher compression factor compared to lossless compression. As human perception of images can tolerate some loss, lossy compression is often used in image compression. For maintaining a reasonable quality, the maximum compression factor is around 20, that is, a 1 MB image is reduced to 50 KB.

The main goal of compression algorithms is to get the best image quality of the compressed image with maximum compression. There are several compression algorithms that are broadly classified as lossless algorithms and lossy algorithms.

As we pointed out, digital cameras using CCD arrays appeared in 1986. By 1995 digital cameras became popular as their prices dropped. By 2005 film cameras disappeared and digital cameras took over. Digital cameras were incorporated into smartphones

In 2013 front facing cameras started appearing in smartphones allowing one to take his or her photo without requesting someone else to take a photo. This was called a 'selfie' and by 2015 the word 'selfie' became a word in the Oxford English dictionary.

around 2000 and by 2010 they became very common. In 2013 front facing cameras started appearing in smartphones allowing one to take his or her photo without requesting someone else to take a photo. This was called a 'selfie' and by 2015 the word 'selfie' became a word in the *Oxford English Dictionary*.

High-end smartphones nowadays have a resolution of 20 megapixels. Each image is stored using 60 MB (3 bytes for each pixel of a colour picture). If 100 photos are to be stored, they require 6 GB. This is quite large. They are thus compressed using a standard algorithm called jpeg by a factor of around 15 thereby reducing the required storage to 400 MB. The quality of the compressed image is very good.

Video

A video is a picture that captures motion. The main difference between a still picture and a video is the need to repeat each picture at least 30 times per second to perceive motion.

Sony released the first all-digital video camera in 1986. It scanned images using CCDs and repeated image capture 30 times a second. The uncompressed digital images were stored which had an enormous number of bits. Ampex released the first compressed video in 1993.

The history of video cameras is quite long. The first working motion picture camera and projector were built by Augustus and Louis Lumiere in 1895. The first electronic video camera was invented by Philo Farnsworth in 1927. An image is focussed in this device by a lens onto a plate coated with a layer of caesium oxide. When light strikes the plate, electrons are emitted that are gathered by a detection circuit and amplified. A competing system used by television cameras was invented by Vladimir Zworykin around 1930. It took another 20 years before colour TV camera was invented in 1950. All these and their successors, portable camcorders, that appeared in the 1970s were analog devices. Sony released the first all-digital video camera in 1986. It scanned images using CCDs and repeated image capture 30 times a second. The uncompressed digital images were stored that had an enormous number of bits. Ampex released the first compressed video in 1993. All these early recorders recorded the data on magnetic tapes or disks.

Following the invention of flash memory[1] video recorders inte-

[1]Flash memory is a semiconductor memory that stores data even when it is not connected to a source of power, popularly known as a pen drive or memory card.

grated them by 2003. A major problem in digital video recording is the enormous number of bits that are generated. 1 GB of storage is required for 1 hour video. Hence, compression is required for economical storage and transmission of video. The primary idea used in video compression is the fact that most of the pixels do not change between successive frames of a video and, therefore, only the changing data need to be stored. Software and devices incorporating the software called video coders/decoders (CODECS) are used for compression. Standards for video compression have evolved over the years starting from the early 1990s in order to incorporate interchangeability of compressed video data among the manufacturers of video equipment. There are two sets of standards – one developed by the International Standards Organization (ISO) and the other by the International Telecommunication Union (ITU). The ISO standards are called MPEG-1, MPEG-2 and MPEG-4. The ITU standards are H.261, H.262 and H.265. These standards are based on the same basic principle but differ in details. MPEG stands for Motion Picture Experts Group. MPEG-1 developed in 1993 standardised digital audio and video storage in CDROMs. MPEG-2 developed in 1995 standardised videos for DVDs and television including high definition (HD) TV. MPEG-4 part 2 was standardised in 1999 mainly to support video on low bit rate (around 64 bps) systems for applications that use the Internet and mobile devices. It also supports synthetic animation and may be considered as a subset of the H.265 standard. MPEG-4 Part 10 was developed in 2003 and corresponds to H.264 advanced video coding. It significantly improved picture quality at low bit rates but required more processing. H.265/HEVC standardised in 2013, supports 8000 dots per inch, 120 frames per second very high-resolution video. It supports parallel processing for compression. H.264 is probably the most commonly used standard that supports video on the Internet and mobile devices providing a compression factor of 100 that requires only 10 MB storage for a 1-hour video [4].

Standards for video compression have evolved over the years starting from early 1990s in order to incorporate interchangeability of compressed video data among manufacturers of video equipment. There are two sets of standards – one developed by the International Standards Organization (ISO) and the other by the International Telecommunication Union (ITU).

Audio

As early as 1928, Harry Nyquist, who was working at the Bell Telephone Laboratories, USA, proved that all the information contained in an analog signal is preserved when only a small number of samples of the signal is taken that is slightly more than $2f_h$.

Audio or voice signal generated by our vocal chords is a continuously varying signal, namely, an analog signal. In order to store and process it using computers, it is necessary to convert it to digital form. As early as 1928, Harry Nyquist, who was working at the Bell Telephone Laboratories in the USA, proved that all the information contained in an analog signal is preserved when only a small number of samples of the signal is taken that is slightly more than $2f_h$ where f_h is the highest frequency component of the signal. For example, the highest frequency that an average person can hear is 22 kHz. Thus, if slightly more than 44000 samples per second of an audio signal are taken, we can reconstruct the original signal with no loss of quality. For example, for storing audio in a CDROM, 44100 samples per second of an audio is recorded by the audio industry.

The amplitude of an audio signal sample is a number (e.g., a voltage level if it is the output of a microphone). This has to be represented in binary for the audio signal to be digitized. The question that arises is – how many bits should be used to represent the amplitude of the signal? The number of bits to be used depend on the incremental loudness levels human ears can perceive. If 16 bits are taken to represent the amplitude, more than 64000 levels can be perceived which is quite adequate.

With the advent of integrated circuits, it was easy and inexpensive to build electronic circuits that could convert an analog audio signal generated by a microphone to its digital equivalent. Such a circuit was called an analog-to-digital (A/D) converter and became available as a chip in the early-1970s.

With the advent of integrated circuits, it was easy and inexpensive to build electronic circuits that could convert an analog audio signal generated by a microphone to its digital equivalent. Such circuit was called an analog-to-digital (A/D) converter and became available as a chip in the early-1970s. It is also required to convert the digital data after it is processed to analog signal that could be fed to a loudspeaker for us to hear. A digital-to-analog (D/A) converter is required for this. An integrated circuit for doing this also appeared in early 1970s. The A/D converter sampled audio signals at a rate determined by Nyquist theorem and digitized the samples to a number of bits determined by the designer.

Digital audio recording standard for CDROM appeared in 1984. It used 16 bits to represent the amplitude of audio signals and 44100 samples per second. Thus, one hour of stereo audio signal if digitized will require $2 \times 441000 \times 60 \times 60 \times 16$ bits ≈ 600 MB. This is smaller than required for storing of video, yet it is quite large. Digital audio files are normally compressed for storage and transmission.

The main principles used in audio compression are – reducing the sampling rate when the audio frequencies are low, taking no samples during silent intervals in speech and music, and reducing the number of bits used for the amplitude of the samples using some properties of the human hearing system and how sound is processed by our brain. Based on these, an algorithm called mp3 was proposed in 1989. This algorithm applies a set of filters to the audio signal and in each frequency band picks the appropriate number of samples. With mp3 compression, usually a compression ratio of 96 is obtained for telephone conversations and between 12 and 14 for CD quality audio. One-hour audio when compressed using mp3 will require 600/12 = 50 MB. In 1993, mp3 compression was standardised as a part of the MPEG-1 standard [4].

The main principles used in audio compression are – reducing the sampling rate when the audio frequencies are low, taking no samples during silent intervals in speech and music, and reducing the number of bits used for the amplitude of the samples using some properties of the human hearing system and how sound is processed by our brain.

Digitization and compression of images, audio and video is a breakthrough in ICT due to the following reasons:

- Humans generate and consume continuously varying signals. The idea that a discrete number of samples of these signals can be selected and digitized for processing, storage and transmissions is novel.
- The idea that digitized image, audio and video data can be compressed while preserving their information content is novel.
- The idea that charge coupled device (CCD) can be used to design digital cameras and scanners is novel.
- Digitizing images has enabled the digital preservation, storage and indexing of old, rare manuscripts in libraries. Researchers all over the world can use these stored and indexed documents from anywhere in the world.

Digitized audio has improved the audio quality. In order to use audio with the Internet, digitization and compression are essential. Gadgets such as iPod and software such as iTunes introduced by Apple Computers that revolutionised music distribution depend on digitization and compression.

- The physical space required to store archival documents has been drastically reduced due to digitization, compression and storage in digital media such as tapes.
- Paper documents degrade. Preservation of scanned and compressed digital equivalents in digital media is more secure and environment-friendly.
- Internet infrastructure has been made more useful because of the availability of a huge number of documents, audio and video material, created, digitized, and stored by individuals all over the world.
- Digitized audio has improved audio quality. In order to use audio with the Internet, digitization and compression are essential. Gadgets such as iPod and software such as iTunes introduced by Apple Computers that revolutionised music distribution depend on digitization and compression.
- Digital video has improved the quality of TV and has led to HDTV. Digitization and compression have enabled viewing live TV broadcasts using mobile devices.
- The video on demand industry owes its existence to digitization and compression of video.
- Industries that made photographic material (films, chemicals, photographic paper) have been disrupted by the advent of digital cameras. A new industry of digital camera manufacturers and manufacturers of CCDs has replaced the old photographic industry.
- Our society has been transformed because of the ease of digital photography, audio and video compression and distribution. Social media platforms such as Facebook depends on the ease of distribution of digitized audio and video.

Our society has been transformed because of the ease of digital photography, audio, and video compression and distribution. Social media such as Facebook depend on the ease of distribution of digitized audio and video.

12. Mobile Computing

The advantages of portable mobile computers were realised quite early with the advent of single chip microprocessors. One of the earliest portable computers was Osborne 1 that was marketed in

1981. This was followed by a version of the modern laptop computer – Grid Compass 1101 – in 1982 that could be run using batteries and had a foldable electro-luminescent display.

The modern laptop became popular in the mid-1990s due to the advent of many technologies such as Lithium-ion batteries, power saving processors, improved liquid crystal displays, and thin high capacity hard disks that could be placed below the keyboard of the laptop. By mid-1990s LANs in most organizations were connected to the Internet. Internet access became essential for all computer users. A mobile laptop that was not connected to the Internet was an 'orphan' and not very useful. Hence, there was a great demand to connect laptops wirelessly to LANs.

A number of radio frequency bands called the industrial, scientific and medical (ISM) bands centred around 900 MHz, 2.4 GHz and 5.8 GHz were reserved by the International Telecommunication Union (ITU) for gadgets that used radio frequencies. These bands were not permitted to be used for communication.

A number of radio frequency bands called the industrial, scientific and medical (ISM) bands centred around 900 MHz, 2.4 GHz and 5.8 GHz were reserved by the International Telecommunication Union (ITU) for gadgets that used radio frequencies. These bands were not permitted to be used for communication.

The Federal Communications Commission (FCC), the telecom regulator of the USA, decided in 1985 to allow the ISM bands to be used for wireless communication in addition to their use in gadgets. These bands could be used by anyone in a restricted range that limited the power of the transmission below one watt, without a licence or payment to the government. Soon ITU allowed ISM bands to be freely used all over the world without a licence for wireless communication. (Now many more radio frequency bands have been opened up for communication in restricted range without a licence or a fee).

Advent of Wi-Fi

Even though ISM bands were free there was no standard that could be used by all the vendors of wireless equipment to build and market systems in these bands. The Institution of Electrical and Electronics Engineers (IEEE) of USA formed a committee to standardise the systems that use wireless transmission using the ISM bands. After extended discussions, a standard called

IEEE802.11 was published in 1997. There were two variants of this standard – IEEE 802.11a that operated in the 5 GHz band and IEEE802.11b that operated in the 2.4 GHz band. A consortium of half a dozen companies decided to cooperate and built a transceiver in the 2.4 GHz band. The transceiver power was around 100 milliwatts and the range around 50 to 100 metres depending on obstructions such as thick walls. The name Wi-Fi was selected for this wireless system to rhyme with hi-fi that was a popular term. By the end of 1998 Wi-Fi devices were available in the market.

Early Wi-Fi systems provided a speed of around 10 Mbps. Two new standards, IEEE 802.11g and IEEE 802.11n in the 2.4 GHz band were announced with speeds of 54 Mbps and 600 Mbps in 2003 in 2009 respectively. A new standard IEEE 802.11ac in the 5 GHz band with a speed of 3466.8Mbps was announced in 2013 and IEEE 802.11ax in the 2.4/5 MHz band is expected to be released in Dec.2018 with a speed up to 10.53Gbps.

In 1999 Apple Computers provided a slot in its laptop computer iBook, for a Wi-Fi transceiver card that was being sold for around USD 100. Wi-Fi enabled laptops could be connected to LANs that had a computer with a wireless transceiver. LANs were normally connected to the Internet. This allowed laptops to access the Internet by connecting to a LAN anywhere in its vicinity. Soon wireless routers to connect to the Internet were developed that were called Wi-Fi 'hot spots'. Laptops and other mobile devices could then connect to the Internet via 'hot spots' that were placed not only in organizations and homes but also in public places such as airports, railway stations and restaurants allowing anytime, anywhere connectivity.

Early Wi-Fi systems provided a speed of around 10Mbps. Two new standards, IEEE 802.11g and IEEE 802.11n in the 2.4 GHz band were announced with speeds of 54 Mbps and 600 Mbps in 2003 and 2009 respectively. A new standard IEEE 802.11ac in the 5 GHz band with a speed of 3466.8 Mbps was announced in 2013, and IEEE 802.11ax in the 2.4/5 MHz band is expected to be released in December 2018 with a speed up to 10.53 Gbps. Laptops were followed by personal digital assistants (PDA), tablet computers, and smartphones, and all of them provide Wi-Fi connectivity.

Cellular Mobile Systems

Wi-Fi is primarily a short-range (≈ 50 m) system enabling Internet connectivity for mobile devices within the range of a fixed wireless hot spot. True mobility requires mobile computing devices such as laptops and tablets to be able to access the Internet while riding in a car or a train. For this, the computing device (laptops/tablet) must be able to access the mobile telecommunication system. The history of mobile telecommunication systems is long. Analog cellular telecommunications system appeared in 1977–78. This was based on cellular technology proposed by AT&T and Bell Laboratories. The cellular system consisted of a number of low powered, broadcast towers each covering a hexagonal cell of a few kilometres radius that meshed to cover the entire area of operation. A mobile telephone could wirelessly communicate with one base station in the cell. When the phone moved from one cell to another the signal was handed over to the adjoining cell. This enabled continuous communication when a mobile phone moved in the area of operation. Interestingly a mobile phone was invented in 1973 four years before a mobile communication system was in place for it to be used. The Federal Communication Commission (FCC) of the USA was at first reluctant to permit numerous mobile towers to be put up. After discussion, it authorized the mobile communication system to be started in 1982. A system called Advanced Mobile Phone Service (AMPS), an analog system, was installed in 1983 in Chicago – the first mobile telephone service in the USA. This analog system is now called 1G (First Generation) cellular system. Interestingly NTT of Japan had launched a cellular telephone service much earlier in 1979.

The change over from analog to digital happened in 1990s with the advent of Global System for Mobile Communication (GSM) standard. This was called 2G and allowed short messaging service (SMS) besides voice. It used circuit switching for both voice and data. The power needed to operate handsets was low and permitted long time intervals between charging the battery used by the handsets. The next evolution, called 3G, was a digital system using packet communication for data that was Internet compatible and was introduced around 1998.

The next evolution was the change over from analog to digital proposal in 1990s with the advent of Global System for Mobile Communication (GSM) standard. This was called 2G and allowed short messaging service (SMS) besides voice. It used circuit switching for both voice and data. The power needed to operate handsets was low and permitted long time intervals between charging the battery used by the handsets. The next evo-

Generation	Year Introduced	Standards (Primary)	Technology	Voice Switching	Data Switching	Data Speed-Nominal
1G	1983	AMPS	Analog	Circuit	None	N.A.
2G	1991	GSM, CDMA, EDGE, GPRS	Digital	Circuit	Circuit	200 kbps
3G	1998	UTMS,CDMA 2000, EVDO	Digital	Circuit	Packet	20 Mbps
4G	2008	LTE-A, WiMax	Digital	Packet	Packet	Max 1Gbps
			(IEEE 802.16m)			Normal 1 20 Mbps

Table 1. Evolution of mobile communication systems.

lution was 2.5G that allowed digital data transmission at a slow rate between 56 kbps and 115 kbps using General Packet Radio Service (GPRS). This was followed by 2.75G System called Enhanced Data rates for GSM Evolution (EDGE) and speeded up data transmission to 382 kbps.

The next evolution called 3G was a digital system using packet communication for data that was Internet compatible and was introduced around 1998. 3G started with data download speed of around 200 kbps. It gradually increased to 3G turbo that allowed networks based on Universal Mobile Telecommunication System (UMTS) standard with higher data speed for download around 20 Mbps. 3G systems are now being superseded by 4G systems. Whereas 3G systems used circuit switching for voice and packet switching for data, a 4G system is a fully packet switched system for both voice and data (voice over IP is used for voice data and IP for data). The wireless technologies used by the 4G system are LTE-A (Long Term Evolution – Advanced) proposed by NTT DoCoMo of Japan and WiMax (IEEE 802.16m) standard.

Table 1 briefly compares the various mobile systems [5, 6].

With the advent of 4G, we now have a fully IP based system that provides an excellent digital mobile communications infrastructure. The next evolution of mobile communication is called 5G [7]. It is expected to be available by 2020. To qualify as a 5G system it should provide 10 Gbps download speed, 1 msec end-to-end round-trip delay, and 99.999 percent availability. A version of the 5G mobile system was demonstrated in the 2018 winter Olympics in South Korea.

Smartphones

With the advent of 2G mobile communication system that could carry high-speed digital data, it was evident that the time has arrived to transform mobile phones from those used only for voice communication and short messages to ones that could have many features of a computer, namely being able to access the web, email service, and perform simple tasks such as those provided by a PDA, namely, a diary, appointment reminders etc. The first company to design a prototype mobile phone that had voice and features of a PDA was IBM which demonstrated the device in 1992 at COMDEX, a trade show. In the later part of 1990s many companies entered the field: Nokia, a Finnish company released Nokia 9000 in 1996 that combined voice phone with a PDA. Research in Motion, a Canadian company, released a PDA phone named Blackberry in 2002. Blackberry offered voice communication, text messaging, email, and web browsing. It had a physical QWERTY keyboard and was a very popular device with corporates as its mobile email service was considered very secure. It was very popular between 2002 and 2007 [8].

When 3G mobile communication system was introduced, it had packet communication for data. The time was ripe to improve mobile phones. Steve Jobs of Apple Computers introduced iPhone in January 2007 and called it a 'revolutionary and magical product'. It was sleek and feature packed. It had a touchscreen with icons and a virtual keyboard. Besides voice communication, it could perform most of the basic tasks of a laptop such as email, fax, web search, sound recording, camera, and a host of other applications. He called it a 'smartphone'.

When 3G mobile communication system was introduced, it had packet communication for data. The time was ripe to improve mobile phones. Steve Jobs of Apple Computers introduced iPhone in January 2007 and called it a 'revolutionary and magical product'. It was sleek and feature packed. It had a touchscreen with icons and a virtual keyboard. Besides voice communication, it could perform most of the basic tasks of a laptop such as email, fax, web search, sound recording, camera, and a host of other applications. He called it a 'smartphone'. Another revolutionary idea of Steve Jobs was to store in the memory of iPhones small application programs developed by independent developers. These application programs were called 'apps' and transformed the way smartphones are used. Some typical apps are a scientific calculator, a game and a browser. Icons can be created for each app and displayed on the screen of the phone. An app is invoked by touching the icon on the screen.

Meanwhile, Google released a free OS named Android for mobile smartphones. This allowed many companies to design smartphones with Android as their OS and enter the market bringing down the cost of smartphones. A huge market developed for apps built over the Android OS. Samsung, a Korean company, that uses Android, is currently one of the main competitors of Apple's iPhone. Apple followed up iPhone with a tablet computer called iPad in 2010 whose unique features are a 10 inch touchscreen, Wi-Fi connection, and all the features of a laptop. It is portable and new versions have 4G connectivity making iPads usable with mobile infrastructure both for data processing and video calls. The major disadvantage of Apple products is their high cost.

Mobile computing is a breakthrough in ICT due to the following reasons:

Wi-Fi hotspots have permitted anytime, anywhere, access of laptops and other mobile devices to the Internet that has allowed businesses to have uninterrupted access to their databases.

- The idea that the license-free ISM wireless band can be used to connect laptops wirelessly to a LAN is novel.
- Wi-Fi hotspots have permitted anytime, anywhere, access to the Internet enabling businesses to have uninterrupted access to their databases.
- The idea that computing, Internet access, voice and video calls could all be combined and provided by a hand-held mobile device from anywhere and at any time is novel.
- A number of computing jobs, such as accessing an organizations' database, application programs and communication with

colleagues can now be done while on the move from anywhere in the world. This has greatly improved the efficiency of professionals.

- Mobile devices such as smartphones are not only used for voice calls today. They are now also used to email, web search, take photographs, use new services such as WhatsApp, video call, play games, shop using e-commerce platforms, find routes with GPS and maps, among many novel applications.

- The advent of smartphones opened up the creation of a plethora of apps (application programs). Many apps appeared spawning new innovations. Apps worth billions of dollars are now sold every year.

- A new industry of manufacturing smartphones and tablet computers which are sold in hundreds of millions has now been established. Desktop computers are slowly disappearing.

Mobile devices such as smartphones are not used any more primarily for voice calls. They are now used to email, web search, take photographs, use new services such as WhatsApp, video call, play games, shop using e-commerce, find routes with GPS and maps, among many novel applications.

13. Cloud Computing

Massachusetts Institute of Technology (MIT) had an IBM 704 mainframe computer in 1957 which was used in batch mode. A user had to punch a deck of cards containing program and data, submit it to the computer centre which would put it as part of a batch of programs. The user then had to wait for almost a day to get the results of computation printed out. As a program would invariably have one or more errors, three or four attempts were needed to obtain one useful result from the computer. John McCarthy, who had just then joined MIT wrote a note to the Director of the Computer Centre suggesting that teletypewriters from the offices of the faculty be connected to the computer permitting interactive time-shared use of the computer by several persons. This would permit every user to write a program and immediately execute it, correct errors, and rerun it until the results are obtained. The idea was revolutionary at that time. A number of changes in the hardware as well as the operating system of computers were required. Fernando Corbató of the Computer Centre at MIT took up the challenge and designed a system with

John McCarthy, who had just then joined MIT wrote a note to the Director of the Computer Centre suggesting that teletypewriters from the offices of the faculty be connected to the computer permitting interactive time-shared use of the computer by several persons. This would permit every user to write a program and immediately execute it, correct errors, and rerun it until the results are obtained.

his team named Compatible Time-Sharing System (CTSS) that became operational in 1961. When computers could be time-shared, John McCarthy wondered in a speech he gave in 1961, why cannot they be converted someday as a public utility, similar to telephones, that can be accessed and used remotely by many. He opined that such a system could become the basis of a new and important industry. Technology in the 1960s was not mature enough to implement his idea.

The phrase 'cloud' came from the common practice of enclosing a group of computers connected to the Internet in the shape of a cloud in the diagrams used by computer professionals. Cloud computing has become a reality due to the rapid growth of computer and communication technologies, changes in management philosophy, and the availability of excess computing capacities with giant corporations such as Amazon and Google.

A somewhat similar computing utility became a reality in 2006 when Amazon, an e-commerce company, started a service called Amazon Web Services. This service provided computing as well as storage 'on demand' to several customers simultaneously via the Internet. Such a service is now called 'cloud computing'. The phrase 'cloud' came from the common practice of enclosing a group of computers connected to the Internet in the shape of a cloud in the diagrams used by computer professionals. Cloud computing has become a reality due to the rapid growth of computer and communication technologies, changes in management philosophy, and the availability of excess computing capacities with giant corporations such as Amazon and Google. The most important technological improvement which has given rise to cloud computing is the rapid increase in communication bandwidth at a competitive cost that has permitted fast and inexpensive access to computers connected to the Internet. Communication bandwidth available in the world has been doubling every year at constant cost. This implies that customers can use a computer halfway across the globe as if it is in the next room. Bandwidth increase enables one to transport massive amounts of data almost instantly and cost-effectively to remote computers connected to the Internet. It also enables interactive use of computers regardless of its location.

Besides rapid improvement in communications, CPU speed has also been doubling every 24 months at constant cost, and disk storage capacity has been doubling every 15 months again with no increase in cost. The increase of CPU speed has enabled interactive use of computers by multiple users in a time-shared mode, each user under the illusion that the entire CPU power is available to him or her. The increased disk capacity at no increase in cost has made it practical to store many programs and large databases of multiple users on the disks of large clusters of computers at remote locations. Hardware infrastructure of cloud computing systems now consists of a large cluster of inter-connected CPU boards and a set of disks. Such systems are called warehouse-scale computer systems.

Concurrently there have also been improvements in software systems. Time-shared and virtual machine OS introduced by IBM in the mid-1960s have steadily improved. Virtualization was an important development as it introduced a software layer on a computer's basic OS so that it mimics the programming environment of any desired computer. Virtualization is essential to allow multiple users with diverse applications requirements to use a group of servers available on the cloud. Complex virtualization software and software to manage multiple virtual machines called a 'hypervisor', has been developed by many companies during the last decade.

In addition to advances in computer hardware, software, and communications, management philosophy has been changing in the last decade. Managers have been concerned about the mounting cost of the computing infrastructure and IT staff in their organizations. It was realized by the top management that computing is not the 'core activity' of their organizations. Management trend currently is to 'outsource' non-core activities to service providers. The availability of computing as a utility allows organizations to pay service providers only for what they use. This eliminates the need for organizations to budget huge amounts to buy and maintain large computing infrastructure.

In addition to advances in computer hardware, software and communications, management philosophy has been changing in the last decade. Managers have been concerned about the mounting cost of the computing infrastructure and IT staff in their organizations. It was realized by the top management that computing was not the 'core activity' of their organizations. Management trend currently is to 'outsource' non-core activities to service providers.

Amazon had established computing and communication infrastructure to cater to the maximum demand of their e-commerce business including high demands which are seasonal, for example, during Christmas. The computing facility installed by Amazon was not fully utilized. There were long periods of only 10% utilization of the installed systems. Amazon saw a business opportunity of selling the excess computing infrastructure it had built up to other organizations that need computing power. In 2006 Amazon floated a business called Amazon Web Services which sold computing infrastructure, on demand, to customers who used the Internet to access the infrastructure.

Meanwhile, Google became a very successful company providing free information search service on the Internet and had to maintain a large computing infrastructure to cater to the maximum expected demand. It started a free email service – Gmail – in 2004 for all its customers using this infrastructure. In 2012 it launched Google drive that gave 15 GB free storage to its customers and free office productivity applications (such as word processing). Google also saw a business opportunity to sell their excess hardware capacity. It started Google compute engine as a paid cloud service in 2012. Besides Amazon and Google, many companies such as IBM and Microsoft and new companies such as Rackspace have entered the cloud computing business.

Nowadays there are a variety of cloud services such as – IaaS (infrastructure as a service) in which the provider of the service gives CPU power and storage on demand, PaaS (platform as a service) in which besides the hardware infrastructure, software infrastructure such as an OS, programming languages, and application program development and deployment tools are given to the customer, and SaaS (software as a service) in which a software applications are provided on demand.

Nowadays there are a variety of cloud services such as – IaaS (infrastructure as a service) in which the provider of the service gives CPU power and storage on demand, PaaS (platform as a service) in which besides the hardware infrastructure, software infrastructure such as an OS, programming languages, and application program development and deployment tools are given to the customer, and SaaS (software as a service) in which a software applications are provided on demand.

There are also a variety of clouds namely, private clouds, public clouds, and hybrid clouds. Private cloud consists of infrastructure maintained by a provider for the exclusive use of an organization. Public cloud is an infrastructure maintained by a provider and shared by many organizations. Hybrid cloud is the combination of a private and public cloud. In a hybrid cloud, an organization normally uses a private cloud. However, some services such as the organization's publicly accessible website may be hosted in a public cloud, or if there is an unexpected high demand for computing power or storage, some non-critical applications may be shifted to the public cloud [9].

All the developments described above led to the emergence of cloud computing which is a breakthrough in ICT due to the following reasons.

- The idea that computing facilities can be provided on demand and can be accessed by customers at anytime from anywhere on 'pay for what is used' is novel.
- It has revolutionized the way computing is done. Many organizations do not have to the incur capital expenditure to set up in-house data centres as CPU power and storage can be bought as and when required from cloud computing providers. It is no more necessary to license expensive software for long periods. The required software is often available 'on demand', and one can use it and pay only for what components of the software are used and for how long.
- It has permitted many entrepreneurs to test their ideas without having to invest in large infrastructure. If an idea is found to be marketable, the business can be quickly scaled up as more computing power can be hired on demand. This has catalysed many innovations.
- Cloud computing has allowed companies that maintain large computer centres to meet their occasional maximum demand and rent out their excess capacity when not in use.
- Big data analytics and deep learning that has sparked the use of AI in many novel applications would not have been possible but for the availability of inexpensive cloud computing power.
- An entire new industry of cloud computing providers and services based on the cloud computers has emerged.

14. Deep Learning

Deep Learning is a branch of the broad field of AI. AI has a long history that started soon after computers became reasonably powerful and reliable. In 1955 John McCarthy along with Marvin Minsky, Nathaniel Rochester, and Claude Shannon submitted a proposal to the Rockefeller Foundation, USA, to sponsor

Deep Learning is a branch of the broad field of AI. AI has a long history that started soon after computers became reasonably powerful and reliable.

a research project on ‘artificial intelligence’ (a word coined by McCarthy) in the summer of 1956 at Dartmouth College in the USA. It stated that “the study was to proceed on the basis of the conjecture that every aspect of learning and any other features of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve the kind of problems now reserved for humans, and improve themselves”. This meeting led to groups of researchers initiate work in this new exciting field of machine learning, and natural language synthesis and understanding by computers [10].

A spectacular achievement of AI was the defeat of Gary Kasparov, the World chess champion, in 1997 by a computer program that was executed in IBM’s supercomputer named Deep Blue. Another significant achievement of AI was the defeat of the human champion in a TV quiz show named Jeopardy! by IBM’s supercomputer Watson in 2011.

Early work in AI tried to discover algorithms to solve difficult problems such as game playing (e.g., chess) in which there is a vast number of choices, and one has to find a good strategy to play rather than use brute force that will be inefficient and take enormous time. A number of groups initiated research on natural language understanding by computers. Another typical application of AI was to design programs called ‘expert systems’ that would emulate and mimic human skills such as diagnosing faults in machines. These systems exhibit intelligent behaviour by mimicking a human problem solver, learn from errors, improve their behaviour, and also explain how the programs arrived at the conclusions. There have been continuous innovations in this method of problem solving by AI. A spectacular achievement of AI was the defeat of Gary Kasparov, the World Chess Champion, in 1997 by a computer program that was executed in IBM’s supercomputer named Deep Blue. Another significant achievement of AI was the defeat of the human champion in a TV quiz show named Jeopardy! by IBM’s supercomputer Watson in 2011. The uniqueness of this win was the requirement for the computer to ‘understand’ natural language, have an encyclopaedic knowledge of a huge number of fields, and analytic abilities [11].

At the same time, another group of researchers started working solving AI problems by modelling the human brain on a computer. The brain model was based on the work of biologists. These ‘brains’ were trained to solve problems by trial and er-

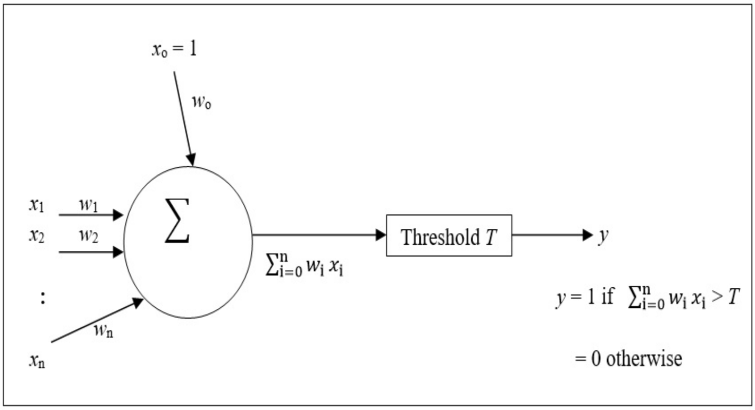


Figure 1. An artificial neuron.

ror after examining a large number of examples and solutions. This method has been found to be appropriate for ill-defined or unstructured problems. A typical example is recognizing a face which is done very easily by humans. On the contrary, it is very difficult to design algorithms for computers to recognize faces.

In 1943 biologists Warren McCulloch and Walter Pitts had modelled a human brain as a collection of neurons. The human brain is thought to have 100 billion nerve cells called neurons that send signals to other neurons using what are known as synapses. Ex-

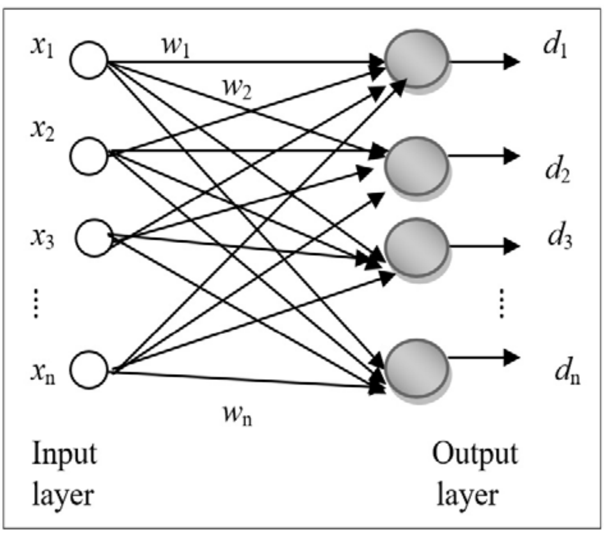


Figure 2. A perceptron. Training inputs x_is are given to the input layer and the desired outputs d_is are obtained in the output layer. Each link has a weight. The weights are adjusted by trial and error.

ternal stimuli from organs (smell, touch, sight, sound, etc.) are accepted by dendrites that create electrical impulses that travel through the network of neurons. The neurons decide whether to send the impulses to other neurons to take necessary action or to ignore them.

McCulloch–Pitts neuron model was used in 1957 as the basis to model an artificial neuron by Rosenblatt, a psychologist (see *Figure* 1). The input to an artificial neuron is a set of bits x_i that are multiplied by a set of numbers called weights w_i. There is one input called a bias input x_0 that is set to 1. If $\Sigma^{n}_{(i=0)} w_i x_i$ is greater than a threshold T, the neuron fires and gives an output 1 else a 0. He built a machine called a 'perceptron' using a network of artificial neurons called the artificial neural network (ANN). A perceptron had an input layer, a single layer of artificial neurons and outputs of this layer (see *Figure* 2). Rosenblatt trained the perceptron to obtain a pattern defined by a vector d_i at the output layer by adjusting the weights w_i systematically in such a way that the output vector d_i equals the desired binary vector for given inputs x_i s at the input layer. As building computers was costly in the 1950s, Rosenblatt built a perceptron with only a single layer of artificial neurons. Perceptron was oversold, particularly by the popular press, with the implication that it could solve many difficult recognition problems [12].

The best formulation of how the hidden layers improve feature recognition by adjusting the weights of the connections using an algorithm called back propagation and gradient descent was published by Rumelhart, Hinton, and Williams in 1985.

In 1969, Marvin Minsky and Seymour Papert wrote a book titled *Perceptrons* and showed that a single layer perceptron had serious limitations. They formulated a theory describing how perceptrons computed and showed that it was impossible to train a perceptron to learn the exclusive OR function. The book was a damper on research in ANN.

Perceptron used a very simple model of a neuron and used only a single layer of neurons which was a serious limitation. New models of ANNs with a more sophisticated model of neurons (see *Figure* 3) and multiple layers of neurons between the input and the output layers called 'hidden layers' were proposed independently by many researchers (see *Figure* 4). Hidden layers could find 'features' in the input data and allow succeeding layers to work with these features rather than with raw data. However, the best formulation of how the hidden layers improve feature recognition by adjusting the weights of the connections using an algorithm called backpropagation and gradient descent was published by Rumelhart, Hinton and Williams in 1985. In 1989 Yann LeCun with co-authors showed how multilayered ANN using back propagation algorithms could recognize handwritten postcodes called ZIP codes used in the USA (similar to PIN codes in India).

After these initial successes in pattern recognition research, further work on ANN slowed down as it was not possible to train neural networks (i.e., to find correct weights of connections) to recognize more difficult inputs such as speech. Methods of training ANNs with more than two hidden layers were also found to be difficult and time-consuming. The problem of training many layered networks was solved in a paper titled 'A Fast Learning Algorithm for Deep Belief Nets' by Hinton, Osindero and Teh in 2006. They showed that many-layered ANN, could be trained well if the weights of the links to the neurons are initialized systematically rather than using random weights. The system used by them was to train each layer using unsupervised learning (i.e., dividing outputs into clusters with similar properties) and use these weights as the starting point for supervised learning (i.e., learning to match

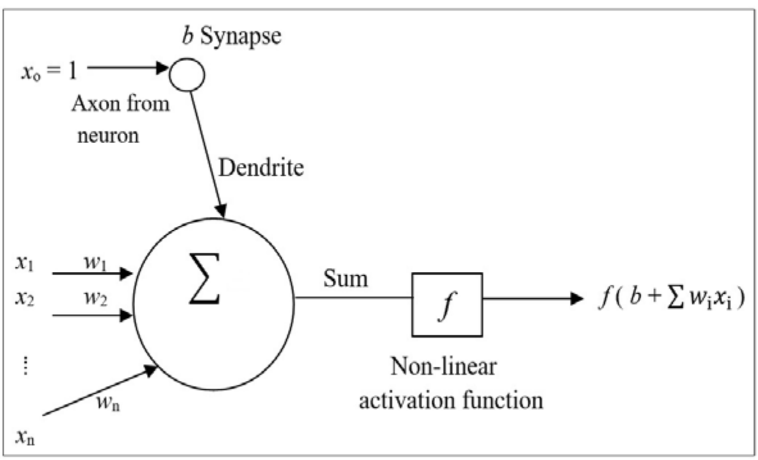
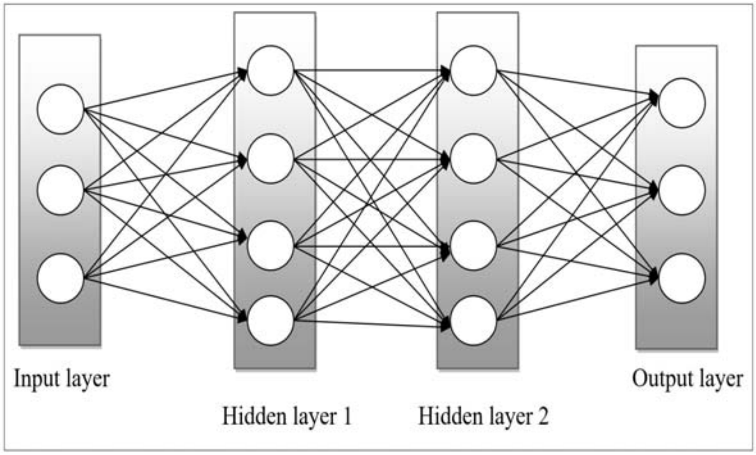


Figure 3. A more sophisticated model of a neuron. The activation function f is a nonlinear function. If $f(b + \Sigma w_i x_i) > T$ (a pre-set threshold) the neuron fires and the output is 1, else it is 0. (Source: http://cs231n.github.io/neural-networks-1/)

Figure 4. A multilayered neural network. (Source: http://cs231n.github.io/neural-networks-1/)



pre-specified outputs).

Meanwhile, researchers postulated and experimented with many types of ANNs variously known as convolutional neural nets, recurrent neural nets, neural nets with long short-term memory, and more recently generative adversarial networks. The main differences between these ANNs are the configuration of the arcs connecting the neurons and the activation function. These networks use different training regimes and exhibit specific strengths in their recognition properties.

Coupled with the appearance of fast special purpose processors was the emergence of 'big data'. A large number of real-life data such as huge number of emails, Facebook data with pictures, YouTube videos, images loaded by users in hosting sites such as Flickr, and digitized speech data from audiobooks are now available with the explosion of the use of the Internet.

The process of training complex multilayered ANNs or 'deep networks' was too slow for practical use until the appearance of faster hardware specifically, general purpose graphics processing units (GPGPUs). GPGPUs did computation in parallel and could be used to solve linear simultaneous equations much faster than general purpose microprocessors. Coupled with the appearance of fast special purpose processors was the emergence of 'big data'. A large number of real-life data such as huge number of emails, Facebook data with pictures, YouTube videos, images loaded by users in hosting sites such as Flickr, and digitized speech data from audiobooks are now available with the explosion of the use of the Internet. This provides real-life training data for researchers to train deep networks. Training deep networks with a huge amount of real-life multimedia data is called deep learning [13].

Deep learning got an impetus when companies such as Google and Microsoft started using deep neural networks along with improved training algorithms to solve real-life problems such as speech recognition, image classification, generating images similar to given images, predicting next letters of words using the previous sequence of letters, generating next word in a sentence while typing a sentence, language translation, handwriting recognition, and similar difficult problems that were earlier intractable. One of the spectacular successes was a deep learning program called AlphaGo that beat the World champion in GO, a very complex East Asian board game, in October 2015. AlphaGo used a tree search algorithm to find its moves based on knowledge previously learned by a deep learning algorithm by extensive training, by examining a large number of games played by humans, by playing against humans, and by playing against earlier versions of itself trained to play GO. It was a big challenge for computers till then to beat humans in GO as the number of moves in each board position is enormous and very clever strategies are required to play the game.

The combination of big data, GPGPUs, and the huge computing resources available in the 'cloud' have made deep learning algorithms solve many practical problems. In spite of its ability to solve difficult recognition problems, deep learning has some major flaws. One of the most glaring flaws is the inability of deep learning networks to explain how the problem was solved. The other is the non-robustness in the solution. In other words, a small perturbation of inputs can sometimes lead to incorrect recognition. These flaws can lead to difficult ethical and legal problems in cases such as when a driverless car using deep learning meets with an accident leading to grievous injuries or death of a human.

One of the most glaring flaws is the inability of deep learning networks to explain how the problem was solved. The other is the non-robustness in the solution. In other words, a small perturbation of inputs can sometimes lead to incorrect recognition. These flaws can lead to difficult ethical and legal problems in cases such as when a driverless car using deep learning meets with an accident leading to grievous injuries or death of a human.

To summarise, the idea of using ANNs to solve diverse problems, even though very old and often discredited, has now become practically useful. This has happened due to the availability of large volume of training data combined with the availability of fast GPGPUs, better neuron models, many layers of neurons with diverse architecture, and superior training algorithms.

Deep learning is a breakthrough in ICT due to the following reasons:

- The idea of modelling the way the human brain works using ANNs with many layers of neurons and systematically training them is novel.
- The use of large ANNs with many layers of neurons has changed the way many AI problems are solved.
- Difficult problems that need solutions such as automatic translation of one language to another, recognizing faces, recognizing handwriting, and similar tasks are now being solved with deep neural networks.
- It is expected to herald a change in the way AI is used in the society. Driverless cars are in the horizon. Robots with deep learning are expected to carry out many tasks previously considered to be the exclusive domain of humans such as diagnosing illness from medical images, answering routine user queries, and real-time translation of one natural language to another.
- It is feared that AI with deep learning will change the nature of routine work causing disruption in employment patterns.
- A new industry of designing specialised hardware to build deep learning networks, designing algorithms to train deep networks for many new tasks, and discovering novel applications of deep learning networks is emerging.

Driverless cars are on the horizon. Robots with deep learning are expected to carry out many tasks previously considered to be the exclusive domain of humans such as diagnosing illness from medical images, answering routine user queries, and real-time translation of one natural language to another.

Conclusions

In this article, I have identified and described five breakthroughs (10–14) in ICT since 1955 (besides the others in the previous articles) that I have selected based on the six criteria listed in the introduction. Minicomputer was a possible candidate on my list, but I did not include it in spite of its meeting a number of criteria as it had a short life and did not lead to societal changes. Cryptography and its applications was another technology that I could have included, but I was not convinced that it had the same impact as other breakthroughs in my list. A common characteristic of all the breakthroughs in ICT is the long gestation period. Very good ideas took a long time to be realised as products as a number of different technologies had to mature. This series of articles would be of historical interest to many young students and professionals, and I trust it will kindle a debate among senior professionals on my selection of breakthroughs.

A common characteristic of all the breakthroughs in ICT is the long gestation period. Very good ideas took a long time to be realised as products as a number of different technologies had to mature.

Suggested Reading

[1] History of Search Engines: From 1945 to Google Today, www.searchenginehistory.com

[2] T B Rajashekar, Web Search Engines – How to Get What You Want from the World Wide Web, *Resonance*, Indian Academy of Sciences, Vol.3, No.11, pp.40–53, 1998.

[3] Lawrence Harte, *Introduction to IP Video; Digitization, Compression and Transmission*, Althos, USA, 2007.

[4] V Rajaraman, *Introduction to Information Technology*, 3rd Edition, PHI Learning, New Delhi, 2018, pp.41–88 for introduction to image, audio and video acquisition, digitization, and compression.

[5] Richard Frenkiel, A Brief History of Mobile Communication, http://www.winlab.rutgers.edu/ narayan/Course/Wireless_Revolution/vts%20article.pdf

[6] History of Mobile Communication, https://www.tutorialspoint.com

[7] Brad Reed, A Brief History of Smartphones, *PC World*, June 18, 2010. https://www.pcworld.com/article/199243/a_brief_history_of_smartphones.html

[8] U N Kar and D K Sanyal, A Sneak Peek into 5G Communications, *Resonance*, Indian Academy of Sciences, Vol.23, No.5, pp.555–572, 2018.

[9] V Rajaraman, Cloud Computing, *Resonance*, Indian Academy of Sciences, Vol.19, No.3, pp.242–258, 2014.

[10] V Rajaraman, John McCarthy – Father of Artificial Intelligence, *Resonance*, Indian Academy of Sciences, Vol.19, No.3, pp.198–207, 2014.

[11] Raman Chandrasekar, Elementary? Question Answering, IBM's Watson, and the Jeopardy! Challenge, *Resonance*, Indian Academy of Sciences, Vol.19, No.3, pp.222–241, 2014.

[12] Andrey Kurenkov, A Brief History of Deep Learning, www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning-

[13] Jason Brownlee, What is Deep Learning? Machine Learning Mastery, August 16, 2016. https://www.machinelearningmastery.com/what-is-deep-learning/

Address for Correspondence
V Rajaraman
Supercomputer Education & Research Centre
Indian Institute of Science
Bengaluru 560 012, India.
Email: rajaram@iisc.ac.in