5G and Beyond The Future of Connectivity

Dr. MANVI SHARMA

Xoffencer

5G AND BEYOND: THE FUTURE OF CONNECTIVITY

Dr. Manvi Sharma

Xoffencer

www.xoffencerpublication.in

ISBN-13: 978-81-970527-2-9 (paperback)

Publication Date: 25 February 2024

MRP: ₹550/-

ISBN

9 788197 052729

Published by:

Xoffencer International Publication

Behind shyam vihar vatika, laxmi colony

Dabra, Gwalior, M.P. – 475110

Cover Page Designed by:

Satyam soni

Contact us:

Email: mr.xoffencer@gmail.com

Visit us: www.xofferncerpublication.in

Author Details

Dr. Manvi Sharma

Dr. Manvi Sharma is Ph.D in VLSI domain from Thapar Institute of Engineering and Technology, Punjab, India, M.Tech in ECE from Indira Gandhi Institute of Technology (now IGDTUW) and B. Tech in EI from APJ Abdul Kalam Technical University, India. She is also certified in Artificial Intelligence and Machine Learning from IIT Kanpur. She has several Research papers in SCI, Scopus indexed journals and conferences, also has some patents and copyrights on her name. Her area of interests are Communication systems, Mechatronics and IoT.

Preface

The text has been written in simple language and style in well organized and systematic way and utmost care has been taken to cover the entire prescribed procedures for Science Students.

We express our sincere gratitude to the authors not only for their effort in preparing the procedures for the present volume, but also their patience in waiting to see their work in print. Finally, we are also thankful to our publishers **Xoffencer Publishers, Gwalior, Madhya Pradesh** for taking all the efforts in bringing out this volume in short span time.

Abstract

Research on fifth-generation wireless communications is outperforming the expansion of wireless communications as a result of ever-increasing data amounts, higher data rates, and the proliferation of mobile devices. As a result of the adoption of new technology to equip future millimeter band wireless communication systems through massive multi-input multi-output (MIMO) with extreme base station and device densities and an unprecedented number of antennas, 5G networks are characterized by ubiquitous connectivity, extremely low latency, and very high-speed data transfer. These characteristics are the result of the adoption of new technology. A one-of-a-kind catalyst for kindling new sorts of information technology and driving innovation, 5G has developed into a revolutionary technology. Not only has 5G arisen as a separate motivator, but it has also evolved as a new motor for stimulating industrial upgrading and maintaining economic growth. Because 5G is more than just an improvement over previous generations, it requires a new category of materials that includes metals, polymers, ceramics, functional materials, nanoparticles, and composites. Materials that fall into this category include nanoparticles. Because of the development of 5G network systems and the widespread use of 5G terminals, there has been a significant increase in the demand for particular components to meet the growing demand. A number of components are included in this category. These components include antennas for the base station, filters, shields against electromagnetic interference, protective coatings and sealants, solutions for thermal management, and high-frequency printed circuit boards. It should come as no surprise that technological advancements are still being made, with improved plans being formed for its application in a world that is totally connected, as well as upgrades to suit the demands of the future. All of this encompasses 5G and beyond. These initiatives include the development of innovative technical systems as one of their tasks. It takes a significant amount of effort on the part of scientists and engineers in order to produce advancements in the design of components, the manufacture of materials, and other relevant technical improvements.

Contents

CHAPTER 1

INTRODUCTION

1.1 EVOLUTION OF 5G TECHNOLOGY

The capabilities of mobile networks have been significantly enhanced as a result of the introduction of 5G radio. Over the course of their existence, mobile networks have primarily connected the mobile devices of users, which include computers, tablets, and smartphones. The introduction of 5G infrastructure will put the constraints of the existing mobile broadband service to the test in terms of the data speeds, capacity, and availability of the service.

Additionally, 5G will make it possible to provide new services, such as linking industrial Internet of Things devices and making it easier to communicate with mission-critical systems. For 5G, the bar has been set exceptionally high, with promises of data rates of up to 20 gigabits per second (Gbps), capacity upgrades of up to a thousand times, ultralow latency, great dependability, and flexible platforms for device connectivity. It is anticipated that 5G would alter practically every facet of society due to its emphasis on enhancing efficiency, productivity, and safety. The majority of the time, the telecom operators and manufacturers developed and constructed 4G networks in order to accommodate the situation of smartphone usage. The concept of 5G networks is already attracting a great deal of interest from a variety of other stakeholders, including other industries and NGOs. The aforementioned groups are eager to acquire knowledge regarding 5G networks in order to make the most of the opportunities that these networks present. 4G was developed with the primary intention of making human interaction easier. With 5G, every gadget will be connected to one another.

Considering that the Nippon Telegraph and Telephone Public Corporation (NTT) introduced the very first cellular mobile communication service in December 1979, mobile communication technology has progressed in tandem with other technical advancements ever since. Voice calls were the most frequent mode of communication throughout the first two generations of mobile networks, which were known as 1G and 2G. Additionally, basic e-mail was additionally available during this time period. On the other hand, beginning with the third generation of mobile technology, mobile devices will be able to share data communications such as "i-mode" and multimedia information, which includes files of audio, video, and photographs. As a result of the advent of high-speed connection capabilities that exceed 100 megabits per second (Mbps) utilizing the Long-Term Evolution (LTE) technology, the popularity of smartphones has experienced a remarkable spike ever since the fourth generation (4G) of the technology was established. In addition to that, there are a number of new styles of communication that involve the use of multimedia.

The present iteration of 4G technology is still in the process of developing into LTE-Advanced, which will enable transmission speeds that are very near to one gigabit per second. The Internet of Things (IoT) and artificial intelligence (AI) are two examples of newly developed social and industrial services that are expected to be made available by the next generation of wireless networks, which is expected to be 5G and 6G. According to the theory, 5G and 6G will also increase the quality of the multimedia communication services that are now available by providing new technical features such as high capacity, low latency, and enormous connectivity.

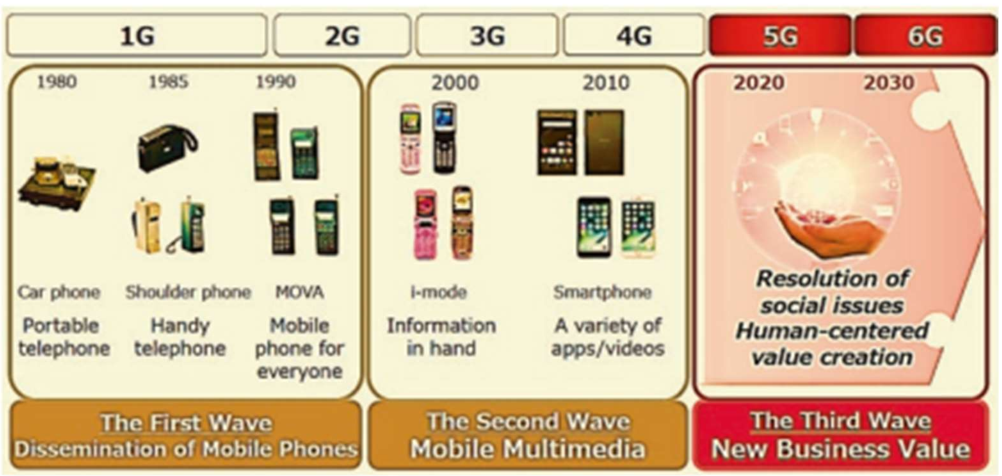


Fig. 1.1 Advancements in Mobile communications technology and services

Source: Adapted from Nakamura 2020 with permission from IEEE

As can be seen in Figure 1.1, the mobile communication system undergoes a significant technological update every ten years. On the other hand, the services that are offered by mobile communications go through significant modifications in cycles that typically span twenty years. As a result of this, we can anticipate that the "Third Wave" that 5G ushered in will expand into an even greater wave in the 2030s. This is because of the breakthroughs that have been made in 5G and sixth-generation (6G) technology, which will come to the advantage of both enterprises and the general population. The capabilities of mobile networks have been significantly enhanced as a result of the introduction of 5G radio.

Over the course of their existence, mobile networks have primarily connected the mobile devices of users, which include computers, tablets, and smartphones. The introduction of 5G infrastructure will put the constraints of the existing mobile broadband service to the test in terms of the data speeds, capacity, and availability of the service. Additionally, 5G will make it possible to provide new services, such as linking industrial Internet of Things devices and making it easier to communicate with mission-critical systems. For 5G, the bar has been set exceptionally high, with promises of data rates of up to 20 gigabits per second (Gbps), capacity upgrades of up to a thousand times, ultralow latency, great dependability, and flexible platforms for device connectivity.

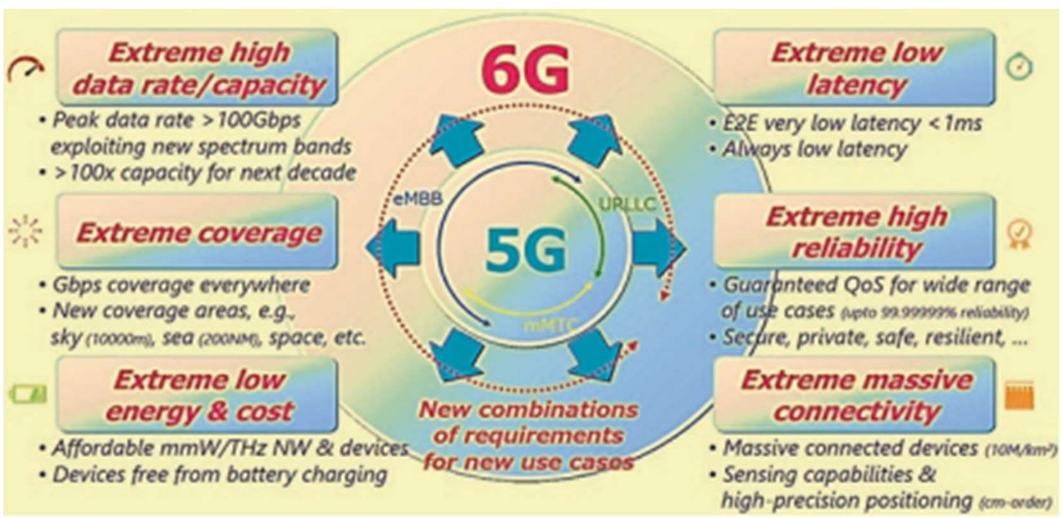


Fig. 1.2 Requirements for 6G wireless technology

Source: Adapted from Nakamura 2020 with permission from IEEE

It is anticipated that 5G would alter practically every facet of society due to its emphasis on enhancing efficiency, productivity, and safety. The majority of the time, the telecom operators and manufacturers developed and constructed 4G networks in order to accommodate the situation of smartphone usage. The concept of 5G networks is already attracting a great deal of interest from a variety of other stakeholders, including other industries and NGOs. The aforementioned groups are eager to acquire knowledge regarding 5G networks in order to make the most of the opportunities that these networks present. 4G was developed with the primary intention of making human interaction easier. With 5G, every gadget will be connected to one another.

The prerequisites that must be fulfilled before 6G can make full use of wireless technology through the evolution of 5G are depicted in Figure 1.2. In addition to raising the criteria for 5G itself, the criteria for 5G have been expanded to include other requirements that were not previously taken into consideration more generally. In addition, similar to the situation with 5G, it is not necessary to fulfil all of the needs simultaneously; instead, new combinations of requirements will be necessary for the development of future new use cases.

1.2 5G TECHNOLOGY COMPONENTS

The 5G wireless communications system is a consolidated system that integrates a wide variety of radio access technologies. The individual components of the primary technological components are depicted in Figure 1.3.

1.2.1 5G Spectrum

The current 4G network is unable to support certain frequencies and spectra, but the 5G network will be able to do so. Fifth-generation mobile radio technology is designed to function on any frequency band between 400 MHz and 90 GHz. This range of frequencies is ideal for mobile radio. The low bands are required for coverage, but the high bands are required for high data rates and capacity. Both of these requirements are necessary. In the initial 5G networks to go live, Time Division Duplex (TDD) frequencies were utilized between 2.5 and 5.0 GHz, Frequency Division Duplex (FDD) frequencies were utilized below 2.7 GHz, and Time Division Duplex (TDD) millimeter wave frequencies were utilized between 24 and 39 GHz. Figure 1.4 illustrates the three primary spectrum possibilities that are available. The millimeter wave spectrum may have a large bandwidth that can reach up to one to two gigahertz at frequencies that are higher than twenty gigahertz. This makes it possible to increase the data rate to anywhere between 5 and 20 gigabits per second, which is ideal for exceptionally high-capacity mobile broadband.

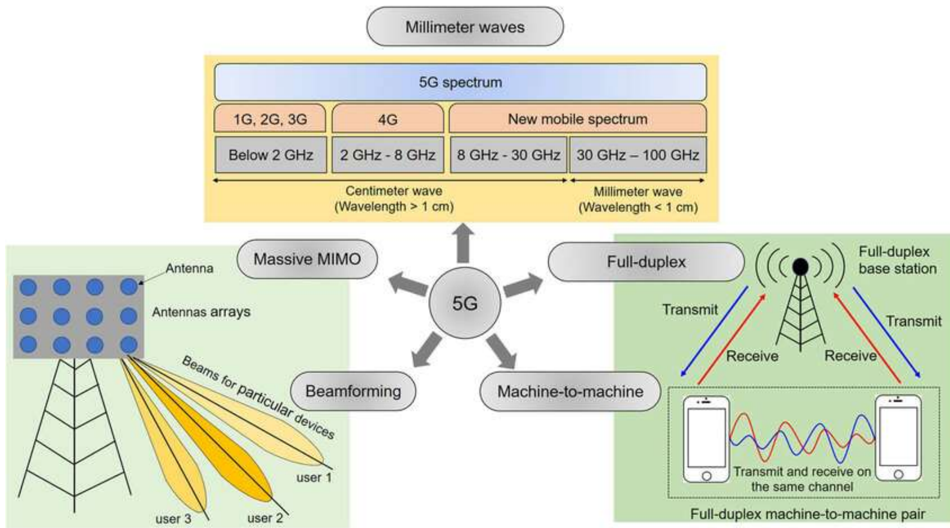


Fig. 1.3 Key 5G Technology components including millimeter waves, Network slicing, Massive MIMO, Beamforming, full-duplex, and machine-to-machine cloud-optimized architecture applied in 5G wireless communications making it smarter, faster, and more efficient

Source: Modifed from Hao et al. 2020 with permission from Walter de Gruyter and Company

Millimeter waves, in contrast to conventional bandwidths, are perfect for transmission across shorter distances because of their extremely narrow bandwidth. In proportion to the magnitude of the attenuation that represents the existing situation, the

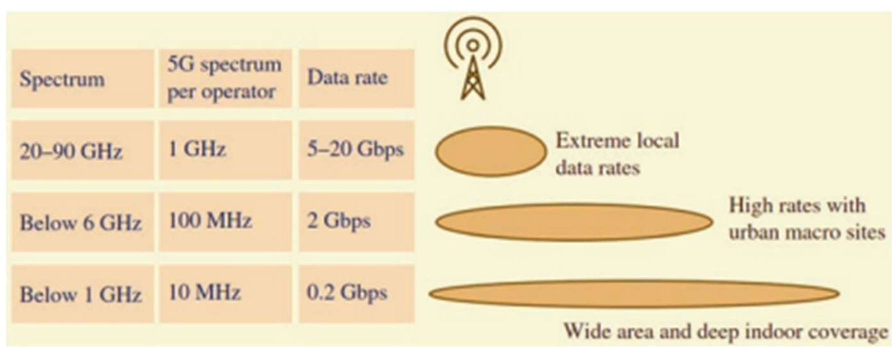


Fig. 1.4 Utilization of all Spectrum Options by 5G

Source: Adapted from Holma et al. 2019 with permission from John Wiley and Sons)

In the case of waves that travel shorter distances, the amplitude of the signal decreases as the distance between the two points increases. In addition, millimeter wave signals are far more sensitive to obstructions than microwave signals are. This is because millimeter wave signals have a lower diffraction and a higher specular transmission. When compared to other types of waves, millimeter waves are more likely to be obstructed by structures such as buildings, trees, and other vegetation, as well as by adverse weather conditions. Metal objects in close proximity have the potential to simply interfere with the 5G wireless signal that is being transmitted by mobile terminals in the vicinity. Within a distance of hundreds of kilometers, macro base stations are required to be situated in order for 4G wireless communications to function properly.

However, when it comes to wireless communications at the 5G level, base station towers have a significantly shorter range and can be designed to be smaller in size. If we take into consideration the limited range of 5G, it is likely that macro base stations will continue to link in the sub-6 GHz band. In addition to this, a multitude of nanoscale base stations are established and connected; these stations are composed of nanodevices and operate on millimeter wave frequencies. 5G relies on a flood of smaller base stations to complement the capabilities of traditional cellular towers, in contrast to prior generations of wireless communications, which relied on base stations that were dramatically growing. An effective solution for the millimeter wave transition is provided by these compact base stations that are grouped in close proximity to one another.

Because of this, it is possible that in the future, users may be able to purchase 5G wireless communications that offer unlimited data storage, call volumes, and data broadcasts. Because of the tremendous advancements that have been made in the signaling, administration, and accounting systems of 5G communications, these networks are now able to handle a wide variety of applications that go beyond mobile broadband.

1.2.2 Massive Multiple-Input Multiple-Output (MIMO) Antennas

Massively multiple input multiple output (MIMO) systems are characterized by the utilization of antenna arrays in macro base stations. These systems make it possible to precisely focus the energy that is transferred to mobile devices. It is necessary for a massive MIMO system to have a network consisting of hundreds of antennas in order to guarantee that a large number of user terminals will receive uninterrupted service. In a huge MIMO system, every user has the ability to alter the amount of power they broadcast in accordance to the number of antennas that are present at the macro base stations. Their system is now capable of achieving the same degree of performance as an equivalent system with a single input and a single output as a result of its implementation. With the advent of large-scale MIMO, gains in connection reliability and coverage have been made possible, along with advances in capacity, data speed, and energy efficiency that are many orders of magnitude better than those that were previously possible. By taking the numerical average of terminals that have been randomly arranged, it has been shown that approximately 95% of terminals are capable of handling a throughput of 21.2 Mb/s each. On the other hand, array antennas are capable of delivering a downlink throughput of approximately 20 Mb/s for a thousand terminals.

However, beam steering that makes use of millimeter waves and enormous MIMO antenna arrays would necessarily result in an increase in the number of major complaints from antenna systems. On the other hand, the common antennas that are present in portable devices, such as 4G terminals, are not equipped to accommodate millimeter waves. There is a possibility that the antennas for 5G wireless communications will be impacted by the components that are located in close proximity to them, such as the battery and the protective cover, when they are integrated into a physical device like as a phone. When operating at frequencies that span the low band to the high band, 5G antennas can be reduced to the size of micrometers or even nanometers on the spectrum. As a result, it is not completely out of the question that portable devices could integrate an extremely large number of antennae.

1.2.3 Network Slicing

It is necessary for the physical and protocol layers of 5G to have designs that are flexible in order to accommodate a variety of use cases and frequency bands, while simultaneously optimized for the transfer of energy and spectrum. Because of network slicing, each of the several 5G services will have its own specialized virtual network portion. This will be achievable. By having the capability to slice data in this manner, operators are able to serve a variety of use cases and corporate clients without having to construct specialized networks. In particular, 5G networks are intended to be able to deal with conditions that are exceedingly varied and demanding in terms of latency, throughput, capacity, and availability. From the perspective of a standard network infrastructure architecture, the concept of network slicing provides a solution that is capable of satisfying the requirements of all use cases. A graphical representation of the idea of network slicing may be found in Figure 1.5. There are a wide variety of devices that are able to share the same network infrastructure. These devices include smartphones, tablets, virtual reality connections, personal computers, crucial remote controls, and automotive connectivity.

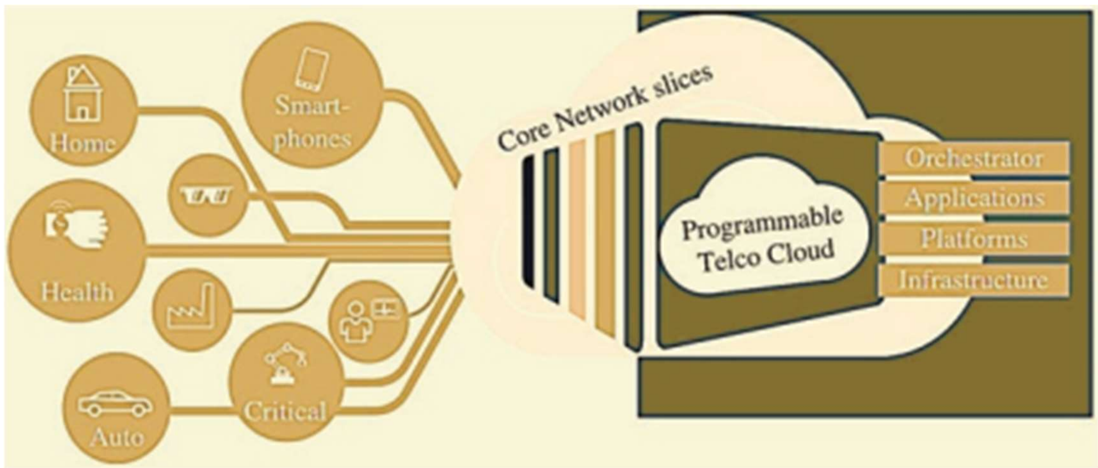


Fig. 1.5 Network slicing concept

Source: Adapted from Holma 2019 with permission from John Wiley and Sons

1.2.4 Dual Connectivity and Long-Term Evolution (LTE) Coexistence

A flawless deployment and user experience are guaranteed by 5G, which is the first radio system to be thoroughly integrated with the legacy radio network. By utilizing dual connectivity, a user equipment (UE) that is capable of 5G can simultaneously connect to both a 5G radio network and an LTE radio network. Dual connection is a solution that could be utilized in this situation. Although it is possible to install 5G on its own, it is more likely that it will be introduced concurrently with LTE in the first phases of its adoption. In theory, a device that supports 5G technology is capable of simultaneously connecting to both 5G and LTE networks. It is possible that the utilization of dual connection might simplify the process of implementing 5G, boost the data rate of consumers, and make it more dependable. Additionally, 5G is intended to work in conjunction with LTE networks, which gives rise to the possibility of frequency sharing and simplifies the process of spectrum reframing.

1.2.5 Support for Cloud Implementation and Edge Computing

In most cases, radio networks have been established with a dispersed architecture, which indicates that all radio processing takes place in close proximity to the antenna of the base station. According to the data, it would appear that the core network is designed in a very centralized manner, with a few number of essential sites throughout the network. Core processing will be more dispersed in future architectures in order to reduce latency, whereas radio processing will be more centralized in order to encourage scalability. This will enable future architectures to be more scalable. The capabilities of mobile networks are expanded as a result of this advancement, such that they now include access to edge cloud servers. Once that is accomplished, these server sites will be able to handle radio and core network operational capabilities.

As can be seen in figure 1.6, the architecture saw development over the course of time. The 5G specifications intend to strengthen the radio cloud by presenting a brand-new interface that will be implemented within the radio network. By utilizing this interface, it is possible for a centralized edge cloud site and a dispersed radio frequency (RF) site to simultaneously share capabilities with one another. Scalability of the network, which is made possible by the use of radio cloud technology, is beneficial in situations such as the addition of a large number of devices that are connected to the Internet of Things.

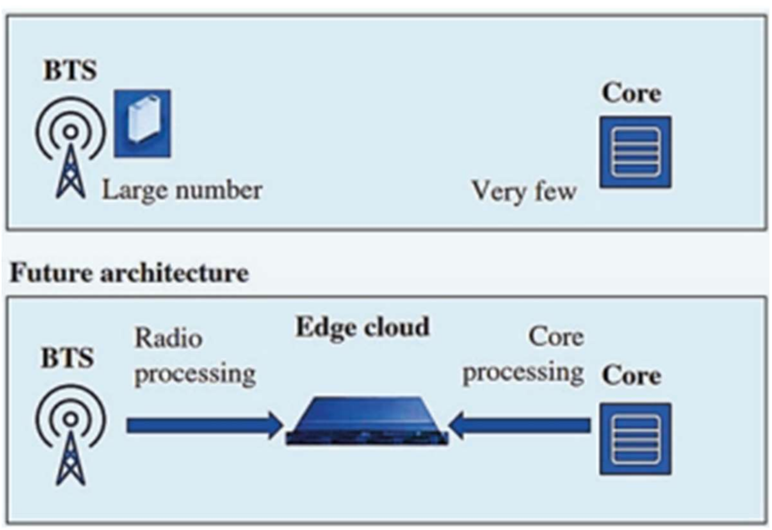


Fig. 1.6 Network architecture evolution

Source: Adapted from Holma 2019 with permission from John Wiley and Sons

A method that is referred to as "edge computing" can be described as the process of bringing computing resources closer to the edge of a network. Mobile computing, also known as multiaccess edge computing, can be conceptualized as the hosting of a small data center that is located significantly closer to the user in the event of advanced business or industrial use environments. In addition to the possibility that this data center is established on the premises itself, it is also feasible that it is located within the main office of a telecom business. It is anticipated that applications that require excessive volumes of data and processing will be able to reap the benefits of edge computing. These applications include those that benefit from extremely low latency. Among the many applications of this technology, some examples include virtual and augmented reality, industrial applications, and robotics control. In accordance with the anticipations, manufacturers operating in the industrial sector can reap the benefits of digital twin technologies by utilizing applications that are enabled by edge computing.

1.3 MATERIALS SOLUTIONS FOR 5G HARDWARE SYSTEM INTEGRATION

It is conceivable to categorize hardware solutions that are capable of functioning in 5G as either User Equipment (UE), Customer Premise Equipment (CPE), Base Station or infrastructure. These classifications are entirely possible. Each of these aspects—the requirements, the dimensions, and the constraints on power—is one of a kind. Miniaturization and decreased power consumption are the driving forces behind user equipment (UE), in contrast to infrastructure gear, which is driven by the desire to meet high gain, communication range, enormous MIMO demands, and broadband needs. According to the findings of a comparison of the technological indicators for these categories, the results are presented in Table 1.1.

Beginning with the early 2G and 3G systems and continuing all the way up to the most recent 5G and 4G LTE systems, the environment of base transmit systems (BTSs) has seen a significant transformation. In order to support 5G, it is necessary to reallocate the spectrum that is currently available, assign new spectrum at frequencies lower than 6 GHz, and allocate spectrum in the millimeter-wave (mm-wave) frequency band. Additionally, the requirements for radio frequency (RF) equipment that are necessary for 5G base stations are substantially different from those that were necessary for earlier generations. 5G typically makes use of higher-order active antenna systems (AASs), which are composed of a multitude of transmit and receive paths (32, 64, and higher) coupled to an array of antenna components. This is in contrast to the 4G deployments that came before it, which primarily concentrated on the range of two to eight transmit routes per sector.

In a perfect world, each antenna would come equipped with its own power amplifier (PA), low noise amplifier (LNA), switch, phase shifter, and variable gain amplifier (VGA). This would be the case from the beginning. In an ideal scenario, the antennas will be connected to RF front-end devices that are capable of managing anything from four to twenty-four antennas on a single device. Eventually, a number of these radio frequency (RF) front-end components will be paired with radio frequency (RF) transceivers, which are the devices that are accountable for the creation, modulation, and demodulation of signals. The implementation of 5G has resulted in the introduction of novel and distinctive requirements for radio frequency (RF) front-end electronics. This has necessitated the development of novel solutions and technologies in order to cater to the requirements of each sector with regard to the capabilities of RF transmit amplification and reception (Rx) hardware.

In a perfect world, each antenna would come equipped with its own power amplifier (PA), low noise amplifier (LNA), switch, phase shifter, and variable gain amplifier (VGA). This would be the case from the beginning. In an ideal scenario, the antennas will be connected to RF front-end devices that are capable of managing anything from four to twenty-four antennas on a single device. Eventually, a number of these radio frequency (RF) front-end components will be paired with radio frequency (RF) transceivers, which are the devices that are accountable for the creation, modulation, and demodulation of signals. As a consequence of this, radio frequency (RF) front-end electronics are being subjected to new and distinct demands brought about by 5G. This innovation is forcing the industry to search for solutions and technologies that are capable of fulfilling the functionality of RF transmit amplification and reception (Rx) hardware capabilities.

Table 1.1 Technology metrics for 5G communication systems

Merits	**Downlink (Base station)**	**Uplink (CPE)**	**User equipment (UE)**
Antenna and module size	70 x 70 x 2.7 mm'	450-1400 mm= Substrate thickness: IS mm	20x5 x2 mm' (Quakomm. QTM052)

Number of antennas	64-256	16-32	4-8
PA power	33 dBm	19 dBm	10-15 dBm (6-8 dBm usually)28 mm CMOSDC power for four elements: 360-380 mW Power consumption per channel <100 mW
Antenna gain	27 dBi EIRP of 50 dBm (IBM)	18-21 dBi for 8 x 4 way patch antenna with grounded rings	20 dal for 2 x 2 antenna array
End•o-end loss	-	-	2.5 dB
Pathloss	135 dB	135 dB	-
Received power	-75 dBm	-90 dBm	-140 dBm•
SNR per RX element	5 dB	-15 dB	6.2 dB'
Rx gain	21 dBi	27 dBi	-10 dB
Rx SNR after gain	26 dB	12 dB	-

Working at a high frequency (above 26 GHz), which is also referred to as millimeter-wave 5G, will have the potential to bring a number of distinct benefits that 5G does not now offer.

In order to adequately handle such a high frequency, it is required to use novel materials and a diverse spectrum of device designs. Low-loss materials that have a tiny dielectric constant and a moderate tan loss, on the one hand, have more possibilities at high frequencies owing to the increasing quantity of considerable transmission loss. This is because of the fact that the dielectric constant is small. A further objective of contemporary package designs is to minimize the amount of signal loss that occurs by including passive components in every corner of the box. Driving at a high frequency, on the other hand, results in the emission of greater heat and necessitates a significant amount of power. In addition to the regulation of temperature, power amplifiers that have a greater gain and power density will be necessary. It is seen in Figure 1.7 that the design of 5G hardware and materials has requirements that are equivalent. Both potential and problems are presented by 5G and AASs in comparison to preceding generations of wireless technology. There are a number of possible opportunities and dangers associated with technologies that pertain to semiconductors, degrees of integration and functionality, packaging, size, and cost, as well as other front-end devices for radio frequency baseband transmitters.

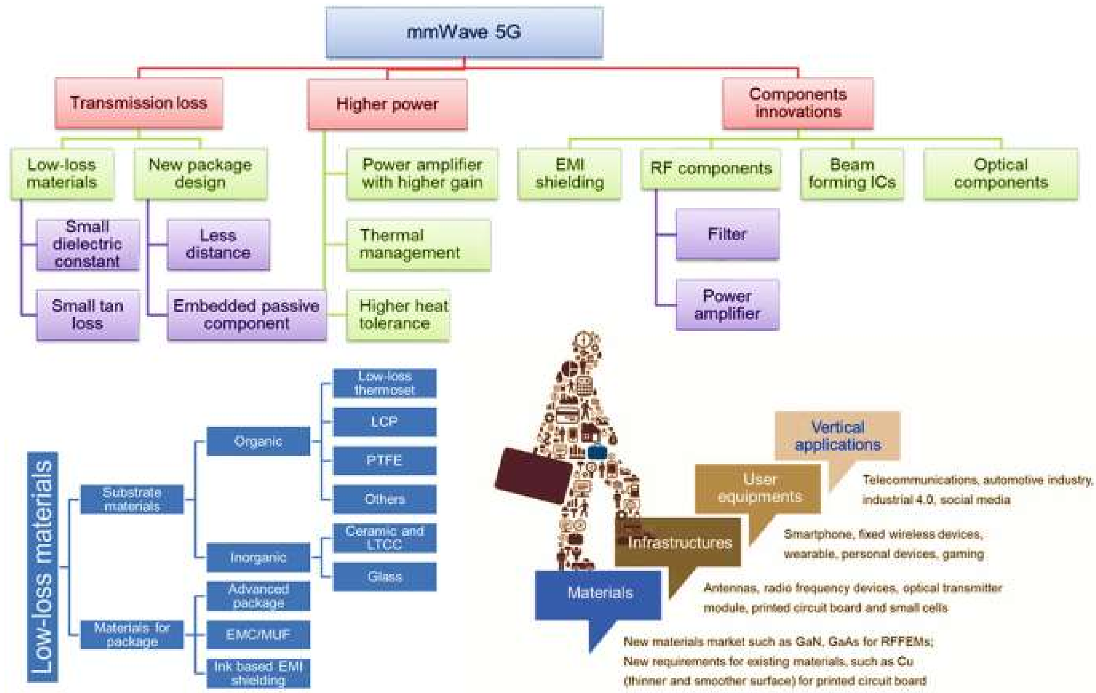


Fig. 1.7 Common requirements for 5G materials and hardware design

Source: Modified from Jiang et al. 2020, Credit: ID Tech EX

1.3.1 Evolution of the Cellular Base Station and its Construction Materials

Figure 1.8 shows the evolution of BTSs, which may be divided into three separate periods. First, a ground-based transceiver station, as shown in Figure 1.8a, is a common feature of older BTS installations. After then, the antenna components are driven by the radio frequency signal, which travels from the ground level all the way up to the masthead via high-level power amplification. A large number of currently operational infrastructure systems rely on this concept. Figure 1.8b shows, however, that a lot has changed, and this representation is more in keeping with many of the systems that have been built recently and are still in use. The base station's entrance is still down on the ground, and that's also where you'll find the channel cards. However, the changeover to radio frequency (RF) has been completed at the very peak of the tower in the farthest radio heads, with antennas placed directly above them. In terms of the system's power efficiency, this technique provides substantial benefits over previous systems while also being far more versatile. And since it helps the architecture become more modular, network providers have greater leeway to adjust to their own needs and requirements.

Figure 1.8c shows the changeover to the real 5G networking standard, with AASs mounted atop the distant radio heads. In this scenario, an AAS encompasses the whole spectrum of radio frequency (RF) electronics, which includes RF power amplification, Rx functionality, filtering, and other related operations. System designs like mMIMO offer exciting new opportunities that will allow for much higher data rates. Figure 1.8c also shows that operators could serve a lot more customers if they installed small cells at frequencies between sub-6 GHz and millimeter wave, which would improve density. Fixed wireless access is an additional potential use case. With the arrival of 5G, this is shifting into a phased array to provide point-to-point solutions. As an additional effect, the frequency range is widened. From 2.7 GHz to around 6 GHz, sub-6 GHz is gradually increasing. Also, millimeter-wave frequencies of 28 GHz and up to 40 GHz have just arrived.

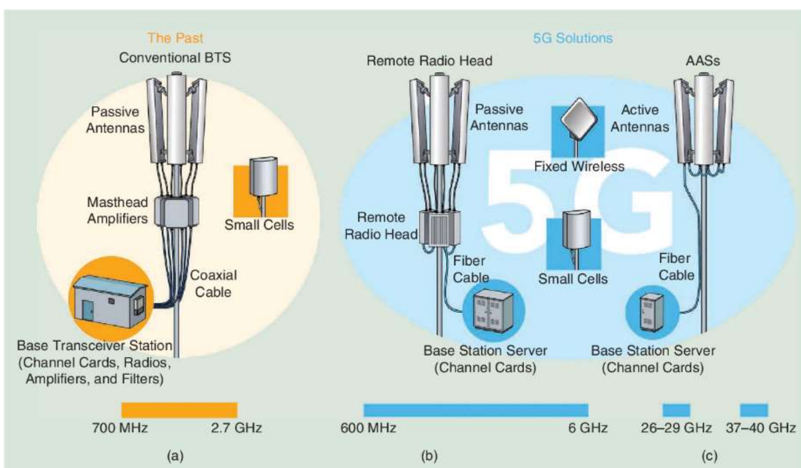


Fig. 1.8 The evolution of the cellular base station

Source: Adapted from Westberg et al. 2019 with permission from IEEE): (a) Conventional BTS, (b) remote radio head, and (c) AASs

Due to the fact that they enable several wireless devices to connect to a single hub, base stations are an important component in the discipline of wireless technology. An antenna and analog-to-digital converters (ADCs) are the components that make up this wireless receiver and transceiver with a short range configuration. These components are responsible for converting RF impulses into analog signals once they have been digitized using digital technology. Utilizing the millimeter wave spectrum will need the installation of large beamforming multiple-input, multiple-output (MIMO) antennas on the base station of the fifth-generation wireless network.

Through the concentration and direction of many beams emanating from this array of antennas, it is possible to identify and target a mobile phone or any other ground-based target. As protective housings, antenna Radames shield the antenna from the elements, including wind, rain, and ice, ensuring its continued operation. One of the most important performance goals for radome materials is to have a low permittivity and a low loss tangent at several gigahertz frequencies. Other important performance goals include reducing costs and improving processing efficiency. A number of materials, including polypropylene (PP), polycarbonate (PC), polyvinyl chloride (PVC), and sheet molding compound (SMC), are often used in the production of radome shells. As a result of their lower dielectric constant (Dk) and the simplicity with which they can be manufactured, they seem to be more suitable for use as radome shells. In addition, antenna Radames are required to have consistently high values for the dielectric constant (Dk) and the dissipation factor (Df), as well as low mechanical strength at low temperatures and high levels of ultraviolet (UV) resistance.

Drivers to 5G Hardware System Integration:

This is the direct reason of the complexity of 5G systems, which is caused by the addition of capabilities to existing gear. As a result of this, the integration, interoperability, and protection of all of the devices and components that are included inside a module need to be more stringent. The situation is comparable to the way that transmissions with higher frequencies have more losses in free space as a result of increasing attenuation. These approaches, which include beamforming and multiple-input multiple-output (MIMO), make it possible for beams to link with target nodes in a manner that is very concentrated and narrow. This may help to offset the effects of attenuation. The use of higher-order antenna arrays that are driven by several active components is what is required to achieve this.

Digital, analogue, and hybrid beamforming methods are the primary classifications that are used to classify MIMO beamforming approaches. When it comes to analogue beamforming, modulated signals are responsible for controlling the amplitude and phase fluctuation at the broadcast end. A combination of the signals that have been received from all of the antennas is performed before the analog-to-digital (ADC) conversion takes place. At the broadcast end, digital beamforming is responsible for regulating the amplitude and phase fluctuations that occur as demodulated signals. The processing of the incoming analogue signals, on the other hand, takes place after the ADC has been completed.

In hybrid beamforming technology, the combination of analogue and digital beamforming types enables a great level of customization in terms of channel layout, low cost, and power consumption, among other advantages. In addition, hybrid beamforming technology offers more flexibility. As a result of this quality, the base station is capable of supporting several users at the same time.

When it comes to mobile systems, analogue beamforming is the way to go since it only needs one RF chain to manage the antenna components. On the other hand, passive components are able to benefit from beam modification. However, digital beamforming may be able to give complete beam control for future ultrawide-band communication. This would be possible with several beams operating simultaneously and a large number of apertures. In addition, software-defined radio capabilities may be included in order to provide intelligent and dynamic band selection as well as spectrum hopping. In comparison to digital beamforming, classical beamforming is more power-efficient and needs more advanced technical implementation. For the purpose of encoding, multiplexing, and digitizing the antenna outputs, this cutting-edge digital beamforming approach makes use of a single analog-to-digital converter (ADC).

After that, it executes demultiplexing and beamforming in the digital domain. By using a recently developed onsite coding method that incorporates frequency and code multiplexing, it is possible to lower the average power consumption of the digitizer by a factor of ten to thirty-two. In-band duplex operation may be implemented by systems via the use of time division duplexing (TDD) or frequency division duplexing (FDD) in order to make the most of the limited spectrum allotment. The interference that occurs between the signals that are being sent and those that are being received may be reduced if the signals are differentiated in terms of time or frequency. To finish any of them, however, you will need twice as much time or effort as you would normally. For the purpose of overcoming this constraint and making the most of the 5G spectrum, technologies such as simultaneous transmit and receive (STAR) systems were developed. In order to achieve full duplex functioning inside the band, it is necessary to eliminate interference between the strong transmitter and receiver. It may be possible to accomplish this objective by the use of a mix of filtering in the analogue processing to reduce Tx harmonics and digital signal processing to remove multipath domains, in addition to cross-polarization array components.

In the process of attempting to achieve the desired level of signal strength, the selection and incorporation of transceiver (TRx) technologies are determined by the number of antenna components as well as the angle of the antenna gain. It is possible to combine the RFFE, control, and calibration circuits onto the transceiver dies of the CMOS, SiGe BiCMOS, or GaAs technology. These circuits are connected to the antennas, passive components, and power circuitries inside the package by means of package vias (TPVs), micro bumps, and routing layers that are included within the package. For component counts that are lower, GaN is the material of choice; for component counts that are higher, SiGe BiCMOS is the material of choice; and for bulk CMOS, the situation is completely different. Highly integrated wireless transceivers have been able to deliver single-chip CMOS systems as a result of technological breakthroughs ranging from deep submicron to 28 nanometers. CMOS and SiGe power amplifiers come with power amplifiers that are tuned for 25 dBm power and 30% efficiency.

On the other hand, GaAs power amplifiers reach 30-35 dBm, while GaN power amplifiers go over 45 dBm. It is possible for a 64-element array to reach 240 W with up to 5 W of power from each PA. In order to lessen the magnitude of the error vector, it is necessary for every PA to function in the linear domain. This presents a greater challenge for antennas with 64 arrays. It is necessary to have integrated thermal management in order to deal with the higher power dissipation that is brought about by the combination of analogue and digital processing. This is because the predicted efficiency of 8-dB back-off PA modules is forty percent.

1.3.3 Materials and Electronic Components for 5G Packaging Technology

When it comes to the electrical packaging of 5G systems, a unified module is required. This module is responsible for housing all of the system's components, including passive ones, as well as wireless, analogue, and digital operations. These systems are great examples of the rising trend toward heterogeneous integration, which is a process that is becoming more prevalent. Because of the reasons that are detailed below (Watanabe et al., 2021), this is becoming an increasingly significant component of 5G communication:

(a) Antennas to be integrated with transceiver ICS and related passive and RF power divider networks;
(b) Antennas to be integrated with sub6 GHz (FR1) in the near future as packaging technologies advance;
(c) Filters, diplexers, broadband power amplifiers, and switches to be integrated with new mm-wave bands (FR2); and
(d) Additional emphasis on miniaturization and component integration through add-on modules to the existing RFFE. It is essential to arrange the front-end module and the transceiver in close proximity to one another in order to get further reductions in size and losses. For the purpose of accomplishing this objective, antennas are included into the radio frequency (RF) module, and concurrently, a heat dissipation solution is modelled simultaneously.

Keeping the temperatures of the active components at a level that is under control is the goal. When power amplifiers are integrated with antenna arrays, there are a number of difficulties that need to be resolved, including those pertaining to size, cost, and performance. The development of novel low-loss materials to reduce conductive losses, the modeling of circuit, device, packaging, and thermal solutions, and the manufacture of multilayers with fine-line features and precise layer-to-layer registration are all items that are required in order to handle these challenges. A further emphasis on the significance of separating the different circuit components is brought about by the development of innovative technologies for integrating three-dimensional packages. With the deployment of such high-power amplifiers and a massive antenna array in millions of base stations, the issue of cost in high-volume manufacture is brought into sharp focus.

Packaging Requirements for 5G Systems:

Baseband modules and antenna-integrated transceiver modules are the two primary forms of system-level packaging that are used in millimeter-wave electronic technologies. In these kinds of packaging, the connections between the components, which may include printed circuit boards (PCBs), antennas, and integrated circuits (ICs), must adhere to a great deal of stringent standards. Particularly in the field of analog electronics, impedance regulation is a need that is seen as being of great significance. Interconnecting antennas with transceiver integrated circuits (ICs) in mm-wave antenna-in-package (AiP) systems is the best way to meet the goal of achieving minimum insertion loss and acceptable return loss throughout the frequency range of interest. Form factor that is connected to the connection mechanisms is another essential need that must be met. In accordance with the information shown in Figure 1.9, wire bonding and fip-chip connections are two of the most often used methods for connecting components. Alternately, the new method might be described as fan-out packaging or incorporation of integrated circuits. The interconnection properties of fip-chip and fan-out interconnections, such as fne pitch and low electrical parasitic, are gaining importance in RF/mm-wave packages, such as baseband modules and antenna-integrated modules. This is despite the fact that these types of connections were initially designed for high-performance computing or mobile processor applications.

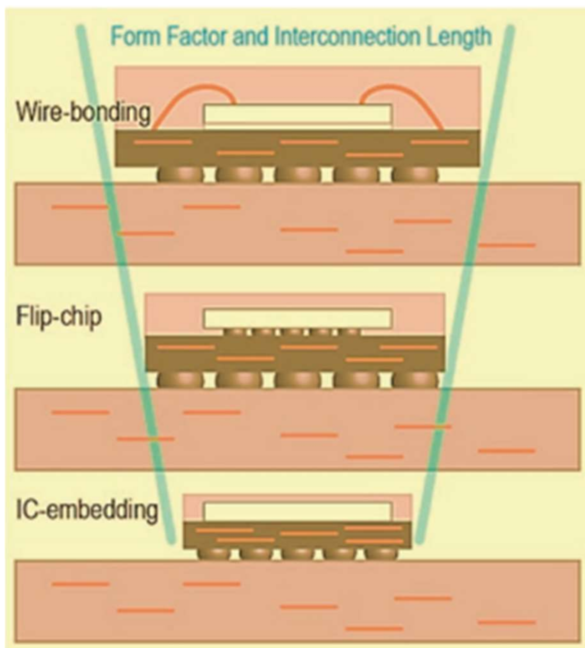


Fig. 1.9 Trends in the miniaturization of RF/mm-wave technology via connector assembly approaches. The most conventional low-I/O connection method is wire-bonding; flip-chip connectivity is possible down to 25 µm; and fan-out or IC-embedding interconnection is also an option.

Source: Adapted from Watanabe et al. 2021 with permission from IEEE)

Wire-bonding connections is a significant challenge in system-on-package (SoP) design. This interconnection causes a notable decrease in the performance of RF/mm-wave system chains owing to the influence that bond wires have on signal loss and impedance discontinuity. This is the case regardless of whether or not the packaging industry has developed into a more cost-effective sector overall. In comparison to wire bonding, the fip-chip approach has a superior performance when it comes to connecting components. The reason for this is because the length of the bond wires is longer than the height of the bump when they are joined after they have been connected.

Additionally, the fip-chip approach offers more than 800 I/O points, which makes it an excellent choice for compact form factors. For the first iteration of the fip-chip approach, solder bumps with a diameter ranging from 75 to 200 μm were used. It is worth noting that solder caps on copper pillar connections have the capability to attain sizes that are smaller than 40 micrometers. Not only did the copper-pillar approach make it possible to achieve high-density input/output (I/O), but it also helped to decrease conductivity loss. There has been a significant amount of interest in the development of fan-out or IC-embedding packaging. As mm-wave packages, fan-out wafer-level packaging (FOWLP) and embedded wafer-level ball grid arrays (eWLB) are attracting a significant amount of attention from a variety of individuals.

As a result of the fact that this embedding approach removes the need for wire bonding, which causes severe high-frequency loss and parasitic, as well as increasing the footprint for high-pin count dies, it is no longer necessary to utilize such dies. The procedure of compression molding is used in order to position integrated circuits for transceivers inside a reshaped wafer that has been formed. In order to disperse the baseband signals, many redistribution layers, also known as RDLs, are constructed. Additionally, through-mold vias are used in order to build vertical linkages. One of the possibilities is the development of panel-level embedding techniques that are based on glass or laminate. Another possibility is the technology that allows integrated circuits to be embedded at the waver technology level.

Both the core modem and the antenna-integrated package schematics are shown in Figure 1.10. The vast majority of these instances are examples of modules that incorporate a number of different strategies from each of the three different sorts of connections. One possible method of connecting the memory to the logic or modem die that is located on top of the printed circuit board (PCB) is to use a bond wire. Fip-chip interconnects, on the other hand, are essential for the high-pin modem die in order to provide a clean signal and reduce the amount of time required for signal delay. Recently, the fip-chip technology has emerged as the industry standard for the assembly of mm-wave phased antenna arrays that include integrated circuits that are coupled to one another. There are two factors that are responsible for this: the improved supply chain and the decreased production costs. Therefore, the conductive materials that are used are determined by the sensitivity of the constructed dies to the conductive loss that is brought about by interconnections as well as the number of pins. The C4 (fip chip) bumps are the alternative choice, whereas the copper pillars are the first option.

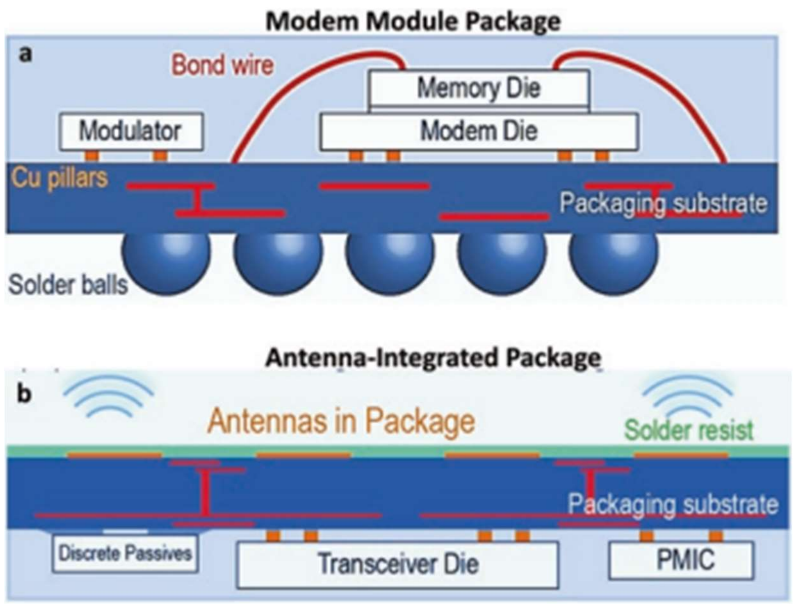


Fig. 1.10 Examples of system-integrated mm-wave packages

Source: Adapted from Watanabe et al. 2021 with permission from IEEE): (a) modem SoP; (b) antenna-integrated package

Dielectric Materials for 5G Module Packages:

As can be seen in Figure 1.11, 5G mm-wave modules need a stack of logic-memory components, surface-mounted passive components, transceiver integrated circuits, power management integrated circuits, and antenna arrays. It is common for substrates that are used for mm-wave packing to have a thickness that falls anywhere between 0.15 millimeters and 1.2 millimeters typically. To a large extent, the variance may be attributed to laws regarding antenna packing. The use of substrates that are thinner than

300 micrometers is feasible for the production of packages that do not include antennas. However, modules that have antennas integrated in are required to have a greater thickness. This is because a thicker substrate enables a greater bandwidth when antenna arrays are used, which is the reason why this is significant. The increase in the amount of space that exists between the ground plane and the antenna patches is the cause for this. Module thicknesses may range anywhere from half a millimeter to two millimeters, and the range of module thicknesses is determined by the height of the mold. However, one of these factors is the thickness of the packing substrate. In addition to the thickness requirement for the antenna, the entire package area and thickness are determined by the components and interconnects of the system (Figure 1.11).

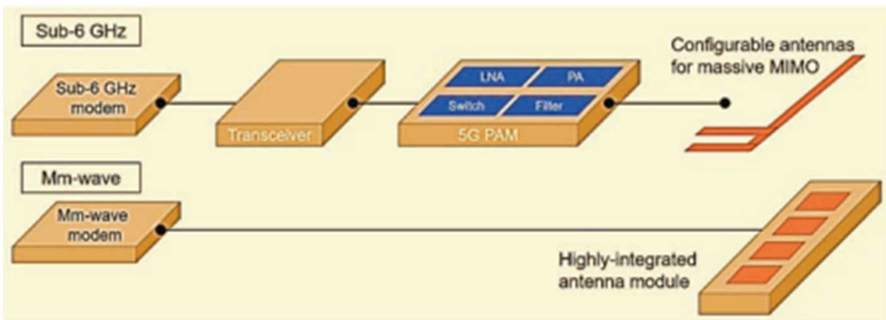


Fig. 1.11 An example of tight integration in RFFE architectures for 5G electronics

Source: Adapted from Watanabe et al. 2021 with permission from IEEE)

The selection of dielectric materials is of the utmost importance in order to achieve the requisite electrical performance of the integrated circuits (ICs) and antenna arrays that are used in the packaging of 5G module components. An improved link budget, low signal dissipation, high signal or power density, the desired frequency response of passive components, a small element footprint, a lower module thickness, high antenna efficiency and EIRP, beam width, angular coverage, energy consumption, and miniaturization of antennas are some of the benefits that can result from the strategic utilization of dielectric materials that have been thoughtfully designed. The production costs of substrate materials should be maintained to a minimum, and the moisture absorption values, dissipation factor (Df), and dielectric constant (Dk) should be as low as possible. In order to construct their circuit board substrates, 4G networks make use of polytetrafluoroethylene (PTFE), while 5G networks make use of alkali-free glass fiber base cloth.

At least for the time being, the majority of antenna integrated package platforms are constructed on organic substrates that are composed of many layers. The manufacturing procedures that are utilized to build multilayered organic substrates for millimeter-wave modules are conceptually similar to those that are used to manufacture printed circuit boards (PCB). Using this technology, it is possible to manufacture in large quantities while still maintaining compatibility with the supply chain that is already in place, which also makes it cost-effective.

Table 1.2 Low-loss dielectric materials currently used for 5G substrate technology

Materials		**Dk**	**Df**	**Reported frequency (GHz)**	**Major suppliers**
Organic	Bismaleimide Triazine (BT)	3.4	0.004-0.005	10	Mitsubishi
	Polyphenylethers (PPE)	3.25-3.4	0.002-0.005	1-50	Panasonic. Risho Kogyo
	Liquid-crystal Polymer (LCP)	2.9	0.0025	10	Rogers. Murata
	Polytetralluoro-ethylene (PTFE)	2.2	0.0009	10	Rogers. DuPont
Inorganic	Low-temperature cofired ceramic (LTCC)	6	0.0018	60	Hitachi Metals. Kyocera. TOK
	Borosilicate glass	5.4	0.005-0.009	10-60	AGC. Coming. Schott. 3DGS. NSG
	Fused silica	3.8	0.0003-0.0004	10-60	AGC. Coming. Schott, 3DGS. NSG

Should include devices that are used on a daily basis. In the millimeter wave frequency ranges, organic materials are developed to have a tan δ that is lower than the standard

FR4, which has a tan δ that is larger than 0.02. As stated by Watanabe et al. 2021, the prepregs and copper-clad laminates (CCL) that are employed in the process of producing multilayered organic substrates are composed of four unique kinds of polymers. The first substance is called bismaleimide triazine (BT), followed by polyphenylene ether (PPE), liquid-crystal polymer (LCP), and finally polytetrafluoroethylene (PTFE). A high glass-transition temperature (Tg), a low water absorption rate, and a low dielectric constant (Dk and Df) are characteristics of glass-cloth PPE substrates, in contrast to glass-cloth epoxy resin (such as FR4). Materials such as Risho Kogyo's CS-3376C and Panasonic's MEGTRON6 (core and prepreg) are examples of substrates that are based on polypropylene. It is possible to create multilayered organic substrates by laminating the BT-based core with build-up dielectrics. This process is quite similar to the formation of PPE resin. It is possible to achieve low CTE, little shrinkage, and high peel strength with copper by using laminates that are based on BT.

There is a wide variety of Dk (3.7-8) and Df (0.0003 for fused silica to 0.006 for alkaline-free borosilicate) that may be found in the glass substrates. Furthermore, it is worth noting that these substrates may be tailored according to the components that are packed, possess a smooth surface, have great dimensional stability (2 μm for 20-mm substrates), and have the capability to manufacture fine pitch through vias. In addition to this, they are not affected by temperature or humidity. Despite this, difficulties with glass packaging continue to exist as a result of unpreparedness in the supply chain and immaturity in the process, which ultimately results in higher pricing. In addition, working with glass has a number of complications owing to the fact that it may be either delicate or robust, and it is also tough to manipulate throughout the process.

When the substrate materials are taken into consideration, the manufacturing of RDL and microvias becomes increasingly significant for the construction of high-density interconnects with a thickness of 5 micrometers.

According to the data shown in Table 1.3, the industry is now engaged in the process of developing materials for millimeter-wave (mm-wave) packaging that possess low loss or low tan δ. Epoxy, a material that is often utilized for low-loss dry films, is generally accompanied with the use of silica or ceramic fillers in order to reduce the dissipation factor (tan δ). There are a variety of technical issues that are brought about by the fillers, such as a limited process window, poor adherence to substrate materials, and dependability in severe settings. The use of fillers results in an increase in Young's modulus and a reduction in CTE. A further technique for the production of dielectrics involves the use of liquid-based dielectrics, such as polyimides. When it comes to wafer operations or build-up layers on one side of the substrate, the use of liquid-based dielectric materials is considered to be more convenient in comparison to dry films. As a result of this, they perform better than dry films. The development of nonfilter photosensitive or photo imageable low-loss liquid-based dielectric materials that are especially designed for 5G mm-wave applications is under underway. The objective is to create fine features that are smaller than 5 µm. The following table provides a list of some of the most important producers of liquid-based dielectric materials that have a low loss capability.

However, the manufacture of these materials in large quantities is still quite some time away. Because it is difficult to produce films with thicknesses more than 40 micrometers, the fact that these materials are spin-coated or slit-coated adds another layer of complexity to the situation.

Table 1.3 Low-loss build-up dry flms for high-density signal routings used in high-frequency circuits

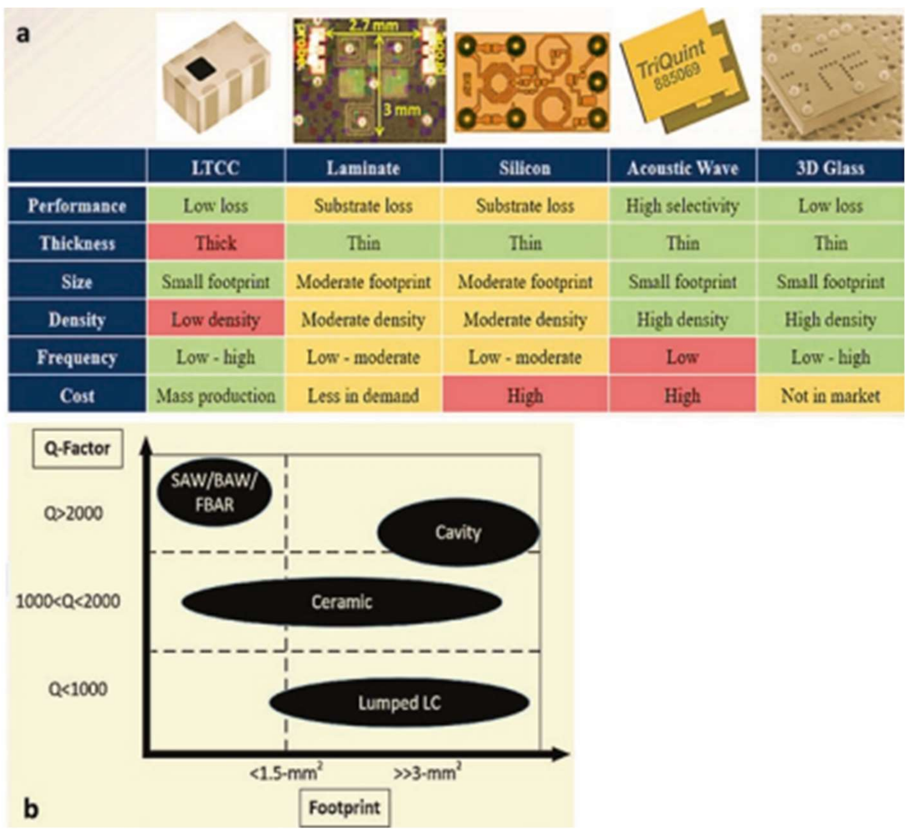
Material supplier	Dk	Df	Applied frequency	CTE (ppm/K)	Tg (°C)
Ajinomoto	3.3	0.0044	5.8 GHz	20	153
DOW	2S7	0.0032	1 MHz	63	250
Hitachi Chemical	33	0.0034	5 GHz	17	233
Sekisui	33	0.0037	5.8 GHz	27	183
Taiyo Ink	33	0.0025-0.003	5-60 GHz	20	160

Table 1.4 Low-loss photosensitive (Photo-imageable) dielectrics for build-up dry films for high-density signal routings used in high-frequency circuits

Material supplier	Dk	Df	Applied frequency	Minimum line width/ space (L/S)	Elongation (%)
DOW	2.65	0.0008	<20 GHz	18 pm	8
Hitachi Chemical	2.4	0.0018	10 GHz	-	-
Sekisui	2.6	0.0048	<40 GHz	8 pm	>50
Taiyo Ink	2.9	0.0030	1 GHz	30 pm	-

Discrete Lumped Circuits for sub6 GHz 5G Bands:

LTCC, which stands for low-temperature ceramic fused ceramic, turned out to be the first passive component to be developed. The use of these components as surface-mount technology (SMT) components for capacitors, inductors, and resistors was extremely widespread. A comparison of five different substrate technologies for passive components is shown in Figure 1.12a. These technologies are divided down according to characteristics such as size, thickness, and performance metrics.



	LTCC	Laminate	Silicon	Acoustic Wave	3D Glass
Performance	Low loss	Substrate loss	Substrate loss	High selectivity	Low loss
Thickness	Thick	Thin	Thin	Thin	Thin
Size	Small footprint	Moderate footprint	Moderate footprint	Small footprint	Small footprint
Density	Low density	Moderate density	Moderate density	High density	High density
Frequency	Low - high	Low - moderate	Low - moderate	Low	Low - high
Cost	Mass production	Less in demand	High	High	Not in market

Fig. 1.12 (a) Evolution of passive component technologies; and (b) comparison of various filter technologies at 1 GHz in terms of performance (Q-factor) and footprint

Source: Adapted from Watanabe et al. 2021 with permission from IEEE)

In addition, the equation that illustrates the connection between the footprint and the filter performance (Q-factor) may be found in Figure 1.12b. Acoustic wave technologies have outperformed the classic lumped and distributed LC networks in the sub-6 GHz range. This has made it possible to realize high-performance filters, resonators, oscillators, and delay lines; among other things. In addition, 4G and LTE networks have heavily used these technologies in their operations.

Distributed Components for mm-Wave:

In addition to couplers, baluns, and filters, other examples of the planar passive components that are manufactured are power dividers and couplers. These components have made significant development in a number of areas, including the decrease of their size and the improvement of these performance metrics. Conductor-backed coplanar waveguides (CBCPWs), microstrip, and substrate-integrated waveguides (SIWs) are some of the transmission lines that are the most suitable for use. Long-term continuous compounding (LTCC), organic laminates, ultrathin laminated glass, and high-tech integrated fan-out wafer-level packaging (InFO WLP) are some of the technologies that are used in the manufacturing process of passive components.

Antenna Systems in Package:

High radiated power, beamforming, a large signal-to-noise ratio, scanning in elevation and azimuth directions over a vast range, and other capabilities of a similar kind are among the numerous capabilities that are enabled by the highly integrated radio access solutions that are a part of the 5G wireless communication system. Part of these remedies is the use of phased-array antenna technology and promising transceiver front-end technologies. while it comes to 4G and LTE capabilities, antenna-in-package (AiP) devices are more feasible than discrete antennas while operating at mm-wave frequencies. There is a direct correlation between the size and pitch of the antenna components and the wavelength, which is the explanation for this phenomenon. Patch and dipole antennas are used extensively in antenna systems that are designed to interact with packages. The proportionate increase in gain that happens with an increase in the number of antenna elements is equal to ten times the logarithm of the number of antenna elements.

According to Qualcomm QTM052, a gain of 20 dBi is claimed to be provided by an antenna array that is 2-by-2 in size. Each individual element contributes 5 dBi to the overall gain, which is comprised of 6 dBi for the sum, 6 dBi for the construction of the array, and 3 dBi for the single polarization. The configuration of mm-wave antenna arrays may be done in a number of different methods, as shown in Figure 1.13. The first possibility is to attach antennas directly to the printed circuit board (PCB), as seen in Figure 1.13a. Not only is it more cost-effective than the other option, but it also comes with a number of additional advantages that are worth mentioning.

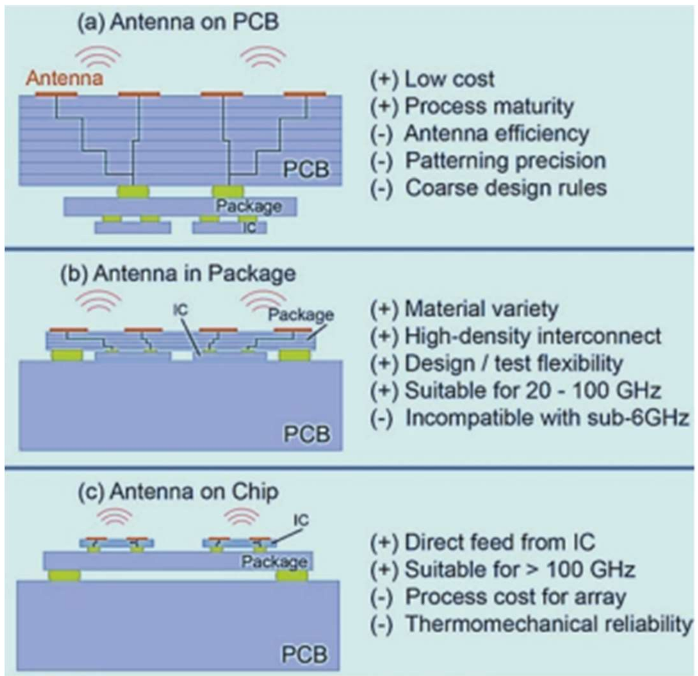


Fig. 1.13 Three approaches for mm-wave antenna implementation

Source: Adapted from Watanabe et al. 2021 with permission from IEEE): (a) Antennas on PCB; (b) antennas in package; and (c) antennas on chip or wafer

Chances to be taken advantage of as a consequence of the supply chain's maturation. The manufacturing process and tolerances are the two significant problems that are linked with antennas that are manufactured on printed circuit boards (PCBs). Due to the relatively coarse design constraints, designers are unable to plan fine structures in transmission-line widths, spacing, via diameters, pitch, or with perfect layer-to-layer alignment. This constraint prevents them from planning fine structures. It is due to the fact that the criteria are not very exact that this is the case. Based on the facts, the implementation of AiP, which is the optimal technique for 5G mm-wave applications, is shown in Figure 1.13b. On the other hand, Diagram 1.13c illustrates the implementation of the antenna in a direct placement onto the integrated circuit wafer.

Although there are numerous benefits associated with the antenna-on-chip technology, such as the ability to integrate directly with other front-end circuits and a reduction in feedline loss and parasitic, there are also a number of problems that occur with this method. Problems with design flexibility, yield and cost, process scalability for large arrays, and antenna efficiency are only a few examples of the challenges that are encountered.

1.3.4 Nanomaterials for Nanoantenna's in 5G

For wireless communications at the 5G level, antennas that are bigger, more adaptable, more gain-adjustable, and steerable are required. In the previous generation of wireless communications, for one thing, the use of the spectrum was more restricted than it is now. Conventional antennas are unable to accommodate the new high frequency due to constraints in both their production and installation processes. When it comes to shapes that are more compact, this is especially true. Metallic nanoparticles are a component that is often found in conducting inks that are used for antennas.

In the event that these particles come into contact with air, water, or oxygen, the antenna will oxidize and degrade at a rate that is quicker than that of bulk metals. A significant amount of surface area is possessed by these particles, which is the explanation behind this. Graphene and carbon nanotubes (CNTs) are two examples of nanomaterials that are connected to carbon and have the potential to be used in the production of antennas that have dimensions that are smaller and thinner. This is due to the fact that the diameter of the antenna has an inverse relationship with the bandwidth that is accessible. On top of that, these antennas have the capability of producing signals at very high frequencies. Around two orders of magnitude smaller than the on-chip antennas that are now in use, nanoantenna's that are compatible with 5G networks are currently being developed.

Communications in the terahertz band are an example of a promising wireless technology that has the potential to satisfy the requirements of 5G and beyond. The reason for this is that the data throughput of terabits per second cannot be maintained by the use of the complete consecutive bandwidth that is available for millimeter waves. In accordance with the information shown in Figure 1.14, metamaterials, metallic nanoparticles, graphene, and carbon nanotubes (CNTs) are all excellent options for millimeter wave and even terahertz band frequencies. Graphene, carbon nanotubes (CNTs), metamaterials, and metallic nanoparticles are all examples of materials that possess electromagnetic properties that can be tuned. Both the permittivity and the permeability of these materials are characteristics that are significantly impacted by the microstructure of these materials. Nanoantenna's have a number of properties, one of which is radiation effectiveness. Radiation effectiveness is a measurement of the efficiency with which the antenna converts electric impulses into electromagnetic ones.

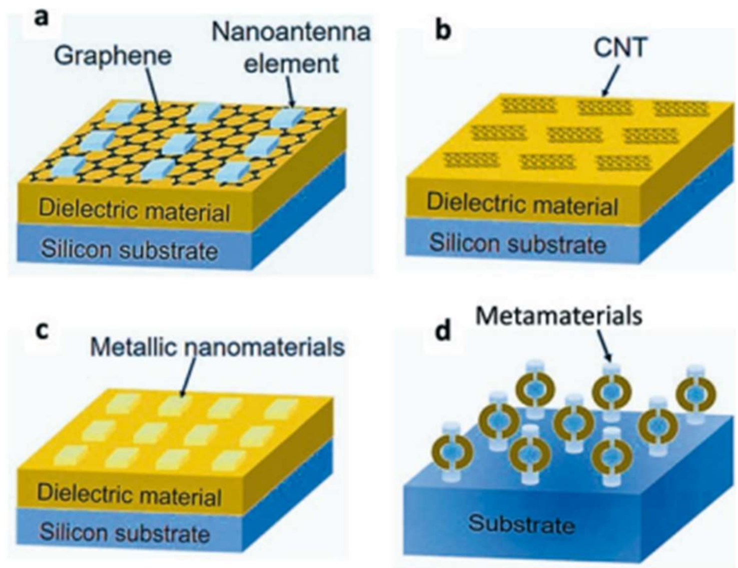


Fig. 1.14 (a) Nanoantenna's based on graphene, (b) nanoantenna's based on carbon nanotubes, (c) nanoantenna's based on metallic nanomaterials, and (d) nanoantenna's based on metamaterials are demonstrated in the following schematic design

Source: Modifed from Hao et al. 2020 with permission from Walter de Gruyter and Company)

At the same time as the radiation pattern illustrates the radiation that the antenna emits in relation to the directions that the electromagnetic waves are traveling, the radiation directivity illustrates the degree to which the radiation is focused in the direction that emits the most radiation. In order to provide you with a concept of size, the following characteristics are shown by nanoantenna's that are made from nanomaterials:

1. The radiation properties of graphene-based nanoantenna's can be controlled by managing their chemical potential, in the same way as the conductivity of

graphene can be modified by modifying the electrostatic bias voltage and chemical doping. When compared to graphene, nanoantenna's that are built on carbon nanotubes display a greater degree of plasmonic loss and a lower degree of tunability. Both of these characteristics are the result of the curvature effect. However, metallic nanoantenna's have radiation properties that are difficult to alter, and SPP waves that go through them have significant ohmic losses from the process.

2. In terms of electromagnetic radiation efficiency, nanoantenna's made from graphene perform better than their counterparts made from copper and carbon nanotubes. Additionally, the radiation efficiency of a single carbon nanotube (CNT) is much lower when compared to that of nanoantenna's that are made completely of copper. In light of the fact that copper is not a very excellent conductor, nanoantenna's built of copper do not absorb a significant amount of radiation. Using a single carbon nanotube (CNT) as a nanoantenna results in a low radiation efficiency. This is due to the fact that the nanoscale radius generates a significant reactance, which is caused by both classical and quantum processes.

   Graphene-based nanoantenna's have a higher directivity than carbon nanotube (CNT)-based nanoantenna's. On the other hand, in the low terahertz band frequency range, CNT dipole antennas have a higher directivity than copper nanoantenna's.
3. The ability to activate or deactivate the gain of graphene-based nanoantenna's may be achieved via the control of the electrical gate voltage. Modifying the gate voltage is the sole method that can be used to alter the architecture of the nanoantenna's array. Despite the fact that there are several varieties of nanoantenna arrays, this particular one does not make use of any switches or phase shifters in order to alter the radiation pattern. For 5G wireless communications, nanoantenna's built of graphene show a great deal more potential than those made of carbon nanotubes (CNTs) or metallic materials. This is due to the fact that graphene nanoantenna's have superior radiation properties, such as high radiation efficiency and directivity.
4. Antennas that are constructed using metamaterials are more compact and minuscule in comparison to conventional antennas. Furthermore, these antennas may be improved from their current state by raising their gain, creating multiband frequencies for antenna operation, and improving their bandwidth, among other qualities. In order to avoid the limitations of nanoantenna's, such

as their restricted bandwidth and efficacy, nanoantenna's that are based on metamaterials might solve the problem. It is possible that nanoantenna's constructed on metamaterials would increase the controllability and promise of nanoantenna radiation characteristics.

5. This is due to the fact that metamaterials cannot be generated in naturally existing materials, which means that they cannot possess unique electromagnetic properties. As a result of their versatility, metamaterials may be used in a variety of different ways to perform the many jobs that nanoantenna's are capable of having successfully completed. A good illustration of this would be an antenna array that has metamaterials that are placed in such a way that they encircle the nanoantenna elements, hence increasing the antenna gain. There is a possibility that metamaterials might be used as a superstrate in order to surround the radiation surface. Because of this, the nanoantenna's will be able to achieve a greater bandwidth than they now do.

Additionally, it is of the utmost importance to manage the energy consumption of nanoantenna's in accordance with the norms that are to be considered acceptable. Because of this, the operation of nanoantenna's may be significantly influenced by the heat that they generate via electromagnetic loss, which may result in an environment with higher temperatures. This is the reason why this particular situation exists. This imaginary fraction of the permittivity, which is not constant and is temperature sensitive, is what determines the dielectric loss component of electromagnetic loss. This component is a component of electromagnetic loss. As the temperature increases, the imaginary component of the permittivity increases as well, which results in a greater dielectric loss. Nanoantenna's make it possible to utilize nanocomposites that are capable of high-temperature absorption, which in turn minimizes the amount of energy that is used. By way of illustration, consider the combination of carbon nanotubes (CNTs) and graphene with ceramics such as silicon carbide (SiC) and silicon oxide (SiO_2). These ceramics have excellent thermal stability, high thermal conductivities at higher temperatures, low oxidation, and high strength. The surface of graphene and carbon nanotubes (CNTs) may be deposited with components such as metallic and ceramic nanoparticles.

1.4 CHALLENGES IN 5G AND BEYOND – 6G

The terahertz (THz) gap frequency range is where it is anticipated that 6G will operate, which will enable it to use channel bandwidths and frequencies that are much bigger than those provided by 5G. It is projected that this would result in data transfer rates that are far higher than the 10–20 Gbps that 5G is capable of achieving, with speeds ranging from 100 Gbps to 1 Tbps. The revolutionary flexibility of 5G will undoubtedly be expanded upon by the use of artificial intelligence (AI) in 6G, which will build on the foundation of 5G's adaptability. The rationale for this is that 6G will be able to make advantage of the creative freedom that 5G offers. Within the context of the introduction of 6G services, it is anticipated that further use cases will make significant development. Two use cases that will connect hundreds of billion devices and provide real microsecond latency are enormous URLLC and huge MTC. Both of these use cases are examples of huge URLLC.

The majority of the typical processes that are utilized for RF packages will not function when dealing with radios operating at sub-THz and THz frequencies. The fact is that this is the case for many of the strategies. According to Figure 1.15, the integration of heterogeneous packages is given a high level of priority in communications that operate at sub-THz or 6G frequencies. In order to achieve this objective in a methodical manner, the components of the system, which include waveguides, active devices, low-loss interconnects, and precision antenna arrays, are being improved. The technologies that are used at low frequencies will become obsolete when the frequency increases into the THz area because of the many problems that are associated with signal loss, size, and materials. There is a much greater frequency in the THz range, which is the cause for this. Current methods of attachment, such as wire bonding or solder bumps, are too arduous to be considered viable options.

Consequently, this indicates that multimode, radiation, and reflection excitation all have a detrimental effect on the performance of electrical systems. The skin has a thickness of around 120 nanometers when it is subjected to 300 gigahertz. Additionally, in order to minimize the amount of conductor loss that occurs during higher frequency packing, shorter interconnects and flat surfaces are important. It is of utmost importance to provide precise impedance matching and high-precision manufacture with a tolerance of less than 1 μm in order to guarantee the least amount of radiation loss possible. In order to effectively characterize potential materials at such high frequencies, it is necessary to take into consideration the fact that dielectric loss is more prevalent in sub-THz or THz bands. The phenomenon is brought about by the fact that some bands are more prone to experiencing dielectric loss than others. A frequent and essential component of THz applications is the antenna, which is also a common component.

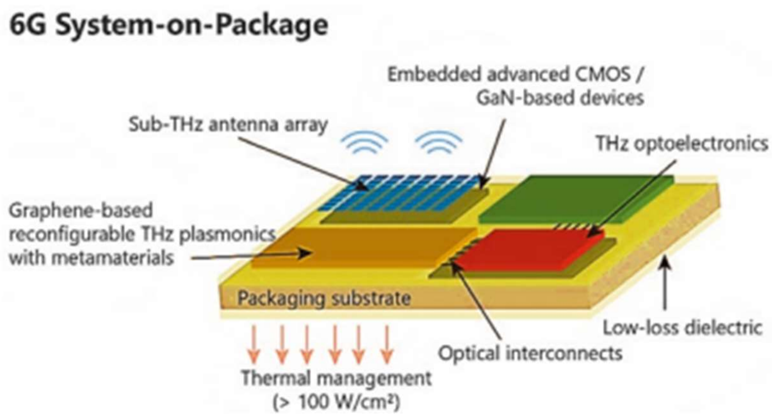


Fig. 1.15 Conceptual diagram of heterogeneously-integrated quasi-optical THz package for 6G communications

Source: Adapted from Watanabe et al. 2021 with permission from IEEE

One of the most important aspects of AiP, which is a technique that is feasible, is the incorporation of antennas or radiators into integrated circuit structures. There are many different components that may be included into the antenna. Some examples of these components are radiation sources and detectors constructed of graphene, lenses, THz antenna arrays (for example, 1024 elements in 1 mm2 at 1 THz), and intelligent meta surfaces that have space-time-frequency coding and low-loss interconnects. Figure 1.15 is a schematic representation of a device that is equipped with antennas and has a heterogeneously integrated quixotical THz 6G package embedded inside it. This apparatus may be set up to emit sub-THz or THz waves into the environment by using either the on-chip antenna or a reconfigurable AiP. Both of these configurations are available. When dealing with systems that operate at high frequencies, it is very necessary to address technological concerns such as the control of heat and the shielding of electromagnetic interference.

CHAPTER 2

EVOLUTION OF 5G SEMICONDUCTOR TECHNOLOGIES

2.1 INTRODUCTION

Currently available on the market are a multitude of semiconductor technologies that are capable of functioning at frequencies of up to 90 GHz. Taking into account the many applications that may be found for millimeter wave transmission systems, each of these choices comes with its own individual set of advantages and disadvantages. Through the use of primary devices that utilized single-function GaAs pHEMT MMIC technology, it was possible to demonstrate high-performance communication lines that were capable of working at frequencies that reached 86 GHz. In succeeding generations, multifunctional GaAs, GaN, and other compound semiconductor chipsets have also evolved. These chipsets give equivalent degrees of baseband to RF capabilities. Additionally, highly integrated silicon-based chips (SIGe, SIGEBICMOS, CMOS, and finally FD-SOI) have also arisen.

In order to fulfill the anticipated requirements for 5G fronthaul and backhaul capacity in the future, it will be necessary to make use of an extremely large quantity of contiguous spectrum bandwidth. As a result of the existence of this spectrum at frequencies higher than 90 GHz, it is absolutely necessary to create semiconductor technologies that are able to function at frequencies higher than 90 GHz and up to 300 GHz.

When the quantity of components reaches several million per year, you may anticipate cheaper costs and improved degrees of integration if you apply CMOS technology. Specifically, this is due to the fact that technology that has been scaled up is capable of doing tasks at a considerably faster rate than previous technologies. The transition frequency (fT) of CMOS transistors is increasing closer and closer to 400 GHz, and some innovations, like as the WiGig 60 GHz chipsets that are now available on the market, have shown that CMOS can be used at mmWave frequencies. On the other hand, when it comes to phase noise and noise figure, point-to-point connections perform worse than SiGe or GaAs components across the same distances (for example, more than 100 meters). One of the most well-known advantages of fully finished SOI (FDSOI) technology is its ability to achieve high speed at low voltage. These technologies, along with the common CMOS planar bulk technology, operate together in tandem. In addition, the fact that it is entirely compatible with bulk CMOS technology designs ensures that the process of managing intellectual property will be successful. SiGe-BiCMOS semiconductors being mixed together

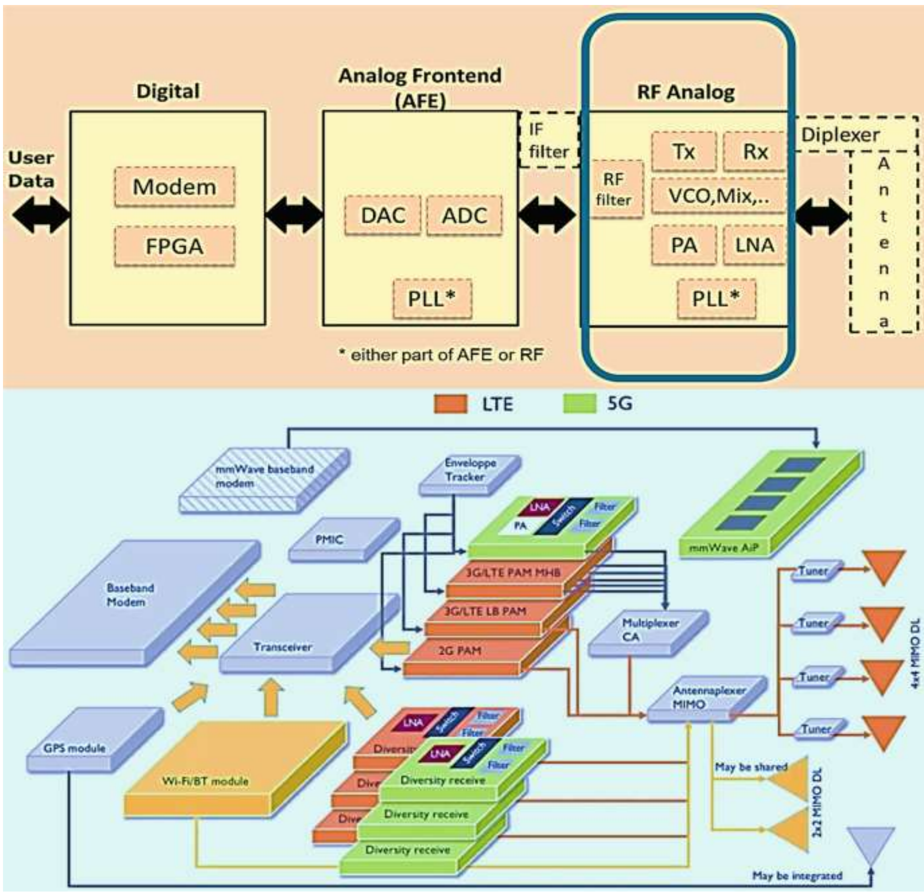


Fig. 2.1 Semiconductor system in 5G

Source: Modifed from Rüddenklau et al. 2018 with open access (Hindawi), and Yole 2019, credit: https://www.3dincites.com/2019/03/ the-5g-revolution-is-pushing-innovations-for-rf-front-end-sip/

The complementary metal-oxide semiconductors and high-performance bipolar transistors are both included onto a single chip. Bipolar transistors provide the high speed and gain that are required for high-frequency analogue sections, but complementary metal-oxide semiconductors (CMOS) are excellent for the construction of simple logic gates that need a minimum amount of power. The integration of digital, analogue, and radio frequency (RF) components onto a single chip is now feasible as a result of the development of cutting-edge SiGe BiCMOS technology, which was made possible by architects, designers, and process integrators. This technology makes it possible to integrate a number of different components into a single system. This simultaneously optimizes power usage while simultaneously reducing the number of external components by a significant amount. Using a technique known as BiCMOS process technology, it is already possible to achieve performance that is equivalent to that of more costly technologies such as gallium arsenide (GaAs).

BiCMOS, which makes use of silicon germanium (SiGe) heterojunction bipolar transistors (HBT) at a certain technological node, makes it feasible to achieve both a higher cutoff frequency and a greater voltage capacity. In contrast to this, CMOS technology is used in large quantities. The usage of a much smaller number of process nodes is required for bulk CMOS devices in order to achieve the same frequencies. This requires design concessions, which nearly always result in materials that are less expensive and a decrease in the overall power performance of the product.

The QFN, eWLB, CSP, and Flip Chip packaging technologies are some of the technologies that are used in the production of RF analogue components. You are able to access these processes at this same now. Modules or bare dies, on the other hand, constitute the majority of the mmWave radio frequency components that are now available for purchase. Examples of eWLB technology being integrated with SiP illustrate the outstanding system-integration capabilities of eWLB technology in the context of assembly and packaging. This encompasses the use of TEV or the RDL in order to include passive components such as transformers, inductors, and resistors into the circuit, in addition to the incorporation of various chips. Add-on antennas are included in the package as a standard feature. In addition, die embedding in laminate and TSV are examples of advancements that enable the integration of technology at the silicon wafer level with technology at the backend. When it comes to silicon interposers, for example, TSV technologies are often used in conjunction with RDL configurations. Due to the exorbitant cost, they are not used by a greater number of people.

Regarding the planning of the development of future chip integration for RF analogue components, there are a number of potential techniques that may be taken into consideration. Numerous elements will, for the most part, determine the result:

a) The system's authorized output power as well as its EIRP after taking into account the antenna

b) A certain amount of phase noise is necessary for the defined modulation scheme (BPSK, QPSK, QAM4,..., QAM256, etc.).

c) Figure of noise

d) The usage of power

e) Size in terms of the area of the printed circuit board and the associated cost.

At frequencies greater than 20 gigahertz, III-V compounds (GaAs, InP, and GaN) perform much better than silicon transistors in terms of low noise performance, linearity, and output power characteristics. The noise figure of the GaAs mmWave low-noise amplifier (LNA) is around 2.5 dB, which is much lower than the noise figure of the SiGe LNA, which operates at a level of 5 dB. With regard to the E-Band, for instance, GaAs have the potential to produce Psat values of more than 30 dBm, whilst SiGe-HBTs can only reach 19 dBm. In contrast, radio frequency application-specific integrated circuits (RF ASICs) constructed of silicon enable the integration of several application-specific capabilities onto a single silicon chip in a consistent and high-yield manner.

In addition, they provide an alternative to GaAs, which would make it far more difficult, if not impossible, to include the many calibration procedures that are required to take into consideration RF impairments. One of the most crucial aspects to take into consideration is the degree of integration. The chip becomes more specific when there is a high degree of integration in the original design, which may result in an increase in the amount of time required for development. One of the benefits is that it simplifies the assembly of modules and reduces the amount of time spent on production testing. A good compromise would be to use compound semiconductors for the front ends of high-end applications such as E-band high power and QAM256. These front ends include the low noise amplifier (LNA) of the receiver input and the power amplifier of the transmit output. Silicon semiconductors would be used for the lower frequency mixed signal functions and control/digital components. A further suitable middle ground is the use of silicon semiconductors for mixed signal processing at lower frequencies.

2.2 EFFECT OF CMOS TECHNOLOGY SCALING ON MILLIMETER WAVE OPERATIONS

Scaling theory, which was established by Mead and Dennard, makes it possible to implement a "photocopy reduction" technique for simplifying the process of lowering the size of features in CMOS technology. Even if the dimensions of MOS transistors are reduced, scaling theory ensures that the field strengths in these transistors remain unchanged over the course of several generations of the process.

The scaling of transistors is a significant factor that plays a significant role in the development of high-performance microprocessors and memory. As a result of the reduction of CMOS integrated circuit technology node scaling by thirty percent, the following benefits have been achieved:

(1) A forty-three percent increase in maximum clock frequency, which was made possible by a thirty percent reduction in gate delay;
(2) A twofold increase in device density;
(3) A thirty percent reduction in parasitic capacitance; and (4) a sixty-five percent reduction in energy and fifty percent reduction in active power per transition, respectively.

Additionally, CMOS scaling is having a positive impact on applications like as high-frequency radio frequency (RF) transmission and millimeter-wave (mmWave) communication. Therefore, this is due to the fact that scaling has a significant impact on the transition frequency (fT) of the MOSFET. A CMOS technology with a 28 nm process was used in order to achieve a frequency of 250 GHz. It is essential to have a significant value of fT when designing high frequency mmWave circuits. This is one of the most crucial requirements. Some of the advantages of scaling include a reduction in the noise figure, which is one of the benefits. The first component that is used in the construction of the RF receiver front end circuit of almost every communication system is a low noise amplifier, also known as an LNA. During the design process, one of the most critical objectives for a low noise amplifier (LNA) is to achieve a low noise figure. There are a number of obstacles involved with scaling, despite the fact that it is obvious that scaling CMOS technology has made it possible to create mmW circuits on silicon.

1. As an example, the frequency at which 5G must operate is 28 GHz, which is twenty times greater than the frequency at which 4G must operate, which is 2.6

GHz. On the other hand, the transition frequency of the MOSFET only increases by nearly five times, going from fifty gigahertz in the 180 nanometer process to two hundred and fifty gigahertz in the 28 nanometer process. As a result of the sluggish expansion of the MOSFET speed in comparison to the operating speed requirements, the circuit design faces a number of obstacles.

2. The inherent gain of the MOSFET and the supply voltage both decrease as the scaling process advances, which results in a reduction in the voltage gain of the amplifier circuit. The intrinsic gain may be calculated using the formula gm/gds, where gds represents the channel conductance coefficient. A process with a 45 nm resolution has an inherent gain of 5, but a process with a 180 nm resolution has an embedded gain of 25. A reduction in supply voltage may have a significant effect on a certain kind of circuit known as low-noise amplifiers (LNAs). These amplifiers make use of cascode design in order to reduce the influence of reverse isolation.
3. Because of the lower supply voltage and slower speed of transistors, inductive components are required for resonance in radio frequency circuits. This is because of the fact that the supply voltage is lower. The fact that passive inductive devices, such as lumped inductors or distributed transmission lines, may take up as much as half of the surface area of a silicon chip is a significant source of frustration for designers. The process of decreasing the size of the on-chip inductor has been a slow one since, according to electromagnetic theory, a large-size inductor is required in order to achieve the required inductance value and quality factor. When compared to its off-chip counterpart, an on-chip inductor may experience much higher levels of transmission, radiation, conduction, and substrate losses. A component of poor quality with a maximum value of around 20 is produced as a result of this.

The following is a list of some of the proposed ways for scaling on-chip inductors that preserve necessary performance criteria such as the inductance value and the quality factor.

- Fig. 2.2 illustrates a variety of inductor shapes, including square, octagon, and circular geometries, among others;
- The various approaches for optimizing layouts, such as shielding, stacking, and nesting, are illustrated in Figure 2.3a–c;
- Methods for improving the quality of the quality factor;

- Methods for minimizing lost property;
- As an alternative to the inductor and transmission line, the utilization of a planar integrated waveguide is being considered.
- The implementation of on-chip inductors is accomplished through the utilization of a variety of materials, including carbon nanotubes and multilayer graphene.

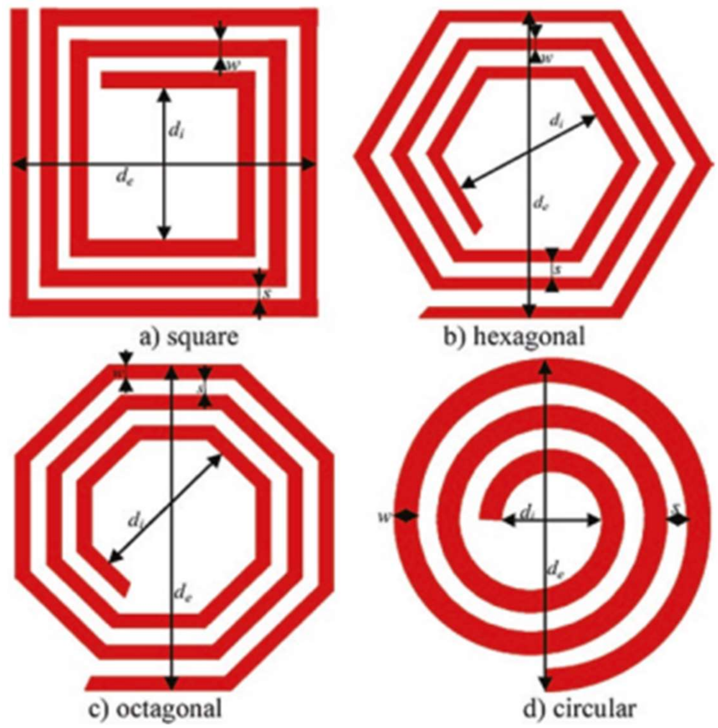


Fig. 2.2 Spiral inductors geometrical shapes used in CMOS

Source: Adapted from Pacurar et al. 2012 with permission from IEEE): (a) Square, (b) hexagonal, (c) octagon and (d) circular inductor

The manufacturing procedure for CMOS spiral inductors is depicted in figure 2.3d respectively. In order to reduce capacitive coupling and increase the quality factor (Q) beneath the inductor substructure, an oxide etching approach is applied. This technique serves as the foundation of the procedure. This method was used to build an octagonal spiral inductor with a strength of 2.2 nanohertz. By employing this technology, it was possible to increase the Q from 10 to 15 at a frequency of 3 gigahertz. When Q was equal to 19, the inductor's greatest resonance frequency, which had previously been 6.5

GHz, shifted to a different value. As a consequence of this, inductors that operate at a wide range of millimeter-wave frequencies can be investigated by examining oxide while simultaneously etching it.

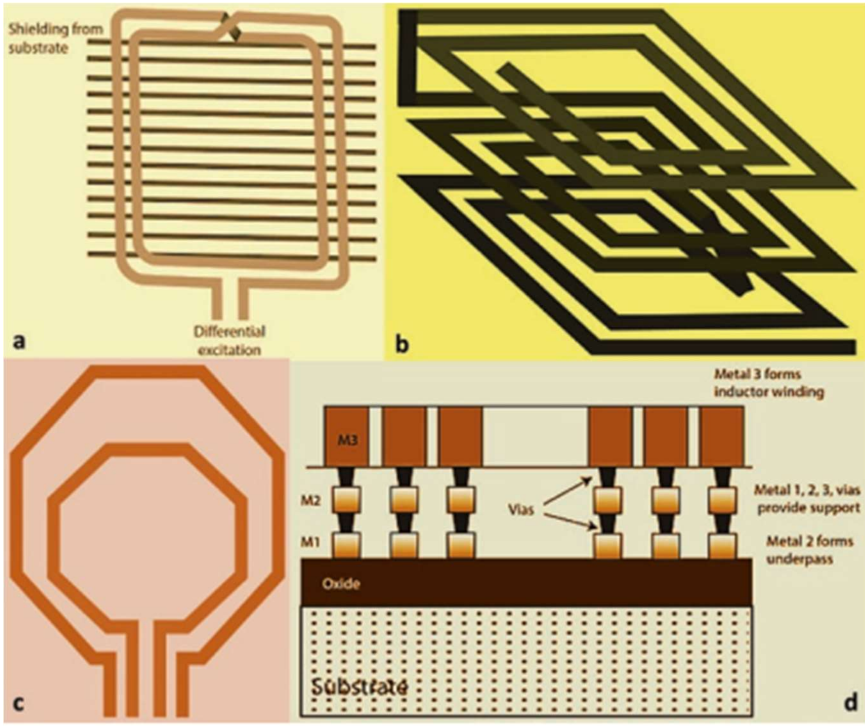


Fig. 2.3 Different layout optimization techniques for inductors

Source: Adapted from Juneja et al. 2021 with permission from Elsevier): (a) Shielding, (b) Stacking, (c) Nesting, (d) Cross sectional view of oxide etched inductor

2.3 DISTRIBUTED AND LUMPED DESIGN APPROACHES FOR FABRICATING PASSIVES

In the building of RF power amplifiers, lumped and distributed passive networks are frequently utilized as matching networks and filters. This is a popular application. A great number of circumstances call for a modification of the impedance at the frequency at which the product is being operated. The utilization of passive networks for the purpose of this conversion is a common procedure. A significant number of radio frequency (RF) applications require that the characteristic impedance of the interconnect transmission lines be fifty ohms. In addition, there are times when it is necessary to meet the requirement of matching the input and output ports of circuits to 50. Networks that transition between several impedance levels in order to match a certain impedance level are referred to as matching networks. Matching networks are a typical moniker for these passive networks because of this reason. Radio frequency (RF) applications make use of the capability of these networks to function as frequency-selective filters. This capability is helpful when dealing with signals that flow across a small wavelength range. On display in Figure 2.4 is the fact that passive components such as inductors and capacitors are responsible for the composition of both lumped-element matching networks and distributed matching networks. Distributed matching networks also make use of transmission lines in their operations.

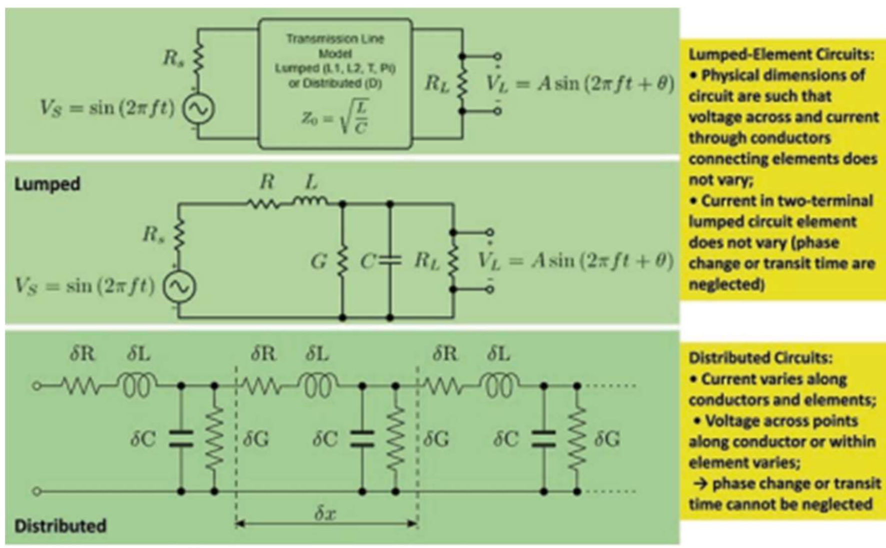


Fig. 2.4 The illustration of a model of a transmission line using either a lumped or dispersed arrangement

2.3.1 Distributed Approach

While operating at mmWave frequencies, where the signal wavelength is comparable to the size of the chip, resonance in the circuit can be achieved through the utilization of transmission lines in a distributed design approach. This provides the opportunity for the circuit to achieve resonance. Shunt conductance (G) is a measurement of silicon oxide losses that result in a capacitive quality factor (Qc), whereas a series resistor (R) in Figure 2.4 represents skin effects-induced conductive losses. Both of these types of losses are referred to as conductive losses. When mmWave transmission lines are constructed on silicon, the quasi-TEM mode of propagation is present in the electromagnetic field. As a result of this, it is possible to obtain precise values of extremely small inductances. In addition to the quasi-TEM mode, there are two other modes that are able to coexist in silicon-based planar transmission lines. Both the slow-wave mode and the skin depth mode are demonstrated in Figure 2.5. Both of these modes are shown in action. The operating frequency and the PSi resistivity are two pieces of information that can be used to determine whether or not propagating modes are present. due a silicon substrate with a low resistivity would exhibit conductor-like qualities, it is undesirable to have a skin depth mode of propagation. This is due of the phenomenon described above. On the other hand, a conventional planar transmission line can have an additional layer added to its construction in order to induce the slow-wave mode of transmission.

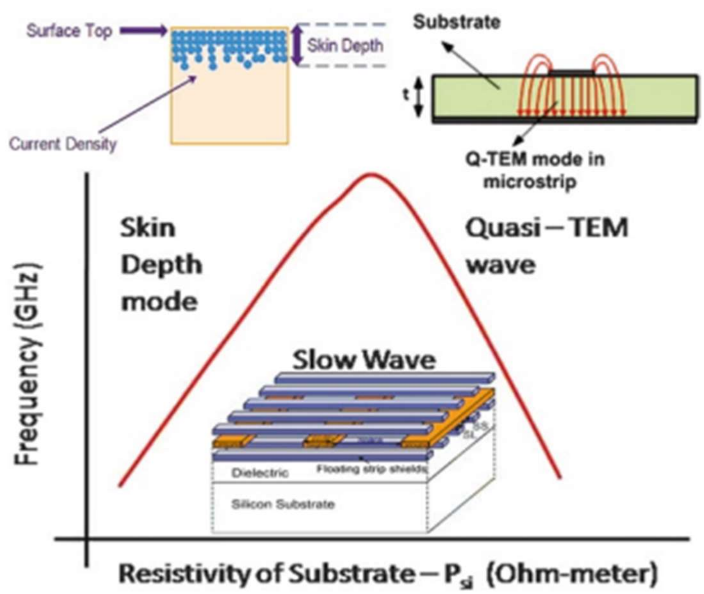


Fig. 2.5 Propagating modes of planer transmission line in Silicon

Source: Modifed from Juneja et al. 2021 with permission from Elsevier

Programmable CMOS allows for the creation of two separate types of transmission lines. In accordance with what is depicted in Figure 2.6a, the first one is the microstrip transmission line, which enables signals to travel through the upper metal layer and the ground plane to be situated below. This particular line is the first one. It is possible to acquire a high capacitive quality factor (Qc) by putting the ground plane of the microstrip transmission line in close proximity to the signal plane using this method. Generally speaking, this is a problem that occurs with microstrip transmission lines. Consequently, this indicates that a microstrip line can be utilized to construct an inductor with extremely low values. An increase in inductance leads to a decrease in the inductive quality factor, which is the reason for this circumstance.

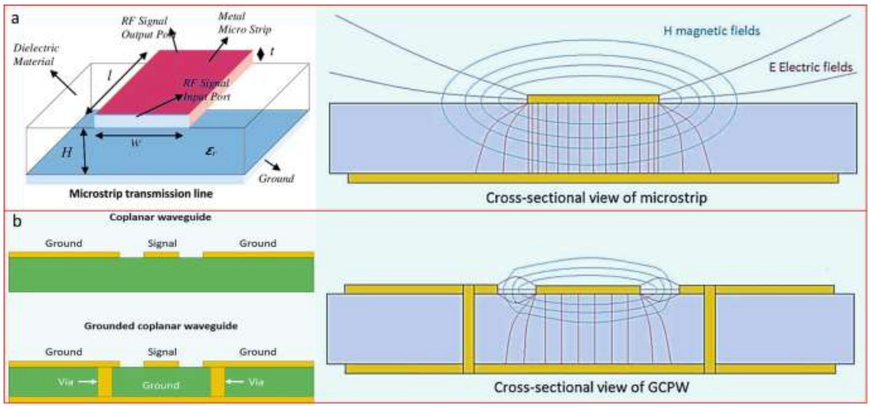


Fig. 2.6 Microstrip transmission line (a) and coplanar waveguide/grounded coplanar waveguide (b)

Source: Modifed from Coonrod 2015, credit: http://www.globalcommhost.com/rogers/acs/tec hsupporthub/en/docs/2015_IMS_MicroApps_microstrip_vs_coplanar_John_C_fnal_0 32315_SE.pdf

The coplanar waveguide (CPW) is the second form of transmission line. It has a higher QL than the microstrip line, but at the same time, it takes up a bigger area than the microstrip line. As can be seen in Figure 2.6b, CPW is comprised of two ground planes that surround the signal line. It is feasible to optimize both Qc and QL by adjusting the distance that separates the ground planes and the signal line. On the other hand, the microstrip line is limited to providing just fixed QL and QC values. Several potential design modifications, such as GCPWs, have been proposed in order to enhance the performance of standard CPWs.

2.3.2 Lumped approach

Illustration 2.2 demonstrates that it is possible to design square, octagon, and circular planer topologies for the installation of lumped inductors. These various topologies are all practical. Both the positive and negative aspects of these topologies are not incompatible with one another. Because it has the fewest number of sides, the square implementation is the one that is easiest to design. It is of poorer quality, but it has a larger inductance value than other options. The square implementation is the configuration that can be created with the least amount of effort. On the other hand, a circular construction is the most challenging to plan, but it has the highest quality factor after it is constructed. If you use an octagonal or hexagonal inductor, you will be able to achieve the same high quality factor; but it will require more room than a square inductor would make available.

You have the ability to make adjustments to one or more of these factors in order to obtain the best possible results with the lumped element's inductance value, quality factor, and resonating frequency.

- Track width of the inductor,
- Inner radius,
- Number of turns and
- Track spacing.

There is a trade-off that must be made between the series resistance of the lumped element and the substrate loss when constructing an inductor that has a compact footprint. This results in a greater quality factor as well as a reduction in the amount of substrate that is lost. The construction of passive components in circuits is done in close proximity to one another in order to reduce the number of impressions that are left on silicon. When this happens, the components are subject to the effects of coupling.

Due to the fact that coupling reduces both the quality factor and the self-resonating frequency of the lumped element, applications that use millimeter waves are required to give this matter special consideration. Using the lumped element technique, a number of different inductor topologies and on-chip passive circuits have been proposed for usage at mmWave frequencies. A lumped inductor/resonator that is constructed using SiGe technology, a bandpass filter (BPF) that is integrated with metal-insulator-metal (MIM) capacitors through the use of the BiCMOS process, a compact bandpass filter that is constructed using silicon-based Integrated Passive Device (IPD) technology, and an active tunable capacitor that is implemented are some of the components that are listed here. Through the utilization of several alternative process technologies, particularly III-V semiconductors, it is conceivable that active and passive mmWave circuits could be implemented more effectively. The passive components of these technologies are made possible by the utilization of a number of different silicon technologies.

2.4 COMPARISON OF SILICON AND III-V SEMICONDUCTORS

Applications that operate at high frequencies at millimeter waves are best served by III-V semiconductor technology, which makes use of materials such as GaAs, InP, GaN, and other materials that are similar to these. Due to this particular reason, mmWave circuits are capable of being constructed using this technology. As a result of the high resistivity of the substrate, GaAs/InP demonstrates enhanced electron mobility, power gain, noise reduction, and body effect reduction. The fabrication of high-Q passives is therefore possible through the utilization of III-V technology. On the other hand, CMOS is characterized by a multitude of intrinsic drawbacks, including a poor power gain, increased noise in the circuit as a result of the high sheet resistance of the polysilicon gate (about 10 Ω/square), and a low resistivity of the substrate (10 Ω-cm), which results in signal loss. It is important to keep in mind that CMOS has a faster heat conductivity than both GaAs and InP. In light of this, GaAs and InP have a lesser device density when compared to CMOS's device density.

Furthermore, GaN possesses a heat conductivity that is comparable to that of silicon, in addition to having an exceptionally high breakdown voltage. The wideband gap Due to the fact that GaN HEMT has a greater power added efficiency in this frequency range, it is an excellent option for power amplifiers that operate at mmWave frequencies. A comparison of the various material properties of silicon, GaAs, InP, and GaN can be seen in Table 2.1, which may be found below.

CMOS technology, on the other hand, has been the subject of substantial research for mmWave applications operating at high frequencies. However, this continues to be the case in spite of the numerous advantages that semiconductor III-V technology might provide. This is largely attributable to the fact that, as a result of developments in CMOS technology, it is now feasible to achieve a transition frequency (fT) that is comparable to that of III-V semiconductors. The function of fT becomes an extremely important performance parameter for transistors when they are operating at such high RF frequencies. Taking into account the financial component is also a crucial factor, since the adoption of the millimeter-wave frequency spectrum for 5G will primarily determine whether or not it is successful in the commercial market. When compared to semiconductor III-V technology, CMOS technology offers a lot of advantages due to the fact that its processes have been around for a long time.

Table 2.1 Typical material properties of GaAs, InP, GaN and Silicon

Materials	**F_t (GHz)**	**p (Ohm-cm)**	**p (cm^2/V-s)**	**Bs (eV)**	**K (W/m-10**	**V_n (V/ cm)**
Silicon	250GHz for 28 nm CMOS	10	13×10^3	1.12	150	3×10^5
GaAs	>500 GHz	10^7 to 10^9	8×10^3	1.42	50	4×10^5
InP	>700 Gib for InP HEMT	10^7	5×10^3	1.35	67	5×10^5
GaN	>500 GHz	$I0^6$ to 10^9	12×10^3	3.4	130	30×10^5

f_T: Transition frequency, ρ: Resistivity, μ: Electron mobility, Bg: Bandgap, K: Thermal conductivity, V_n: Breakdown voltage.

- It is cost efficient
- It has a higher manufacturing capacity
- It offers higher integration
- It has a higher manufacturing yield

Figure 2.7 provides an overview of the fundamentals of CMOS technology as well as a breakdown of the advantages and disadvantages of this technology in comparison to III-V technology. Circuits that use complementary metal-oxide semiconductor field-effect transistors (CMOS) typically include a nonlinear circuit device as an essential component. There is a common connection between complementary metal-oxide semiconductor (CMOS) technology and very large-scale integrated circuits (VLSIs), which are devices that combine millions or even billions of MOSFET transistors onto a single chip or die. The complementary metal-oxide semiconductor (CMOS) technology is the most often used for the production of very large scale integrated circuits (VLSI chips). This is owing to the fact that CMOS technology is characterized by its dependability, low power consumption, relatively low cost, and most crucially, significant scalability. CMOS-based low-noise amplifiers and other component designs have been proposed for use in 5G networking. These designs incorporate a variety of components. The invention of a technique for heterogeneous integration made it possible to make the most of the advantages that semiconductor III-V and CMOS technologies had to offer. InP HEMT and InP HBT chip lets are utilized in circuits operating at frequencies of 1 Hz and lower levels. These chip lets are constructed on CMOS wafers.

The silicon-on-insulator (SOI) and silicon-gate biCMOS technologies are two more silicon technologies that are being utilized for the implementation of mmWave circuit development. The CMOS technology is the semiconductor technology that is now in use. Because of its high transition frequency, which can exceed 300 GHz (for 55 nm fabrication technique), BiCMOS technology is widely utilized in mmWave applications. This results in the technology's widespread adoption. When contrasted with bulk CMOS technology, BiCMOS technology provides superior performance in the realm of radio frequency (RF) and a higher transition frequency respectively. This makes it an excellent candidate for use in 5G mmWave circuits, which is why it is being considered. When compared to bulk CMOS, BiCMOS is characterized by higher power consumption, lower integration densities, and higher production costs.

The silicon-on-insulator, or SOI, manufacturing technology is a second manufacturing process that is based on silicon. It is now being used to create millimeter-wave circuits. Efficiency-wise, it is comparable to that of BiCMOS. In addition, the 28 nm SOI technology has a transition frequency that is greater than 300 GHz, which is comparable to the transition frequency of the BiCMOS process technology. In comparison to the bulk CMOS process, the SOI technology eliminates two of the most significant restrictions. The low resistivity of the substrate and the high sheet resistance of the polysilicon gate are the restrictions that are being discussed here.

It has been abundantly clear that mmWave circuits for 5G applications have made extensive use of a wide variety of silicon technologies, such as bulk CMOS, BiCMOS, and SOI technology. Silicon technologies will continue to be the front-runner when it comes to the design of millimeter-wave circuits for 5G applications, despite the advent of semiconductor III-V technologies that are more efficient. On the other hand, millimeter-wave circuit assembly does not work well with the silicon device models that are now available in EDA tools for low-frequency operations. Such models are not a good fit for millimeter-wave circuit construction.

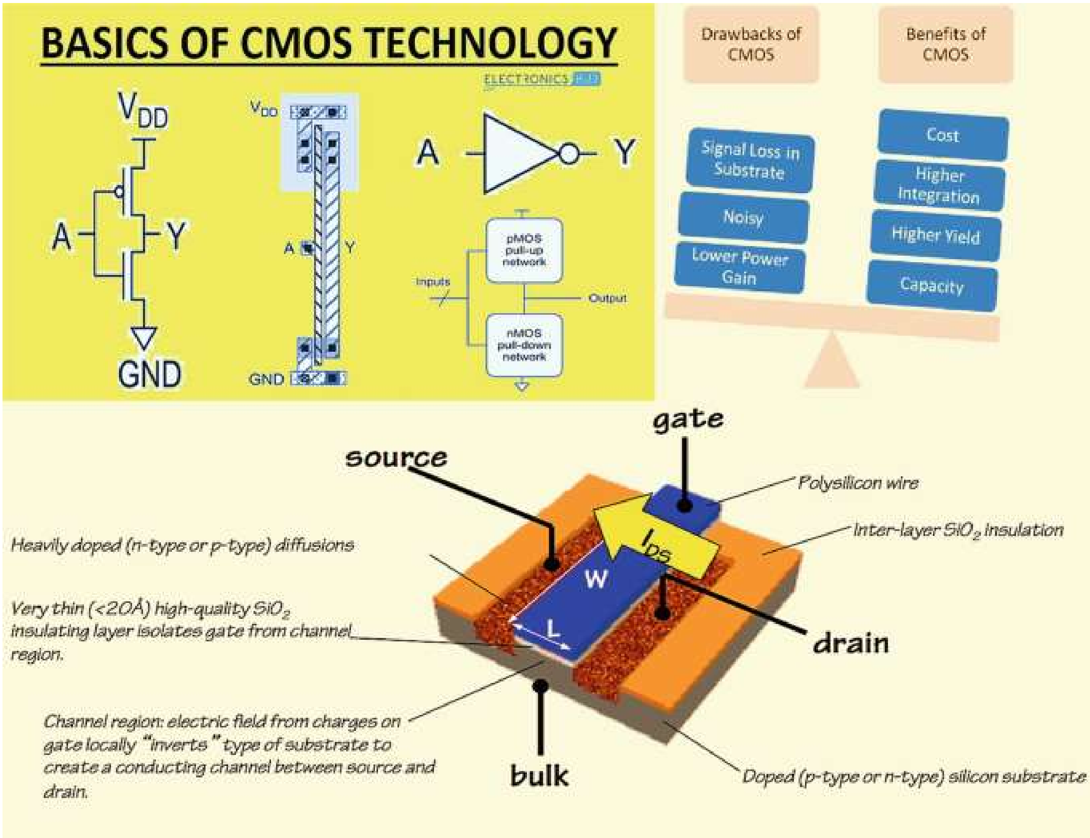


Fig. 2.7 Basics of CMOS technology and its benefits and drawbacks

Source: Modifed from Ward 2017, credit: M.I.T. Department of Electrical Engineering and Computer Science

2.5 TRANSISTOR MODEL DESIGN CHALLENGE IN CMOS TECHNOLOGY

When it comes to the construction of mmWave circuits using EDA tools, the most significant obstacle is the fact that the CMOS transistor models that are offered by these tools are not optimized for high-frequency operations. Black box models with S-parameter expressions and extra parasitic linkages are the types of models that are frequently used at the moment. Because of their dependence on measurements derived from already made devices and their unique shape, which limits design freedom, the present models do not perform up to expectations when using mmWave circuits. This is because of the fact that they are restricted in their shape. The effects of capacitive parasitic, for instance, are taken into consideration in a conventional CMOS transistor model, which is an excellent tool for the design of digital circuit applications. On the other hand, it is common knowledge that led inductance losses and resistive parasitic losses become significant at mmWave frequencies. The maximum working speed at these high frequencies is further restricted by the resistive losses that are produced by the gate, drain, source, and substrate terminals, in addition to the resistance of the channel. Due to this, there is a limit placed on the maximum speed at which the machine can operate.

When it comes to mmWave frequency, lead inductances, which are a type of lengthy connecting wire that creates delay effects, are also highly significant. Additionally, the wiring in question is of utmost significance. Figure 2.8a depicts the standard nMOS small signal model, and Figure 2.8b depicts the mmWave high frequency model, which contains parasitic resistive and inductive components. Both of these models are shown in Figure 2.8. There is a wide range of frequencies that are included in both categories.

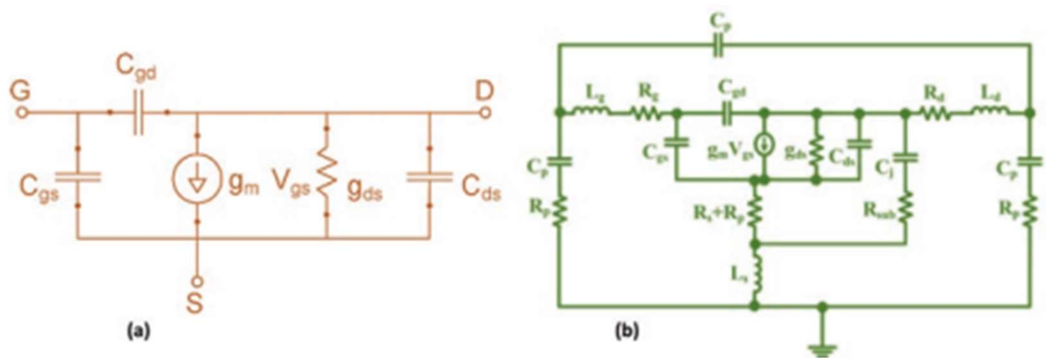


Fig. 2.8 Analyzing two different models of nMOS, one with a traditional tiny signal approach and the other with mmWave high frequency and parasitic elements that contribute resistance and induction

Source: Adapted from Doan et al. 2005; Sialm et al. 2006 with permission from IEEE)

When modeling active devices such as CMOS transistors for use at mmW frequencies, EDA tools commonly adopt the method of introducing parasitic elements into the current models that are accessible in the process development kits (PDKs) of the tools. This method requires the incorporation of parasitic components into the models. Through the utilization of this technology, a number of CMOS models for mmWave applications have been proposed. The development of these models has involved taking into account both internal and external elements. According to the findings of an investigation into the performance of mmWave circuits, the intrinsic factors (Cgd, Cgs, and Cds) had a less impact than the extrinsic parameters (terminal resistances, substrate effects, and parasitic coupling). In addition, it is essential to take into account the fact that the transistor kinds that are now accessible are intended for fixed configuration. In addition to this, it is essential to take into consideration the manner in which particular arrangement characteristics, such as connections to various terminals, influence parasitic sections.

Figure 2.9 presents a selection of layout design concepts that can be implemented for nMOS devices operating at mmWave frequency. Two methods that can be utilized to significantly reduce the impact of parasitic gate resistance are the utilization of a low-resistivity Metal 1 ring for gate connections and the utilization of a large number of gate connections from both sides of the transistor. Because of this, the parasitic gate resistance can be significantly lowered, which is a significant benefit. Another effective strategy for the layout of the mmWave frequency is to make use of a number of short gate fingers throughout the design process.

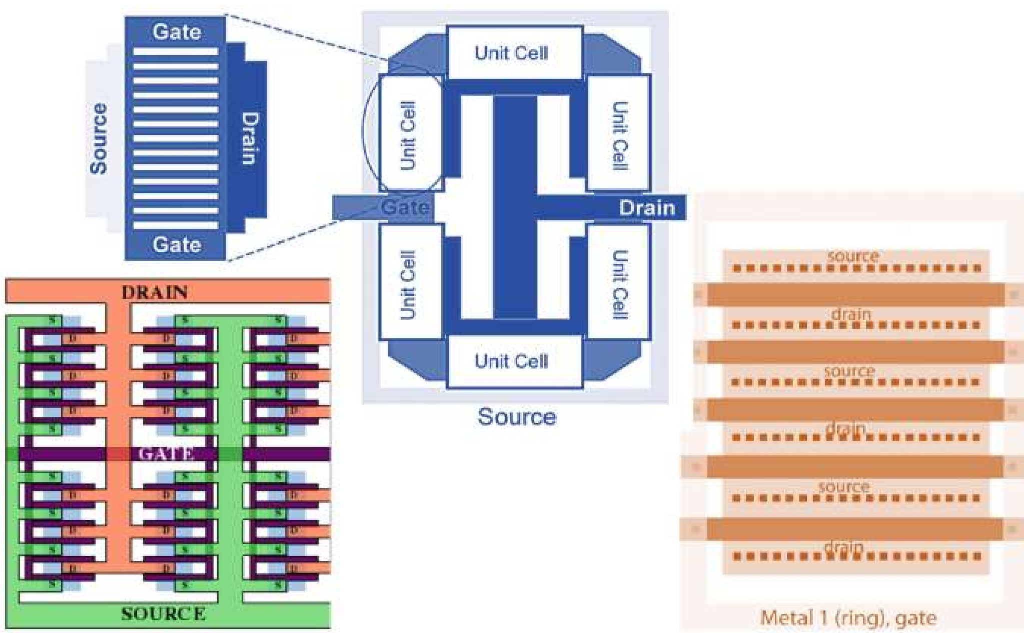


Fig. 2.9 Example layout design approaches for nMOS at mmWave frequency

Source: Modifed from Juneja et al. 2021 with permission from Elsevier)

2.6 GaN and GaN-on-SiC WIDE BANDGAP SEMICONDUCTORS FOR 5G APPLICATIONS

It is necessary to achieve new criteria for data speed, latency, and capacity in order for 5G technologies and infrastructure architecture to be able to effectively supply connections that are capable of meeting these objectives. Micro-level densification of power consumption on individual devices will also be important, in addition to the macro-level densification that will require an increase in the number of base stations. This will be necessary in order to meet the requirements of the network. In order to effectively meet the requirements of a variety of applications, the architecture of the current telecommunications infrastructure requires technology that is capable of providing optimal results. Some of the characteristics that are taken into consideration are heat, velocity, power, efficiency, location, and financial resources. According to the information shown in Figure 2.10, the development of 5G radio frequency technology can make use of a wide variety of semiconductors.

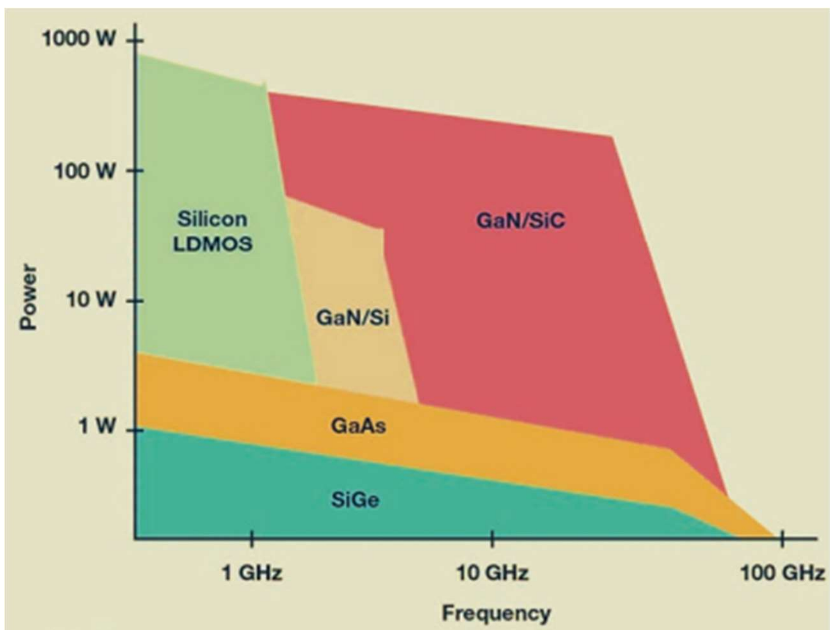


Fig. 2.10 Comparing power and frequency of different semiconductor materials in the microwave range, which includes mm Waves

Source: Adapted from Benson 2017, credit: https://www.analog.com/ en/analog-dialogue/articles/rf-power-amplifers-go-wide-and-high.html)

Despite the fact that it operates at a relatively modest voltage of 2 to 3 volts, silicon germanium (SiGe) technology stands out due to the integration benefits that it provides. Due to the fact that the material operates within a voltage range of 5 V to 7 V, it is ideally suited for use in microwave power amplifiers, where it has found widespread application. Due to the fact that its major usefulness is below 4 GHz, the 28 V silicon LDMOS (laterally diffused metal oxide semiconductor) technology is not as popular for use in broadband applications, despite the fact that it has found some application in the telecommunications industry. GaN technology, which runs from 28 V to 50 V and performs on a substrate that is low-loss and very thermally conductive like silicon carbide (SiC), has made it possible to open up a wide range of new possibilities. At the moment, GaN on silicon technology can only operate at frequencies that are lower than 6 GHz. Because of the radio frequency losses that are linked with the silicon substrate and the fact that it has a lower thermal conductivity in comparison to SiC, the performance of the gain, efficiency, and power is negatively impacted as the frequency increases.

2.6.1 Characteristics of GaN Devices Applied in 5G Technology

As a direct bandgap semiconductor, gallium nitride (GaN) is distinguished from other materials such as gallium arsenide (GaAs) and silicon by virtue of its bandgap width of 3.4 electron volts (eV). The stability of combinations that contain GaN is exceptionally high. Besides being a product that is difficult to work with, it also has a high melting point, which is approximately 1700 degrees Celsius. When subjected to the pressure of the atmosphere, GaN crystals frequently exhibit formations known as hexagonal Wurtzite. As a result of these exceptional characteristics, it has been prepared for deployment in the field of 5G technology in an environment that is optimal.

(a) **High Pressure Resistance and Radiation Resistance:** GaN crystals are able to withstand electric forces that would cause silicon devices to collapse. This is because the chemical bonds that they possess are substantially stronger. Compared to the fields that GaN crystals are able to manage, these fields are far higher. It is radiation resistant and demonstrates great pressure resistance due to the fact that GaN has a large bandgap width. This means that electrons are not easily forced into the conduction band, and the interference signal has little impact on the device. In addition to being able to withstand greater voltages, these qualities also allow the device to generate 5G signals that are stronger than those produced by other devices. Additionally, because of the radiation resistance of the equipment, the signals are protected from signals that could potentially cause interference. Because this helps to keep the 5G signals in an environment that is generally stable, they are able to transfer data with a higher degree of precision.

(b) **High Frequency:** Considering that GaN has a relatively low gate charge, it is necessary to add more gate charge with each switching cycle. The conclusion that can be drawn from this is that decreasing the gate charge will enhance the likelihood of the device reaching the high frequency state. In spite of the fact that GaN is capable of operating at frequencies as high as 1 MHz without experiencing any loss in efficiency, it is unable to maintain its exceptional performance in comparison to other materials. An illustration of this would be the difficulty that silicon has in achieving frequencies that are higher than 100 kHz. Both the strong chemical bond and the high loading capacity of GaN contribute to a significant reduction in the gap that exists between the electrical terminals of the transistor. As a result of this, electrons can be transformed in a shorter amount of time. Furthermore, it is worth noting that the GaN conduction band's lowest point locates itself in close proximity to the Λ point, and there exists a substantial energy differential between this point and the other conduction band energy dips. As a consequence of this, the electron drift rate does not reach saturation in a short amount of time, and it is not simple to generate electromagnetic scattering. In the process of forming a heterojunction, semiconductor materials such as AlGaN are mixed with GaN, which results in the generation of a highly mobile two-dimensional electron plasma. As a consequence of this consequence, the electronic gadgets that it manufactures have switching characteristics that are faster. When it comes to the introduction of 5G technology, which is designed to increase the speed at which calculations are performed, this kind of feature will be quite important.

(c) **High Operating Temperature:** The intrinsic excitation of GaN is lower when compared to that of other semiconductors with tight bandgaps that operate at the same temperature. This indicates that the GaN-based device is capable of transmitting signals with a higher signal-to-noise ratio than other devices. GaN is characterized by a bandgap that is significantly larger than that of normal semiconductors, which is the reason for this phenomenon. As a consequence of this, its working temperature is higher, which is an essential characteristic for the circuit to possess in order for it to function in environments that have higher power and temperature, such as the 5G base station. Furthermore, the outstanding heat dissipation characteristics and high thermal conductivity of GaN become even more obvious when it is utilized in conjunction with a SiC substrate. As a result of these qualities, GaN is able to function exceptionally well in hot environments.

(d) Low Energy Loss: Due to the fact that GaN possesses a high pressure-bearing energy level, it is possible to build a transistor with a shorter distance between each terminal. Consequently, there is the potential for a reduction in the resistance loss as a consequence of this. The high mobility and the high carrier concentration both contribute to a decrease in the resistivity of the material. This ultimately results in electronic devices that have a low conduction resistance, which enables even less energy to be lost due to resistance. It is not completely out of the question to consider the possibility that a 5G signal could generate more power while consuming reduced amounts of energy. As a result of this characteristic, the transmission of signals via 5G may require less power than was previously believed.

2.6.2 GaN Power Integration for MMIC in 5G Technology

Controlling signal power is accomplished through the use of beamforming and multiple-input multiple-output (MIMO) technology in the 5G architecture. This allows for larger over-the-air data capacities to obtain. It is necessary to have a phased-array transceiver system in order to do massive MIMO beamforming. This system must combine a number of radio frequency (RF) circuits into each antenna element. As a result of this, the most important components of the base station design and the handset architecture are the size, cost, and power density of the devices. GaN is an apparent choice for base-station deployment, despite the fact that other semiconductor technologies are better suited to phones for reasons like as pricing, battery voltage, and energy requirements for radio frequency (RF) transmission. The development of switches, low-noise amplifiers (LNAs), and frequency converter circuits has been made possible as a result of ongoing work to modify GaN so that it can operate at higher frequencies and with lower voltages. When the time comes, one or more GaN micromachines (MMICs) will be able to accommodate the multitude of RF chains that will be required.

GaN Power Integration for MMICS:

Components such as inductors, thin-film resistors (TFRs), metal insulator-metal (MIM) capacitors, metal insulator-gate (ON) Schottky gate AlGaN/GaN HEMTs, Schottky barrier diodes, and thin-film resistors are utilized in the manufacturing process of monolithic microwave integrated circuits (MMICs). These MMICs are manufactured using GaN power integration procedures. Figure 2.11 depicts a high breakdown voltage millimeter-wave GaN HEMT integrated circuit (b) and a typical 100 nm GaN-on-Si MMIC process (a) side by side.

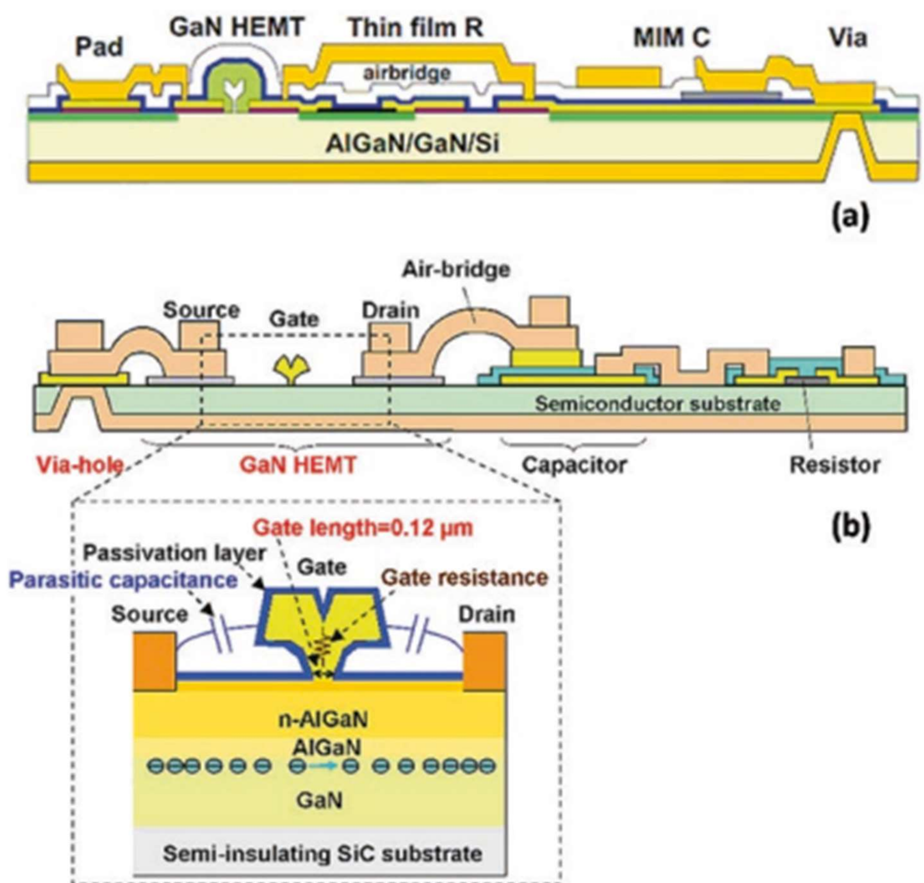


Fig. 2.11 Schematic cross-sections of 100nm GaN-on-Si process for MMICs (a) and high breakdown-voltage millimeter-wave GaN HEMT integrated circuit (b)

Source: Modifed from Fujitsu 2009; credit: Fujitsu Limited; and Rocchi 2016 with permission from IEEE)

Both of these circuits are embedded in silicon. When GaAs devices are incorporated into MMICs, they provide wideband-gap (WBG) solutions that are capable of 500 GHz and have remarkable performance. The aircraft industry, mobile devices, and the backbone of communications are all excellent examples of applications that are ideal for these technologies. The most notable type of GaN micromachining integrated circuits (MMICs) is comprised of high frequency power amplifiers (PAs) that have an operating frequency that is greater than 50 GHz. As an additional component, there are voltage-controlled oscillators (VCOs), components for transmitters and receivers, components for modulators, and high linearity low noise amplifiers (LNAs).

Dialogues are a good example because of their capacity to produce a high-voltage, one-way blocking situation. This makes them an indispensable component in power integrated circuits. In addition, diodes can be utilized as voltage level shifters by making use of the forward voltage loss that they exhibit. A common type of barrier diode is known as the Schottky barrier diode (SBD), which is created by depositing a metal based on nickel on top of an AlGaN or GaN surface. This process is illustrated in Figure 2.12a. Lateral Feld effect rectifiers (L-FERs) make use of a two-dimensional electrochemical channel that is controlled by a Schottky gate between the anode and the cathode. L-FER is a rival that is well-qualified for use in smart power integrated circuit platforms (Figure 2.12b). This is due to the fact that it has a lateral device shape and is process compatible with AlGaN/GaN HEMT systems. Some examples of tri-gate Schottky diodes are shown in Figure 2.12c. These diodes have a lower turn-on voltage. These tri-anode diodes are able to gain access to the multichannel because they make use of the fn sidewalls as a point of Schottky contact.

In addition to this, it demonstrates a great deal of potential for the management of off-state AlGaN/GaN Fin-FET operation. Through the use of GaN integration, it is feasible to convert power to regions where silicon power cannot reach. An increasing amount of technology originating from the field of silicon electronics has been utilized in the integration of GaN. Direct bonding of the individual Si CMOS gate drivers to the components or packaging of the drivers in the same case are both viable options.

At this time, 3D GaN is being installed. Further strategies are now being considered for implementation. On the other hand, in order to integrate GaN, a robust platform that has active and passive components that have been appropriately calibrated and modeled should be constructed. Consequently, we are in a position to reliably estimate the operation of the GaN power integrated circuits. It is of the utmost importance to update

GaN power integration in order to keep up with the continuously advancing technology of discrete GaN devices. For the purpose of making the most of GaN's potential capabilities, the path to GaN power integration is comparable to the path that leads to Si power IC, but it is more expedient.

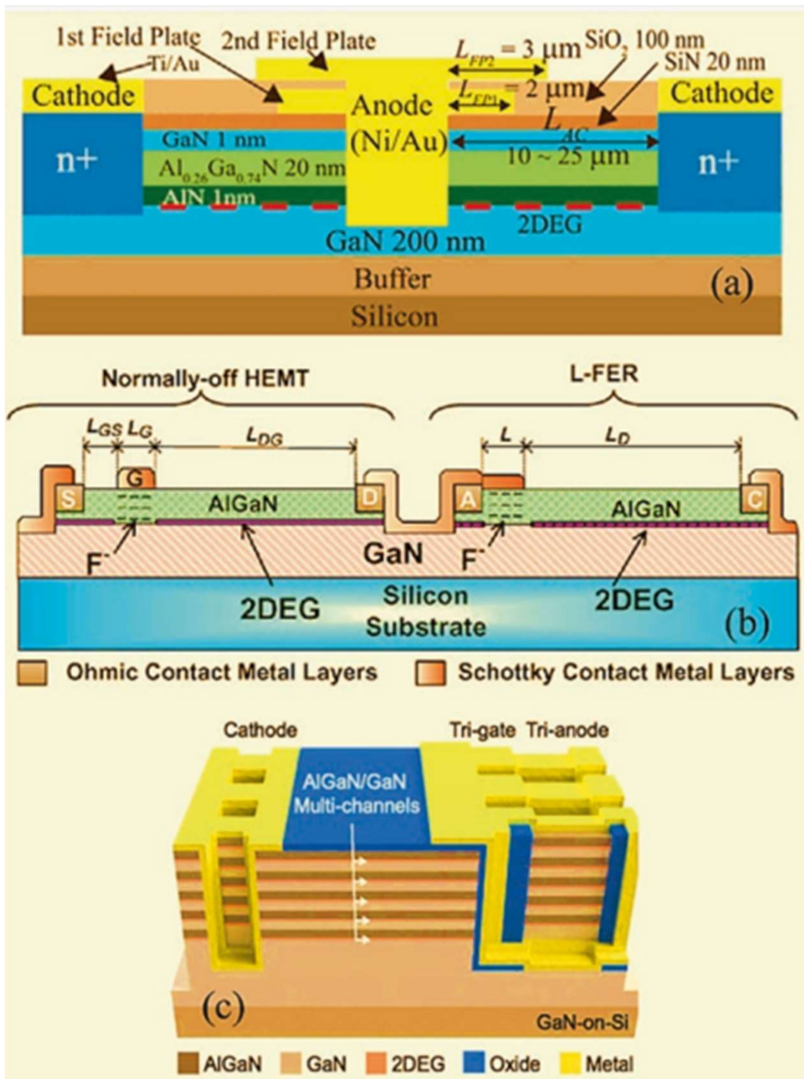


Fig. 2.12 Schematic structure of (a) a typical Schottky barrier diode

Source: Adapted from Zhu et al. 2015 with permission from IEEE), (b) a L-FER with a normally-OFF HEMT. (Chen et al. 2008 with permission from AIP Publishing) and (c) a tri-gated Schottky barrier diode. (Ma et al. 2019 with permission from IEEE)

GaN Base Station Pas:

Considering that each antenna is controlled by its own PA, it is of the utmost importance to meet the power and linearity requirements while simultaneously minimizing variation between cells in MIMO implementations. Small-cell base station PAs that are based on 5G GaN are required because they are vital for making deployment easier. This is because they are lighter, more compact, and less expensive than other base station PAs, all while maintaining high power and efficiency. A comprehensive understanding of the ways in which the unique properties of GaN, such as its breakdown voltage, self-heating, trapping, Feld plate design, and transconductance shape, influence the operating frequency, power, and efficiency (PAE) of the device is required.

Among the most important principles are linearity (which encompasses harmonics, EVM, ACPR, IIP3, AM-AM, and AM-PM), robustness, and transient behavior. Some of the available choices, such as GaAs FETs and Si LDMOS, are not capable of satisfying the demand for high density power; however, GaN will be able to fulfill this requirement. For this to be accomplished, modeling high-power devices using dynamic DC and RF approaches will require a significant amount of effort on the part of the individual.

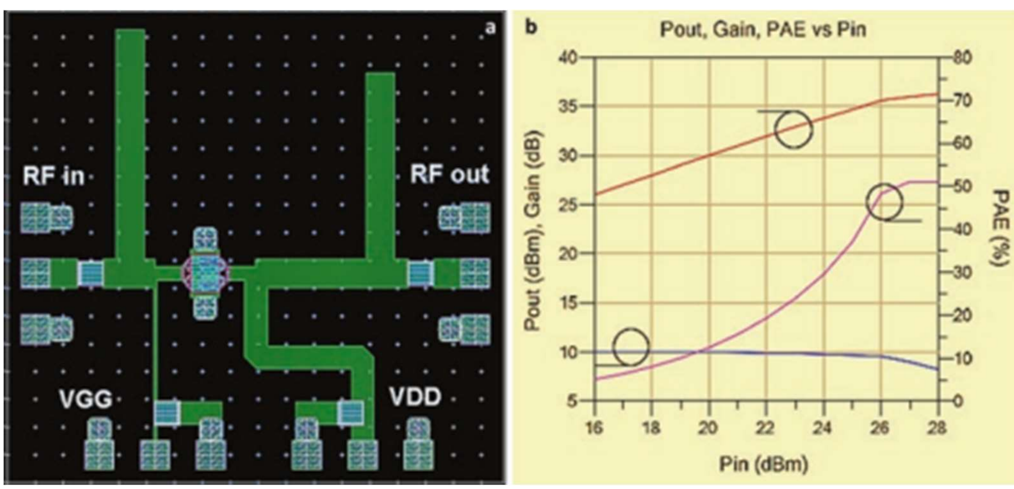


Fig. 2.13 5G PA using a 0.20 um GaN 8 × 100 um FET (a) MMIC and (b) Pout, Gain and PAE performance

Source: Adapted from Yuk et al. 2017 with permission from IEEE)

To construct a 28 GHz Class AB PA MMIC with dimensions of 1.8 mm x 1.7 mm, a single 0.20 μm GaN 8 × 100 μm FET device was utilized. This MMIC was constructed from the bottom up. Diagram 2.13a depicts the layout of the circuit, and diagram 2.13b depicts the performance that was simulated when the circuit was being used. As shown in Figure 2.13b, this demonstration PA is capable of achieving a power amplification efficiency of 51.1% and a small-signal gain of 10.03 dB when the input frequency is 27 dBm. The power output is measured at 36.01 decibels millimeters. In accordance with the design, the current state of GaN technology is capable of producing first-generation 5G systems. This is because the design makes use of the PA output stage of a single MIMO transmitter, which causes this to take place. It is possible to use this PA in conjunction with the Doherty design for base station deployment, which is seen in Figure 2.14a.

GaN Frequency Synthesis:

There will be a significant rise in the number of antenna elements that are utilized in 5G MIMO applications. It will become increasingly difficult to precisely produce and scatter coherent local oscillator (LO) power as the number of transmitters continues to increase. Increasing the low-level power through the use of a public address system (PA) is one straightforward action that may be taken to solve these concerns. Due to the fact that 5G carrier transmissions will first begin in the sub66 GHz range, it is of the utmost importance to provide interoperability with lower frequency cellular bands. The LTE bands, the GSM850/900 bands, and the DCS/PCS bands are all included in this. Consequently, the solution to the challenge of MIMO signal distribution and compatibility is to make use of high power frequency multiplication in order to give sufficient power at the frequencies that are wanted. Using the same methods, it is possible that one day a 5G low-frequency (LO) signal in the millimeter wave region will be generated.

The utilization of high-power GaN frequency multipliers allows for the generation of a low-frequency (LO) signal that is both high-frequency and high-power, derived from a reference frequency that is lower in frequency. As can be shown in Figure 2.14b, the output that is created can be precisely disseminated to each massive MIMO chain when a passive network is utilized. It is possible for GaN devices to multiply frequencies through the use of harmonic augmentation techniques, which enables them to supply power at frequencies that are greater than fT. The development of GaN technology, which enables the synthesis of high harmonics without breaking down, is another promising area for technical advancement. GaN technology has the potential to evolve in the future.

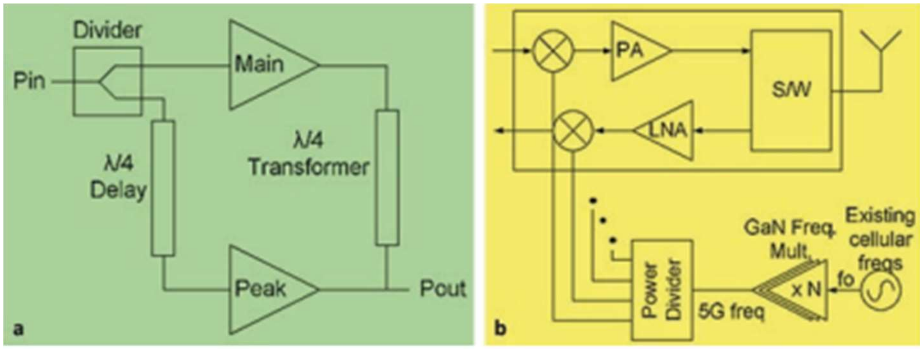


Fig. 2.14 GaN Doherty PA for base stations (a) and Frequency conversion and distribution using high power GaN frequency multipliers (b)

Source: Adapted from Yuk et al. 2017 with permission from IEEE)

Table 2.2 Commercially available GaN foundry services

Foundry service			Discrete
Process	**Bias**	**Frequency**	
0.25 pm GaN-on-SiC	28-40 V	18 GHz, 30 GHz	Y
0.40 pm GaN-on-SiC	28 V, 50 V	8 GHz	
0.25 pm GaN-on-Si	N/A	N/A	N
0.50 pm GaN-on-Si	N/A	N/A	
0.50 pm GaN-on-SiC	N/A	N/A	
0.20 pm GaN (100 mm substrate)	N/A	60 GHz	Y
0.10µm GaN	N/A	>70 GHz	N
0.25 pm GaN-on-SiC (100 mm substrate)	40 V, 48 V	10 GHz, 18 GHz	Y
0.15 pm GaN-on-SiC (100 mm substrate)	28 V	40 GHz	
0.50 pm GaN-on-SiC (100 mm substrate)	65 V	10 GHz	
0.50 pm GaN-on-SiC (76 mm substrate E-mode)	N/A	N/A	N
0.15 pm GaN-on-SiC (76 nun substrate)	N/A	Ka-band	
0.50 pm GaN-on-SiC (76 mm substrate)	40 V	X-band	

GaN technology, which is now suited for commercial application, is attracting the attention of an increasing number of enterprises in the microwave and radio frequency industries. Despite the fact that the development of GaN is primarily focused on the next-generation PA technology, the material is also being evaluated for usage in a variety of circuit applications. This undertaking is doable due to the extensive variety of foundry possibilities that are presented in Table 2.2. When it comes to channel length, the current best practice is somewhere in the range of 0.10 to 0.15 square micrometers. The expansion of GaN will be driven by two significant industries: cellular communications and satellite communications.

CHAPTER 3

5G ANTENNA DESIGN AND PERFORMANCE ENHANCEMENT

3.1 INTRODUCTION

The antennas that are used in 5G devices are extremely important, and they are required to meet certain criteria, including greater gain, decreased radiation losses, and increased bandwidth. Therefore, because there are numerous ways for improving performance that are related to the features of the antenna, antenna design becomes extremely important for 5G devices. The design of antennas for 5G networks can be roughly divided into two types, according to the classification system. structures that are dependent on input/output ports include MIMO and SISO, among other others. Both are classified as wideband and multiband, with the classifications being determined by the way in which each frequency reacts.

When it is necessary to integrate with 5G devices that are responsible for powering the Internet of Things, you have the option of selecting either single-element or multi-element SISO antennas. The presence or absence of a metal ring is the determining factor in determining whether a multielement antenna is categorized as a wideband or multiband MIMO antenna. It is possible to employ massive MIMO antennas in base stations; however, these antennas are not the most suitable choice for mobile devices such as smartphones. Increasing transmission speeds is accomplished by the utilization of carrier aggregation in metal rim antennas that have multiple inputs and multiple outputs (MIMO). In addition to this, design factors such as orthogonal polarization help to improve isolation, which in turn leads to a gain in efficiency.

There is also the possibility of categorization based on their antenna types. In this article, you will discover a comprehensive discussion of each of these antenna types, as well as a suggestion for how to improve their performance. These enhancement procedures have a major impact on the electrical and physical features of an antenna, which ultimately results in an improvement in the antenna's overall performance.

3.1.1 Classification Based on Input and Output

It is demonstrated in Figure 3.1 that the antenna can be loosely categorized according to the input and output ports that it possesses.

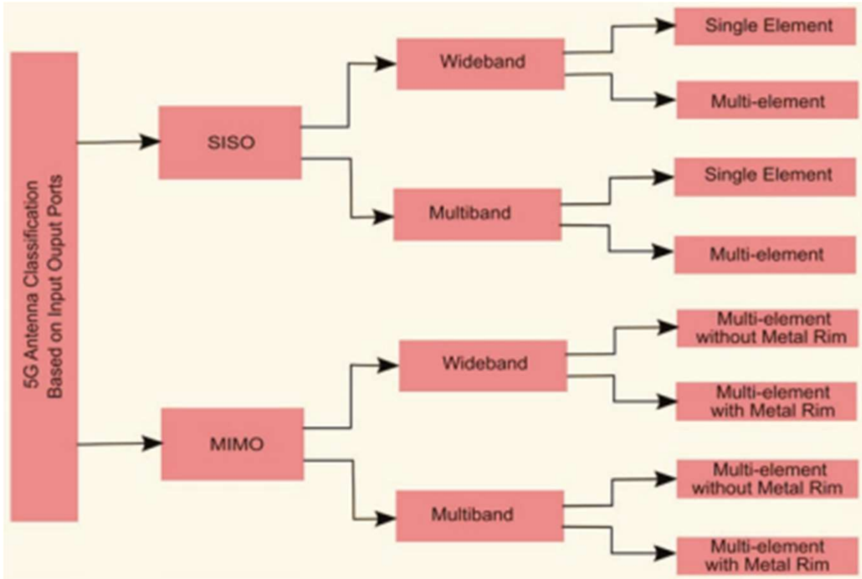


Fig. 3.1 5G Antenna classification based on input output ports

Source: Adapted from Kumar et al. 2020 with open access (IEEE))

(a) **Single Input Single Output (SISO) antenna:** Because of its uncomplicated design and execution, the SISO antenna is well-suited for use in 5G applications not only as a single-element antenna but also as a multi-element antenna. For a single element antenna to attain a high gain, it is necessary for the antenna to have a large dimension. Signals operating at frequencies greater than 6 GHz are subject to increasing propagation losses, which ultimately results in a decline in the quality of the service. Because of this, it is necessary to switch from a single-element antenna to a multi-element antenna in order to get consistent and suitable performance. In point of fact, the primary purpose of a multielement antenna is to maximize the gain of the antenna; however, this comes at the expense of greater design complexity and bulk.

(b) **Multiple-Input Multiple-Output (MIMO) antenna:** On the other hand, interference, multipath fading, and radiation losses are just some of the issues that may arise during wireless transmission. Moreover, the severity of the issue increases in direct proportion to the frequency with which it occurs. Multiple-input multiple-output antennas, often known as MIMO antennas, are extremely important in solving these challenges since they reduce the amount of signal power while simultaneously expanding the transmission range. For this reason, it is possible that 5G networks would employ the MIMO architecture in order to achieve maximum efficiency, the lowest possible latency, and the highest possible throughput. MIMO, which stands for multiple inputs and multiple outputs, allows for the intelligent transmission of a greater number of signals across a network that consists of many antennas, which results in a significant increase in the channel capacity.

In order to reduce the number of antennas that are required, the multiple-input multiple-output (MIMO) technology makes use of multiband antennas. These antennas are capable of performing numerous wireless applications simultaneously. Additionally, MIMO antennas are categorized as either wideband or multiband depending on the frequency band in which they operate. This classification is based on frequency spectrum. A categorization that is even more exact can be achieved by separating the wideband and multiband antennas into those that have metal rims and those that do not have anything at all. Not only does the metal rim antenna give mobile phones a more fashionable appearance, but it also offers great mechanical robustness.

Additionally, in order to get a greater transmission rate, small devices choose the MIMO antenna that has enhanced isolation. Increasing gain, improving isolation (mutual coupling), increasing bandwidth, increasing envelope correlation coefficient (ECC), and boosting efficiency are all possible enhancements that can be applied to a wide range of antenna configurations through the utilization of various engineering techniques. Mutual coupling (MC) is the word that is used to describe the electromagnetic contact that occurs between antenna elements in multiple-input multiple-output (MIMO) signal processing systems. It is possible that the receiver of one antenna will pick up energy from the transmitter of another antenna over the course of this procedure. As a consequence of this, one of the most important aspects of MIMO is the reduction of the reciprocal coupling that exists between the different pieces of the antenna. You can use this mathematical formula to determine the answer to your question.

$$MC_{mn} = 1 - \frac{1}{N}\sum_{m}\sum_{m \neq n} \exp\left(-\frac{2x_{mn}}{\lambda}(\alpha + n\pi)\right) \quad \text{....3.1}$$

where MC_{mn} and x_{mn} the mutual coupling and the distance between the mth and nth antenna elements, respectively, determine the mutual coupling. The coupling level is controlled by the parameter α, whereas the number of MIMO elements is denoted by the letter N. Calculations are often made in the form of scattering parameters, and the resulting value is expressed in decibels.

3.1.2 Classification Based on Antenna Types

A further alternative is to group antennas according to their classifications, as seen in Figure 3.2. The following are some ways in which the several types of antennas that are compatible with 5G might be classified:

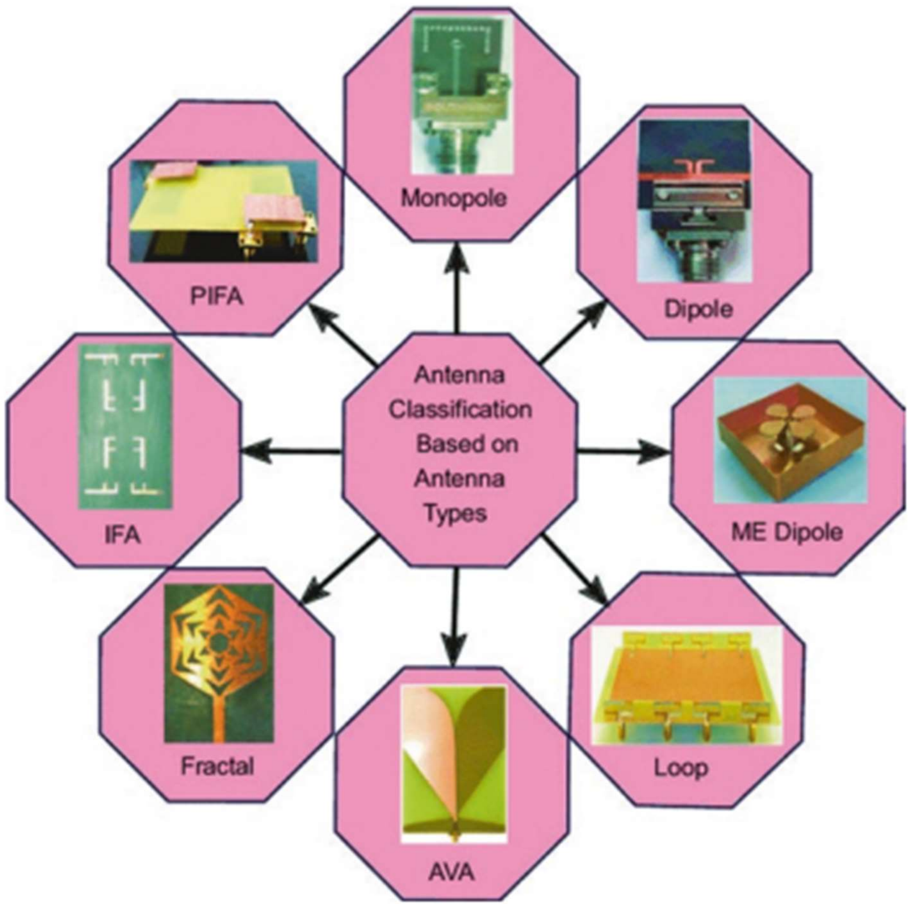


Fig. 3.2 5G Antenna classification based on antenna types

Source: Adapted from Kumar et al. 2020 with open access (IEEE))

(a) Monopole Antenna: The structure is formed by a microstrip line that is straight and measures λ/4 in length. Within the realm of antennas, the resonant working frequency is symbolised by the symbol λ, which is a sign that stands for wavelength. It is possible to modify its fundamental construction into a variety of shapes, including conical, spiral, and others, so that it can be used for a wide variety of applications that adhere to established criteria.

(b) Dipole Antenna: Specifically, it is composed of two microstrip lines that are straight and each have a length of λ/4. Additionally, feeding is supplied between the two microstrip lines. Consequently, the overall length of the dipole antenna is equal to λ/2.

(c) Magneto-electric (ME) Dipole Antenna: Both an electric dipole and a short planar magnetic dipole that is aligned vertically are the two fundamental components that make up this structure. The underside of the substrate is where the magnetic dipole gets its supply of food.

(d) Loop Antenna: There is the possibility that one of its constituent pieces is a ring of whatever shape you can think of, whether it be round, square, rectangular, or something else different. A loop antenna has a lower ratio of wavelength to radius than other types of antennas.

(e) Antipodal Vivaldi Antenna (AVA): The process begins with a ring, which may be shaped in any way you can imagine, from a circular to a square to any other shape you can conceive of. It is possible to lessen the difference between the wavelength and the radius by comparing it to the radius of the loop antenna.

(f) Fractal Antenna: It is made up of a great number of separate instances of the same structure occurring in chronological order. To accomplish this, an iterative mathematical method is utilised as the necessary tool. The fractal antenna can be utilised to create a wide variety of shapes, including but not limited to rectangles, circles, stars, triangles, and even leaves.

(g) Inverted F Antenna (IFA): The fundamental elements consist of a series of repeated repeats of the same pattern in subsequent order. This issue can be resolved by employing a mathematical method that incorporates iterative processes. In addition to the fractal antenna, there are a great number of additional forms that can be utilised; some examples include triangles, rectangles, rounds, stars, and leaves.

(h) Planar inverted F antenna (PIFA): The patch antenna and the ground plane are both components that are included in this part. Shorting pins are used to connect the two sides of the substrate, and the feeding process is carried out from the bottom side of the substrate. Its resonance, which takes place at a quarter of a wavelength, requires less space than other frequency resonances.

Table 3.1 provides a comparison and contrast of the various antennas, drawing attention to the benefits and drawbacks associated with each.

Table 3.1 Advantages and disadvantages of different antenna types

Antenna type	Advantages	Disadvantages
Monopole	Simple to design and fabricate in multi-element monopole antenna design, it can be easily rotated in any direction	Less gain Requires large area of ground Gives poor response in bad weather condition
Dipole	Simple to design and fabricate Receives balanced signal	Less gain Cannot be used for long range communication Low bandwidth
Magneto-Electric (ME) Dipole	High front to back ratio Low side lobe and back lobe level Wide bandwidth Low cross polarization	Design and fabrication are complex Costly
Loop	Easy to design Provides good channel capacity	As single element loop antenna cannot meet the 5G requirements, multi-element loop antenna is required Low gain
Antipodal Vivaldi Antenna (AVA)	Enhances the gain Provides wider bandwidth Gives stable radiation pattern	Requires more space Low gain at lower frequencies
Fractal	It helps to miniaturize antenna size Provides wider bandwidth good impedance matching Provides consistent antenna performance over the operating range	Design is complex Limitation on repetition of fractal design
Inverted F Antenna (IFA)	Smaller in size good impedance matching due to intermediate feeding	Narrow bandwidth Low gain
Planar Inverted F Antenna (PIFA)	Low profile good impedance matching Enhances front to back ratio	Narrow bandwidth Low gain

3.2 PERFORMANCE ENHANCEMENT TECHNIQUES FOR 5G ANTENNA DESIGN

In the process of antenna design, a wide range of performance enhancing tactics have been utilized extensively. Some of the parameter changes that have been targeted by these techniques include a reduction in mutual coupling, an increase in efficiency, a drop in bandwidth, and a reduction in the size of the device. A number of approaches for decoupling and improving antenna performance are displayed in Figure 3.3. These techniques are applicable to both SISO and MIMO antennas, and the targeting parameters for each technique are additionally displayed.

3.2.1 General Antenna Performance

The techniques depicted in Figure 3.3, which enhance the performance of antennas, can be of considerable assistance to the development of new and improved methods for enhancing antennas for 5G networks. The advantages and disadvantages of a number of different approaches are outlined in Table 3.2. The following are some examples of what these methods can look like:

- **(a) Substrate Choice:** In order to ensure that an antenna installation is successful, the first and most important requirement is to select an appropriate substrate. Manufacturers of antennas have access to a diverse selection of substrates, each of which possesses a unique collection of characteristics, such as permittivity's and loss tangents at their disposal. It is necessary to select a substrate that has a low relative permittivity and a low loss tangent if you want to achieve the goals of increasing gain while simultaneously reducing power loss.
- **(b) Corrugation:** When referring to radiators, the term "corrugation" refers to the process of removing a metal component from the edge of the radiator. This component can present itself in a variety of shapes, such as a square, sine, triangle, or rectangle, among others. The end result is an increase in both the bandwidth and the front-to-back ratio of the microphone.
- **(c) Multielement:** By utilising the multielement antenna, it is possible to achieve an even more significant improvement in antenna gain. In addition to this, it provides an increase in both the efficiency and bandwidth of the antenna. An antenna with several elements will perform better than a single-element antenna in situations where the requirements, which may include a wide bandwidth and a high gain, are too severe for the single-element antenna.

(d) **Dielectric Lens:** The gain and directivity of an antenna are both boosted as a result of the dielectric lens's ability to transmit electrostatic radiation in a unidirectional manner. In addition to the fact that the structure and shape of a dielectric lens are both subject to interpretation, the various combinations of substrate materials and substrate kinds are also open to interpretation.

(e) **Mutual Coupling Reduction Techniques:** When developing an antenna that contains a number of components, it is essential to keep in mind that the performance of each component is influenced by the performance of the following components. In order to assist with this matter, a number of different solutions for mutual coupling have been implemented.

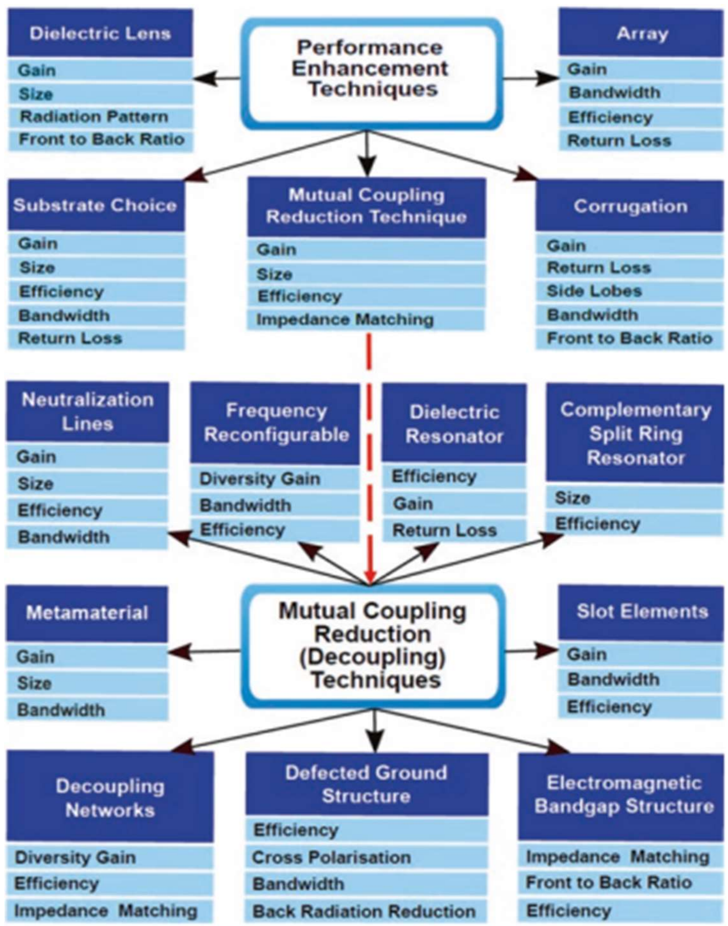


Fig. 3.3 Various performance enhancement techniques for 5G antenna

Source: Modifed from Kumar et al. 2020 with open access (IEEE)

Table 3.2 Advantages and disadvantage of different antenna performance enhancement techniques

Performance enhancement techniques	Advantages	Disadvantages
Dielectric lens	It enhances the gain, improves front to back ratio, provides stable radiation pattern, and radiates the maximum energy in the front direction	It increases the size of an antenna
Multi-element	It improves the gain, efficiency, return loss, and bandwidth	It is difficult to design the feeding network and increases the size of an antenna
Corrugation	It provides improved gain, return loss, and bandwidth. Also, as it reduces side and back lobe levels, the front to back ratio increases	It reduces input impedance
Substrate choice	A substrate having low permittivity gives enhanced gain, efficiency, wide bandwidth, and a compact antenna while a substrate with high permittivity improves the return loss	A substrate having low permittivity is costly and they are not easily available
Mutual coupling reduction	It enhances the gain, efficiency, and input impedance matching. Some mutual reduction techniques reduce the size of an antenna	It increases the complexity of antenna design
Performance enhancement techniques	Advantages	Disadvantages
Dielectric lens	It enhances the gain, improves front to back ratio, provides stable radiation pattern, and radiates the maximum energy in the front direction	It increases the size of an antenna
Multi-element	It improves the gain, efficiency, return loss, and bandwidth	It is difficult to design the feeding network and

		increases the size of an antenna
Corrugation	It provides improved gain, return loss, and bandwidth. Also, as it reduces side and back lobe levels, the front to back ratio increases	It reduces input impedance
Substrate choice	A substrate having low permittivity gives enhanced gain, efficiency, wide bandwidth, and a compact antenna while a substrate with high permittivity improves the return loss	A substrate having low permittivity is costly and they are not easily available
Mutual coupling reduction	It enhances the gain, efficiency, and input impedance matching. Some mutual reduction techniques reduce the size of an antenna	It increases the complexity of antenna design

3.2.2 Mutual Coupling Reduction (Decoupling) Techniques

When it comes to improving the performance of multiple-input multiple-output (MIMO) antennas, certain decoupling techniques are absolutely necessary. The incorporation of these strategies into the design of MIMO antennas is not an option. You are welcome to look over Table 3.3, which provides a comprehensive breakdown of their benefits and drawbacks. The following can be considered the primary components of these techniques for reducing coupling:

- **(a) Neutralization Lines:** The use of neutralisation lines, which function by transmitting electromagnetic waves via a metallic slit or lumped device, can result in a reduction in the amount of mutual coupling that occurs between the components of an antenna. When paired with ground planes, it decreases the footprint of the antenna while simultaneously increasing its bandwidth. A shift in the effective bandwidth is caused by any change in the location of a point on the neutralisation lines. This is because the impedance of that point shifts whenever there is a change in its location.
- **(b) Decoupling Network:** The cross admittance of a decoupling network is transformed into a value that is entirely fictitious when discrete components or transmission lines are added to the network. This method makes use of a planar

decoupling network that is similar to a resonator in order to lessen the amount of reciprocal coupling which occurs. In order to facilitate multielement, dummy load, and linked resonator approaches, pattern diversity technology has been introduced into the decoupling network. When it comes to enhancing isolation, this strategy is successful and does not require a significant financial investment.

(c) **Electromagnetic:** Structure of the Bandgap (EBG) In its most fundamental sense, it is responsible for the transmission of electromagnetic waves. Components that are either dielectric or metallic are used in the construction of the EBG structure, which also features a periodic layout. It is able to generate several bandgaps as a result of its circular resonance characteristic that is independent of periodicity. Very minimal coupling occurs between the EBF structure and itself, making it an extremely efficient structure.

(d) **Dielectric Resonator:** The term "dielectric resonator antenna" (DRA) refers to an antenna that is referred to as such if it comprises a dielectric resonator. DRA is well-known for its high gain and low loss, both of which contribute to its great radiation efficiency. Additionally, DRA is capable of providing good isolation, in addition to showcasing dual-band capabilities.

(e) **Defected ground structure (DGS):** It is the job of the structure to consolidate any slots or faults that may be present on the ground plane of the antenna. There are many benefits associated with DGS, some of which include the potential to offer maximum efficiency, little mutual coupling, and a wide bandwidth.

(f) **Metamaterials:** Certain types of metamaterials include those that are single negative, electromagnetic, double negative, anisotropic, isotropic, terahertz, chiral, tuneable, photonic, frequency selective surface based, and nonlinear. Other types of metamaterials include those that are frequency selective surface based. The qualities that are shared by each of these metamaterials are distinct from one another. When creating metamaterials, it is necessary to use at least two different materials. The use of a metamaterial makes it possible to create an antenna that has a small size, a high gain, a wide bandwidth, and a low amount of mutual coupling.

(g) **Slot Elements:** It is possible to improve impedance bandwidth by using the coupling technique to either the ground plane or the radiation patch depending on the situation. The numerous applications of a slot antenna are a direct result of the several desirable properties that it possesses, including its high gain, efficiency, enormous bandwidth, and mutual coupling value.

(h) Complementary Split Ring Resonators (CSRR): Through the utilisation of the coupling approach on either the ground plane or the radiation patch, it is possible to acquire a more extensive impedance bandwidth. In order to accomplish this, the coupling strategy is utilised. Because of its many benefits, the slot antenna is helpful in a variety of settings. These benefits include its high gain, efficiency, and mutual coupling value, as well as its large capacity for bandwidth. These characteristics are responsible for the versatility of the slot antenna.

(i) Frequency Reconfigurable: Techniques of switching are the foundation of it. Varactor diodes, MEMS switches, and p-i-n are utilised by the reconfigurable antenna in order to improve the envelope correlation coefficient and expand the frequency range. Through the utilisation of a reconfigurable antenna design, it is possible to create an effective system that possesses minimal mutual coupling and high diversity gain.

Table 3.3 Advantages and disadvantage of mutual coupling reduction (decoupling) techniques

Mutual coupling reduction (decoupling) techniques	**Advantages**	**Disadvantages**
Defected ground structure	It is easy to implement, to enhance the bandwidth, to improve the front to back ratio, and to increase the efficiency	Its analysis is the challenging issue
Dielectric resonator antenna	It enhances efficiency, bandwidth, and gain	Its structure is complex
Complementary split ring Resonator	It improves diversity gain and reduces antenna size	It provides low bandwidth
Neutralization lines	It is a compact antenna. It gives wider bandwidth and enhanced efficiency	Its structure is complex
Slot or parasitic element	It enhances diversity gain, bandwidth, and efficiency	It is difficult to design and to decide the position of slot or parasitic element

Frequency reconfigurable	It provides compact size and supports multiple wireless standards. Also, it improves diversity gain, bandwidth, and efficiency	It required external components
Electromagnetic bandgap structure	It provides good front to back ratio and impedance matching	Its structure is complex
Metamaterial	It enhances the diversity gain, bandwidth, and ECC. Also, it is compatible for integration with another components	It is difficult to design and decide the position of managerial unit cells
Decoupling network	It improves diversity gain and impedance matching	It's gain is low and the design is complex

3.3 STRUCTURAL DESIGN AND BUILDING MATERIALS OF 5G ANTENNAS

SISO and MIMO are two broad classifications that can be used to an antenna, with the distinction being made based on the number of input and output ports. In the second place, MIMO antennas can be classified according to the design of their construction. These antennas are classified as multielement antennas if they do not have a metal rim, while those that do have a metal rim are classified as multiband and wideband antennas, respectively. It is essential to make a thoughtful selection of the building materials and structural design of an antenna in order to achieve the highest possible level of overall effectiveness attained by the antenna.

3.3.1 SISO Wideband Antennas

The SISO antennas that are used for 5G applications can be classified as either single-element or multielement, depending on the number of elements that they contain.

Single Element Antenna:

The design, construction, and fabrication of the single element antenna are all simple processes. It is demonstrated in Table 3.4 that a number of SISO 5G antennas have been constructed by utilizing multilayer architectures in conjunction with a single substrate layer. This was all done with the intention of reducing the size of the antenna.

In contrast to an antenna that consists of a single substrate layer, an antenna that has many layers has the potential to increase the gain of a relatively small antenna significantly.

Table 3.4 Comparison of single element antennas (SISO wideband)

Antenna type	Substrate	Size (mm3)	Number of substrate layers	Gain (dBi)	Frequency band (GHz)
Dipole	FR4	40 x 10 x 1	1	2-2.5	3.08-5.15
Antipodal Vivaldi antenna	FR4	40 x 24 x 1.6	1	5-9.53	25-33.4
Circular slot	Nelco NY9220	20 x 16 x 0.508	1	8-9	20-28
Microstrip patch	R04003C, Taconic TLX-9	90 x 96 x 2.878	2	8.59-10.43	3.24-3.8
ME dipole	Arlon 25N	40 x 40 x 10.516	2	6-8	4.98-6.31
Dielectric resonator antenna	Teflon, ceramic, Rogers 5880	75 x 75 x 15.428	3	6-9.2	3.1-5.1
Microstrip patch	RT/Duroid 5880	12 x 12 x 1.02	2	9.5-11	24-34.1

An example of a multi-layered antenna design with circular polarization is presented in Figure 3.4. Figure 3.4a depicts an antenna that is composed of three layers of copper and two substrates made of RT/Duroid 5880 substrates. Both of these components are supposed to be present in the antenna. The meta surface layer is the one that is located at the very top of these three layers. Four square circles measuring four inches by four inches make up its pattern. The radiator patch is located in the epicenter of the copper layer, and the ground is located at the bottom of the layer. The resonance frequency of the antenna is influenced by the square ring of the meta surface, which is a combination of an inductor and a capacitor that is connected in series. Due to the presence of this mixture, the bandwidth of the antenna is impacted. It can be shown in Figure 3.4b that the bandwidth of the antenna, when there is no meta surface present, falls somewhere between 24.6 and 28.7 GHz. The bandwidth of antennas that have a meta surface, on the other hand, can range anywhere from 24 to 34.1 GHz. It is further demonstrated in

Figure 3.4c that the meta surface enhances both the gain (9.5-11 dBic) and the axial ratio bandwidth (24.1-29.5 GHz). Through the utilization of realistic techniques of multilayer and meta surface augmentation, the gain of a miniature wideband antenna can be enhanced.

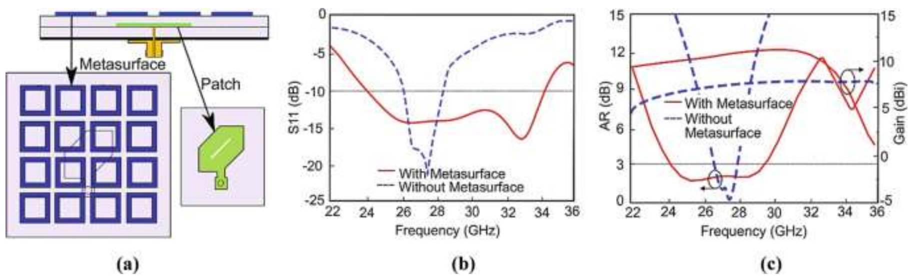


Fig. 3.4 Multilayer antenna design

Source: Adapted from Hussain et al. 2020 with open access (IEEE)): (a) Antenna design; (b) Return loss; (c) Gain and axial ratio

Multielement Antennas:

A wide frequency spectrum, a steady radiation pattern, and a high gain are the three technical characteristics that are considered to be the most critical for 5G antennas. Due to the fact that a single element antenna is unable to satisfy these requirements, it is vital to have a variety of multielement antennas that are created specifically for 5G applications. In this article, we will examine many instances of multielement antennas that are suitable for 5G wideband:

Table 3.5 Comparison of multi-element antennas (SISO wideband)

Antenna type	Substrate	Size (mm3)	Number of substrate layer	Gain (dBi)	Frequency band (GHz)
Antipodal Vivaldi antenna	RT/Duroid 5880	28.8 x 24 x 0.254	1	8.2—13.2	24.04-40.85
Antipodal Vivaldi antenna	RT/Duroid 5881	37.6 x 14.3 x 0.254	1	8.5—10.7	23.41-33.92
Fractal	RT/Duroid 5882	32 x 12 x 0.254	1	7.8—10.9	25.28-29.04

Dipole	K'T/ Duroid 5883	30 x 35.62 x 4.9	4	10.6—12.61	27.12-29.5
Microstrip patch	Taconic TLY-5	96.1 x 50.5 x 1.016	2	13.83—14.31	26.4-28.92
Microstrip patch	RT/Duroid 5880, Acrylic Polymer	32.1 x 37.45 x 2.124	2	10-12	23-32

It is indicated that this table is a 3.5. The size, gain, and bandwidth of single substrate layer antennas are typically considered to be similar to those of other types of antennas. It is necessary to have an exceedingly big antenna in order to obtain the performance increases that are accomplished by a multielement antenna. This is the reason why multilayer antennas are brought into existence. The data presented in Table 3.5 demonstrates that the utilization of corrugations to offer a wide bandwidth results in the formation of a rather compact 1 × 4 multielement AVA.

Both the return loss and the gain that occurred as a consequence of the corrugation process are depicted in Figure 3.5 which can be found here. An illustration of a multielement AVA design with dimensions of 1 × 4 can be shown in Figure 3.5a. In this design, the fat margins of the AVA components are provided with corrugation.

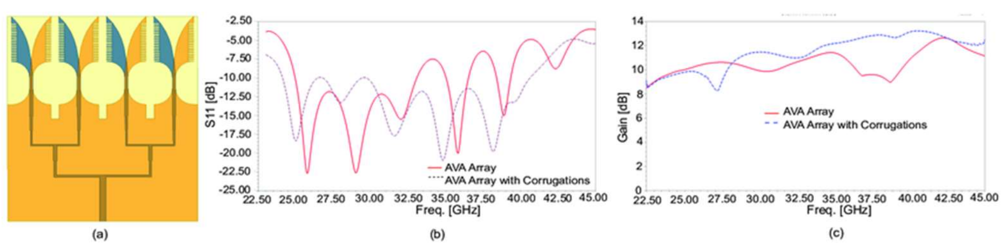


Fig. 3.5 Antenna design with corrugations

Source: Adapted from Dixit and Kumar 2020b with permission from IEEE): (a) Antenna design; (b) Simulated return loss; (c) Simulated gain

Following the incorporation of these corrugations, an inductor (L), a resistor (R), and a capacitor (C) are added to the fat edges. This results in an increase in the length of the electric route that the current takes. Following the addition of corrugation to an antenna, the resonance frequency of the antenna undergoes a change, as demonstrated in Figure 3.5b. Therefore, it is evident that the presence of corrugation in the antenna results in an improvement in the bandwidth of the antenna. When there is a rise in the current density at the edges of the antenna, the performance of the antenna is enhanced. This is because more energy is emitted in the direction of the end-front. Figure 3.5c illustrates the advantage that was achieved in this situation. The strategy of boosting corrugation performance is highly effective in terms of achieving the goals of increasing bandwidth and gain.

3.3.2 SISO Multiband Antenna

SISO multiband antennas can also be classified as single-element or multi-element antennas, which are other possible classifications. The bandwidth and gain that they offer are typically lower than those of other options. Dual band antennas are capable of operating at frequencies of 28 GHz and 38 GHz, despite the fact that their bandwidth is somewhat limited, measuring 3.65 GHz and 2.19 GHz, respectively. Additionally, the gain is reduced as a result of this design, despite the fact that the slot approach was utilized in the construction of the triband antenna. This indicates that SISO multiband is not a suitable solution for use cases using 5G.

3.3.3 MIMO Wideband Antennas

It is possible to categories wideband multiple-input multiple-output (MIMO) antennas as either multielement antennas with or without a metal rim or as multielement antennas.

3.3.3.1 Multielement without metal rim antennas

There are two primary options for MIMO wideband antenna designs that do not include metal rims: multielement antennas and dual element antennas.

3.3.3.2 Dual element antenna without metal rim

As a result of its high gain, enhanced isolation, and high efficiency, dielectric resonator antennas (DRAs) are utilized in the majority of multiple-input multiple-output (MIMO) antennas. One of the most important requirements of the MIMO system is to make certain that the antenna elements are protected from one another to the greatest extent possible. For instance, hybrid feeding mechanisms, frequency selective surfaces (FSSs), and meta surface shields have been utilized in order to accomplish the objective of enhancing the level of isolation between MIMO DRAs at a higher level. The utilization of these strategies is done in order to offset the current displacement that occurs in the components of the antenna structure. Attaching a metal strip to the top surface of the dielectric resonator is yet another method that can be utilized to enhance the isolation of the device. It is possible to transfer a strong coupling field away from adjacent slots with the assistance of an additional metal strip. This will result in a lower ECC value and a higher diversification advantage.

Because of the limited area that is available on the mobile device, isolation in a MIMO system is a very difficult task to do. The neutralization line and decoupling procedures are two methods that are frequently suggested for the purpose of significantly improving isolation. There is also the possibility of achieving high isolation through the utilization of two antennas that are designed to be self-decoupled. It is possible to make this structure by mounting two antenna elements on the same surface in a specific orientation. This results in an increase in both the effective length of the antenna as well as the isolation between its elements. Considering that this self-decoupled antenna architecture has an ECC value that is lower than 0.1, it is predicted to have an efficiency of approximately 58%. There is a possibility that the self-decoupled structure will result in enhanced isolation, less ECC, high efficiency, and a more compact dimension.

An examination of the various MIMO wideband antennas that are prepared for 5G networks may be found in Table 3.6. When evaluating these antennas, some of the performance characteristics that are taken into consideration include size, gain, frequency band, isolation, efficiency, and the type of antenna. The results of this experiment demonstrate that PIFA and ME dipole antennas have the ability to enhance the performance of the MIMO wideband antenna, while DRA and ME dipole antennas have the capacity to boost gain. The self-decoupling antenna will also have a larger footprint than the others, which is another advantage it will have. It has been discovered that antennas constructed employing DRA in the millimeter wave band have a greater gain, a smaller size, a lower ECC value, perfect isolation, and a better radiation efficiency.

Table 3.6 Comparison of dual element without metal rim antennas

Antenna type	Size (mm1)	Gain (dBi)	Isolation (dB)	Frequency range (GH z)	Efficiency (%)
DRA	20 x 20 x 2.54	9.9	24	27.25-2839	-
Monopole	150 x 75 x 0.8	-	17	3.4-3.6	58

PIFA	50 x 100 x 3.00	3	25	2.7-3.6	80-92
ME dipole	60 x 60 x 8	8.2	25	3.3-4.36	89.5

A representation of the DRA structure, the S parameter, gain, and ECC for MIMO DRA may be found in Figure 3.6. As can be seen in Figure 3.6a, a Rogers 5880 substrate is utilized in order to mount two DRAs that are in the shape of rectangles. It is necessary to print a metal strip on top of the DRAs in order to accomplish the goal of improving isolation. As seen by Figure 3.6b, which illustrates the S parameter, the addition of the metal strip significantly improves the sensation of being isolated. As can be seen in Figure 3.6c, we were successful in achieving a high level of diversity gain along with channel capacity. There is a diversity gain in the 28 GHz band that is more than 9.9 dB, and the ECC value is lower than 0.013.

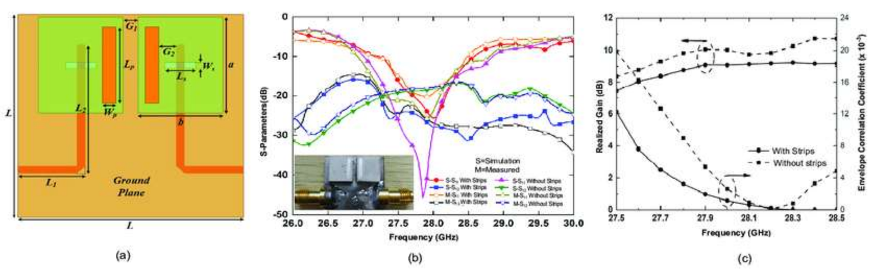


Fig. 3.6 Dielectric resonator antenna design

Source: Adapted from Zhang et al. 2019 with permission from IEEE): (a) Top view of MIMO DRA; (b) S parameters; (c) Gain and ECC

3.3.3.3 Multielement antenna without metal rim

A multi-input multiple-output (MIMO) multielement antenna is utilized in order to enhance the intensity of the signal as well as the transmission range; this is an essential component for any application that involves high-speed communication requirements. Structured monopoles, slot antennas, dual- and tri-polarized antennas, and hybrid antennas are the components that make up these antennas, which are capable of supporting 5G applications.

The benefits of MIMO can be obtained in a number of ways, one of which is by lowering the degree of reciprocal coupling. A number of different approaches have been taken in order to achieve the objective of reducing the level of mutual coupling.

There are many different approaches that fall under this category. Some examples include DGS, neutralization lines, orthogonal polarization, and polarization diversity. Because of the MIMO antenna's use of polarization diversity and its uniplanar design, antennas for 5G networks are more resistant to interference and fading. This is because the MIMO antenna reduces the amount of mutual coupling that has occurred between the antennas.

Table 3.7 presents a comparison of the MIMO wideband multielement antenna to various other wideband antenna applications. For the purpose of this comparison, the following factors are taken into consideration: the frequency range that was utilized, the sort of mutual coupling reduction approach, isolation, ECC, and channel capacity. Using the polarization strategy is often the method that monopole antennas employ in order to achieve high levels of isolation and low ECC values. Utilizing orthogonal polarization is one method that may be utilized to enhance the channel capacity of the SIW antenna. In addition, the utilization of polarization diversity and an F-shaped stub can be of assistance in achieving improved isolation. With regard to these antenna designs, the one that incorporates a microstrip patch offers the highest level of isolation, the broadest bandwidth, the most consistent emission pattern, and the most powerful gain.

Table 3.7 Comparison of multi-element without metal rim (MIMO wideband)

Antenna type	**Frequency range (GHz)**	**Isolation (dB)**	**ECC**	**Channel capacity (bps/Hz)**
Monopole	33-4.2	15	0.1	163
Fatten=	23-4.2	15	<03	-
Monopole	3.4-3.6	10	0.2	35-38
Monopole	5.1-5.9	17	0.01	-
Monopole	235-2.65	123	0.15	38-40
SIW antenna	3.4-3.6	123	0.2	57
Slot	3.4-3.6	173	0.05	40.8
Inverted F	23-7.0	17	0.1	39
Microstrip patch	24.35-31.13	20	-	-

Furthermore, the performance metrics as well as the multielement antenna that was constructed with sixteen elements are displayed in Figure 3.7. The formation of this

16-element antenna involved the use of two multi-element antennas, each of which was constructed using two substrates of the same size. As can be seen in Figure 3.7a, the patch most at the bottom of the stack functions as a radiator, while the patches most at the top of the stack increase the bandwidth as parasitic patches. We were able to rotate the antenna 180 degrees out of phase by making use of feed lines and radiating patches. This resulted in a significant reduction in the amount of mutual coupling that takes place between the elements of the antenna. The frequency plot and return loss S11 are depicted beside one another in Figure 3.7b, both with and without a feeding network.

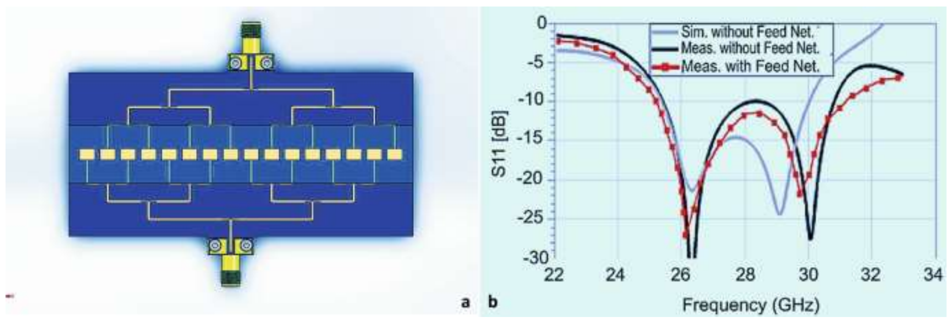


Fig. 3.7 Fabricated multi-element antenna with performance parameters

Source: Adapted from Khalily et al. 2018 with permission from IEEE): (a) Fabricated MSA with parasitic element; (b) Simulated S11 with and without feeding network

The E plane and the H plane gains are those that are produced by this feeding network, which operates at a phase difference that is finite. The impedance bandwidth of this antenna is 5.37 GHz, and it has a gain of 19.88 dBi. This antenna offers remarkable performance. Due to this particular reason, this antenna is suited for use in 5G applications.

Multielement Antenna with Metal Rim

Multiple-input multiple-output (MIMO) antenna systems that are built on metal rims are the foundation of antenna designs for smartphones that are ready for 5G. Through the utilization of a substantial quantity of loop and slot antennas, it is possible to integrate all of the necessary LTE and 5G bands. Metal rim MIMO antennas can either be fixed or reconfigurable, depending on the application that they are being used for. Wideband and multiband applications are both capable of making use of them. The construction blueprints for the metal case of the smartphone are depicted in Figure 3.8.

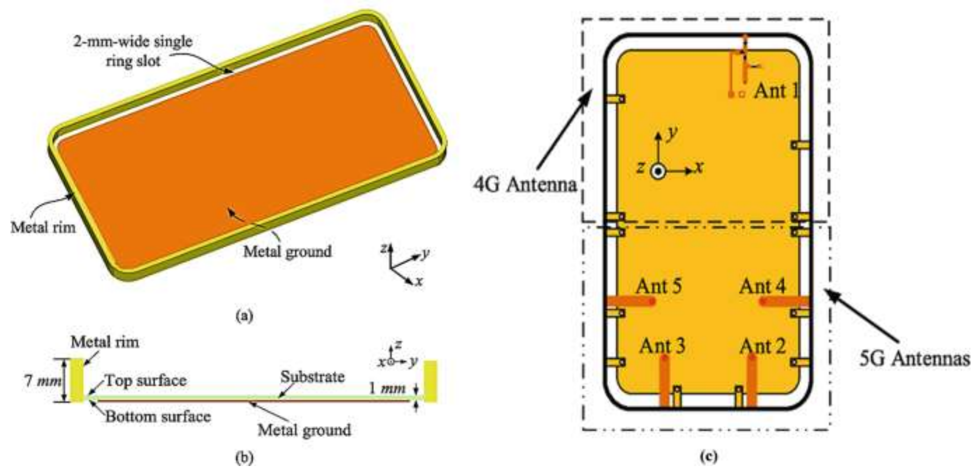


Fig. 3.8 The smartphone metal shell and slot antennas

Source: Adapted from Chen et al. 2019 with permission from IEEE): (a) 3D view; (b) Side view; and (c) Assembly of single ring slot-based antennas on metal-rimmed 4G/5G smartphones

The following three pictures illustrate how antennas based on single ring slots are installed in 4G and 5G devices that have metal rims: In Figure 3.8a, the metal case of the smartphone is depicted from a three-dimensional viewpoint; in Figure 3.8b, the equivalent side view is shown; and in Figure 3.8c, the installation is shown in motion. A substrate, a metal rim, and a single ring slot of two millimeters in width are the components that make up the metal shell. A metal foundation that is located beneath the substrate is completely encircled by a metal rim that is completely and flawlessly intact. When designing antennas, it is usual practice to run them along the perimeter of a metal base.

In addition, the two primary types of MIMO metal rim antennas are the reconfigurable antenna and the fixed antenna. One of the most important characteristics of a reconfigurable antenna is its capacity to change its frequency, which characterizes the antenna as a reconfigurable antenna. Antennas that are permanently mounted and do not permit frequency changes are in contrast to this type of antenna. When compared to fixed-type metal rim antennas, reconfigurable (tunable) antennas have the potential to expand the capabilities of the system in communications and radar applications without increasing the size of the system or the expense of the system. In addition to this, the reconfigurable antenna is capable of operating at a wide range of frequencies and offers a wide variety of polarization adjustments. There are many different materials that can be used to construct tunable antennas. Some of these materials include pin diodes, varactor diodes, microelectromechanical systems (switches, phase shifters), and tunable materials like liquid crystals.

CHAPTER 4

HIGH FREQUENCY FILTER MATERIALS

4.1 INTRODUCTION

The importance of high-frequency filters in wireless goods is expanding, particularly because these items are competing for a smaller and smaller share of the frequency spectrum that is available. Assuring that wireless radio transmitters and receivers receive and process the appropriate signals requires the utilization of a variety of RF/microwave filters. These filters not only shield the devices from signals that fall within their frequency range, but they also prevent signals from reaching frequencies that are either too high or too low. Filters will be essential for next-generation wireless systems that make use of millimeter-wave (mmWave) frequency bands for high-speed communications. This is because filters will eliminate interference to the greatest extent possible. Fifth-generation (5G) wireless technology, in contrast to its predecessors, aims to decrease the amount of energy that is consumed, enhance capacity, minimize latency, and increase the amount of signal traffic density.

In order for networks to accomplish these objectives, it will be required for them to expand their broadband capabilities through carrier aggregation and expand their capabilities into the millimeter-wave spectrum. In order to improve their spatial efficiency, they will require additional tools such as base station density, beam-forming antenna arrays, and massive multiple-input multiple-out (MIMO) antenna technology. The intrinsic radio frequency (RF) front-end components will be confronted with new hurdles as a consequence of the capabilities offered by these technologies. To be more specific, it will be extremely challenging to accommodate a diversified network of mobile devices and base stations, each of which has its own distinct cell size, without resorting to a large number of different filter designs.

It is necessary for systems to be able to deal with both in-band and out-of-band interference in order to accomplish the goal of achieving enormous overlapping of microcells for ubiquitous coverage and wide bandwidths through carrier aggregation capabilities. Similar to the situation that occurred earlier, the implementation of huge MIMO will need the utilization of compact filtering technology. This technique has the capability of lowering the uplink sum rate of maximum ratio-combining (MRC) receivers in situations where there is interference from beyond the operating band. It is imperative that research be conducted into the challenges that are brought about by 5G filters in order to successfully accept these new technologies and move into frequency spectra that have been relatively untapped. In addition to the variables that are driving the physical, electrical, and economic limits for 5G filters, it is necessary to explore the associated simulation technology that will assist designers in physically realizing these components.

4.1.1 Current Status of Mobile Device Filter Technologies

More than sixty filters, the most majority of which are multiplexers, are utilized by the current generation of 4G Long Term Evolution (LTE) handsets in order to accommodate an additional thirty frequency bands. Because of the significant amount of space that these filters take up and the fact that they account for the majority of the mobile ecosystem's RF expenditure budget, manufacturers of these components are under severe pricing pressure to achieve performance criteria while keeping costs low.

This is because of the fact that they are responsible for the majority of the RF expense budget. The technology known as surface acoustic wave (SAW) or bulk acoustic wave (BAW) is the foundation for almost all of these filtering components. Solidly mounted resonators (SMRs) and film bulk acoustic resonators (FBARs) are the two types of resonators that are typically used for making BAW filters. There is also the option of purchasing temperature compensated (TC) versions of both the SAW and BAW components. TC is used as a descriptive prefix in specific situations, such as in the instance of TC-SAW.

In a nutshell, the operation of electrical filters involves the storage of a signal for a considerable amount of time in order to ascertain the rate of change involved in the signal. Because of this, it is possible to differentiate between the various frequencies. Due to the fact that the storage is digital memory, a digital signal processing (DSP) filter is the filter that is the most conceptually easy. Digital signal processors (DSPs) are responsible for carrying out the procedure. These DSPs make use of an A/D converter in order to convert the signal into a digital sequence before storing it in the local semiconductor memory. The next step is for the data to be processed by the math processor, which will remove frequencies that are not requested and allow only the frequencies that are desired to be there. The same objective can be accomplished by using analogue filters, which do not store the signal in digital words but rather in the form of stored energy. As opposed to traditional RLC filters, which store energy in capacitor charge and current, BAW/SAW filters store the signal in acoustic resonators.

This is in contrast to the current and charge that are stored in capacitors. Within the transmission lines or the cavity, electromagnetic (EM) resonance is utilized in order to store the signal. This is accomplished through the utilization of waveguides and cavity filters. When it comes to filters that are based on printed circuit boards (PCBs), it is feasible to combine a number of different responses. Among the answers that are supplied are bandpass filters, lowpass filters, high pass filters, and band stop filters.

A number of different transfer functions are responsible for defining the transition zones of fluorescent filters. For instance, the sudden transition from the passband to the stopband is what distinguishes a Chebyshev filter from other frequency filters. An incredibly minute quantity of spectrum is required in order to go from a condition of maximum signal attenuation to a state of least signal loss. On filters that use a Butterworth or binomial function, the transition from the passband to the stopband is more gradual than on filters that use a different function. It is possible to obtain a passband with minimal loss and very little ripple when compared to a Chebyshev filter, which has transitions that are more rapid. However, in order to transition from filter areas, it requires a wider frequency spectrum.

Despite the fact that it is true that the frequency response of a filter is an average of its responses over all of its spectral areas, the transfer function has a significant impact on the loss characteristics of the filter's passband and stopband. The passband insertion-loss response may exhibit ripples or changes in amplitude as a result of this seamless transition, despite the fact that a Chebyshev filter is capable of making a transition from a passband to a stopband in a quick and smooth manner. Butterworth filters have a passband insertion-loss response that is far more robust than that of Chebyshev filters. In addition, Butterworth filters have the ability to reduce the attenuation of signals that occur at frequencies that are closer to the passband.

4.1.2 The 5G Filter Performance Challenges

Currently, the most prevalent off-chip techniques that are utilized in mobile devices are surface acoustic wave (SAW) filters and bulk acoustic wave (BAW) filters. These filters are an excellent option because they are both compact and capable of handling frequencies up to 6 GHz. In addition, they are the most often used off-chip technologies due to the favorable performance-to-cost ratios that they offer. Unfortunately, there exist filtering techniques that are equivalent for the millimeter-wave spectrum; but these solutions have problems with size, performance, practicability, and availability, among other spectrum features. Over the course of the past few decades, each successive generation of mobile systems has triumphed over its own distinct set of obstacles, beginning with 1G analogue systems and progressing to 2G digital standards, 3G mobile broadband capabilities, and now 4G LTE and LTE-Advanced networks. The implementation of performance increases such as these is necessary in order for technology advancements such as these to become more tangible. In order to address the issue of rising data rates, the Shannon-Hartley theorem provides an explanation of the maximum amount of error-free digital data that can be conveyed across a given bandwidth channel when noise is present.

$$C = M \times B \times \log_2\left(1 + \frac{S}{N}\right) \quad \text{.....4.1}$$

For instance, the MIMO order can be represented by the following: C represents the channel capacity, M is the number of channels, B represents the bandwidth, and SNR represents the signal-to-noise ratio. More specifically, the amount of bandwidth that is accessible in a certain spectrum region is the primary focus of 5G, which is primarily concerned with capacity. Increasing data rates, which are a subset of channel capacity, is the goal of this endeavor.

The 5G Filter Requirements:

In light of the fact that interference mitigation is becoming increasingly vital as a result of spectrum crowding and the requirement to limit or eliminate guard bands as a result of bandwidth utilization, high-performance filtering is absolutely necessary. These two factors contribute to the necessity of reducing interference in the system. Without narrowband filters that have exceptionally steep filter skirts (high selectivity), robust rejection, and negligible temperature drift, the approach cannot function. These filters also have a very low loss. Not only are these criteria quite stringent, but there is also the chance that performance will deteriorate at mmWave frequencies as a result of parasitic growth and substrate losses in the filtering device and its packaging (laminate). It is necessary to implement filter performance measurements that have been specifically designed in order to address the following issues:

1. **Percent Bandwidth:** It is therefore not appropriate for filter technology to restrict the bandwidth of the radio access system. According to the idea, 5G systems should be able to employ a set of radio solutions that are comparable to the mobile high band frequency channel that is already in use, provided that the frequencies of the 5G bands, which range from 3.5 to 6.0 GHz, are sufficiently close to one another. In spite of the fact that the higher frequencies of 6 GHz will put the performance levels of components that are currently available off the shelf to the test, it is anticipated that the fundamental radio designs that are utilised in the wireless communications systems that are currently in use will transition successfully to the higher frequencies of 5G. In light of the fact that SAW filters are already having difficulty functioning in the 2.5 GHz band, the continuously increasing frequency would only make the situation even more difficult for their application. This is the situation with regard to the filters that are under consideration. The field is now free to accommodate BAW and TC-BAW components as a result of this. Due to the performance reduction that the current BAW filter technology faces at higher frequencies, it is unfortunate that these filters might not be suited for use.
2. **Selectivity:** In order to make the most of the available bandwidth, designers need a selectivity that is high.
3. **Insertion Loss:** When designers have a high selectivity, they are able to make greater use of the bandwidth that is available to them.
4. **Size and Packaging:** A number of factors, including issues regarding size and integration, will play a role in the development of filter technology for 5G applications. Concerns like these will be influenced by the way the system is designed. In order to prevent the formation of grating lobes, it is necessary for the antenna elements that make up a phased array to be in close proximity to one another. A half-wavelength distance is significantly less than a few millimetres when millimetre wave frequencies are being used. The plank design is utilised by phased arrays, which are widespread components found in a variety of millimetre wave systems. In Figure 4.1a and 4.1b, the green portions represent the printed circuit boards (PCBs) on which the antennas are located. On the other hand, the blue areas show the circuit "planks" that extend 90 degrees from the array, which are locations where filter space is already restricted. On the other hand, fat-panel topologies are taken into consideration during the design process of base stations. Given that the circuitry is built on the rear side of the antenna printed circuit board (PCB), the design demands for

a significantly more compact spacing between the filtering blocks and the other functional blocks.

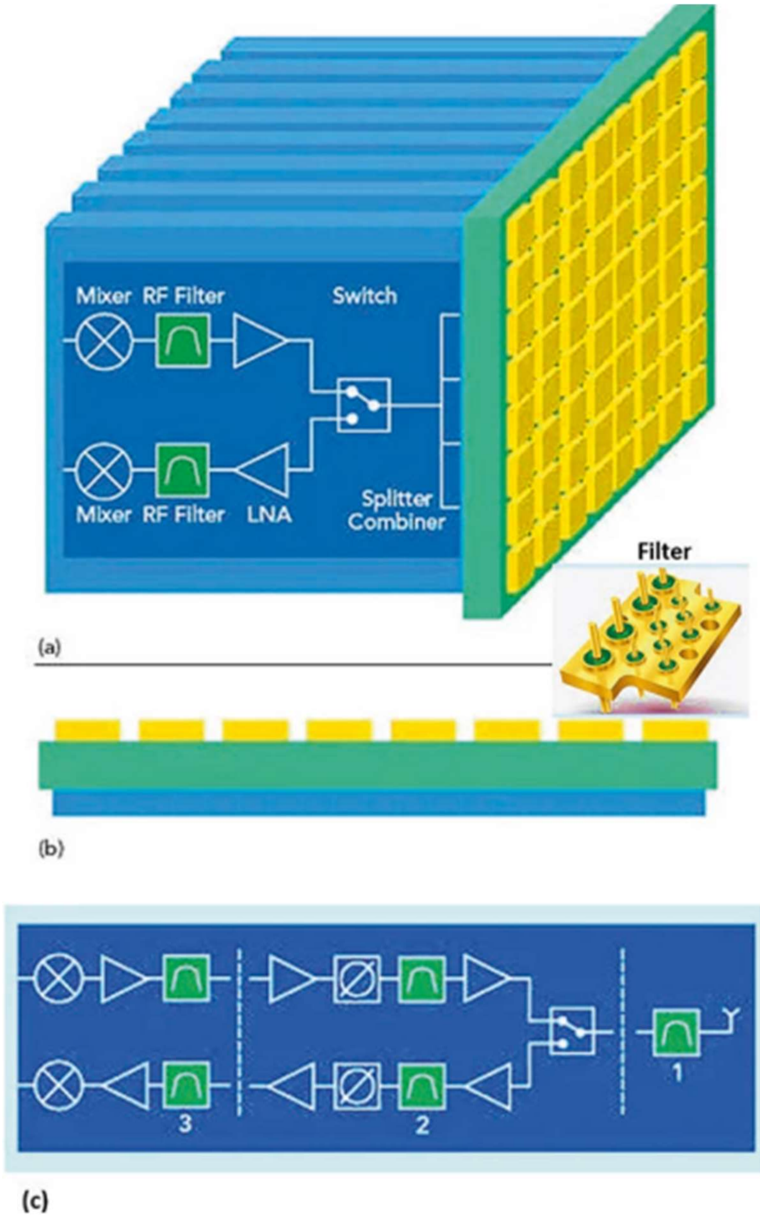


Fig. 4.1 Phased array architectures

Source: Adapted from Matthews (2019), Credit: Microwave Journal): (a) plank; (b) fat panel; and (c) potential radio filter locations

Cognitive and reconfigurable software-defined radio (SDR) architectures, as well as tuneable filters, are two separate approaches to radio design that are garnering a lot of attention. In SDR, all filtering takes place after the analogue-to-digital conversion on the receive side and before the digital-to-analogue conversion on the broadcast side. It is done in this manner in order to improve the effectiveness of the signal. However, despite the fact that this active filter technology consumes a significant amount of power (tens of watts), the current state of the art in silicon semiconductor integration is capable of meeting the filtering requirements. Passive filters, on the other hand, do not become active when there is no power applying to them. The analogue-to-digital converter (ADC) is responsible for converting the entire spectrum that is received. Any potentially powerful out-of-band signals have the potential to have an effect on the front-end amplifier in a single-division receiver (SDR). An adjustable band-select filter may be used to handle out-of-band signals before they are sent to the low-noise amplifier (LNA). Additionally, an anti-aliasing filter that could be modified and placed before the analogue-to-digital converter (ADC) would significantly improve power efficiency.

5. **Temperature Stability:** In order to make effective use of the bandwidth that is available, the filter needs to be able to meet its specifications across a wide temperature range. In environments that are exposed to extremes of temperature and temperature variation, it is possible to deploy systems on a smaller scale. Furthermore, the overall size reduction in systems results in densely populated boards, which can cause the stability of the filter to be affected by heat from components that are located in the surrounding area.

4.1.2.3 Physical Design and Emerging Solutions for the 5G Filters

Off-chip filtering applications in mobile devices have been dominated by SAW and BAW filters for a considerable amount of time. This is due to the fact that these filters have great performance specifications, tiny footprints, and inexpensive pricing in comparison to other possibilities. Unfortunately, due to the nature of their architecture, which involves the utilization of interdigital transducers (IDTs) to process signals inside the form of acoustic waves, the selectivity of SAW and BAW technology decreases at frequencies that are higher than 6 GHz. This renders these technologies unsuitable for use in mm Wave applications. Film bulk acoustic resonator filters, also known as FBAR filters, are a type of BAW filter that can function at frequencies ranging from 5 to 20 GHz while being below the acceptable millimeter Wave ranges.

As a result of their high Q factor, these filters provide minimal insertion loss, which contributes to reasonably good system performance. Furthermore, they are able to be integrated with monolithic microwave integrated circuits (MMIC) and other technologies, which helps to minimize cost, size, and power consumption. As the frequencies climb towards the mmWave range, the signal wavelengths become sufficiently narrow to make it possible to construct filters that are based on electromagnetic principles. There are two types of high-performance filters that are most commonly used between 20 and 80 GHz. These include waveguide filters and cavity filters.

However, the dimensions of these sorts of filters are measured in centimeters' rather than millimeters. Many attempts have been made to reduce the size of these filters so that they can operate at mmWave frequencies. However, the wavelength size of the electromagnetic wave that is being filtered is still quite large in comparison to the physical size requirements of the filter. As a result, it is highly probable that these millimeter-wave filters will be larger than lower-band acoustic filters. This may be acceptable if a different radio architecture can reduce the number of filters that are required. If this is not the case, then an alternate construction method for electromagnetic-based filtering for millimeter-wave radio systems needs to be created.

Figure 4.2a provides a summary of the frequency ranges of filters that are now available to the general public. Acoustic filtering is not as practicable at higher mmWave frequencies, which is why many developers have moved to electromagnetic (EM) solutions. These options include dielectric and cavity waveguide, on-chip, and microstrip (or planar thin film) filters. Within metal or dielectric tubes, waveguide filters are cavities that are either hollow or filled with dielectric. These cavities function as BPFs, preventing certain wavelengths from passing through while letting others to pass through. The military, radar, satellite, and broadcasting industries all make extensive use of waveguide filters, which are characterized by excellent power handling and low loss. Waveguide filters are utilized at mmWave frequencies ranging from 20 to 80 GHz. Waveguides, however, often have dimensions that fall somewhere in the centimeter region (see Fig. 4.2b of the figure). 4.35 millimeters is the $\lambda/2$ array element spacing while operating at 28 gigahertz in free space. It is possible that this technique is not feasible for an array antenna system until manufacturers are able to reduce waveguide widths in a sufficient manner while still meeting the requirements for electrical performance.

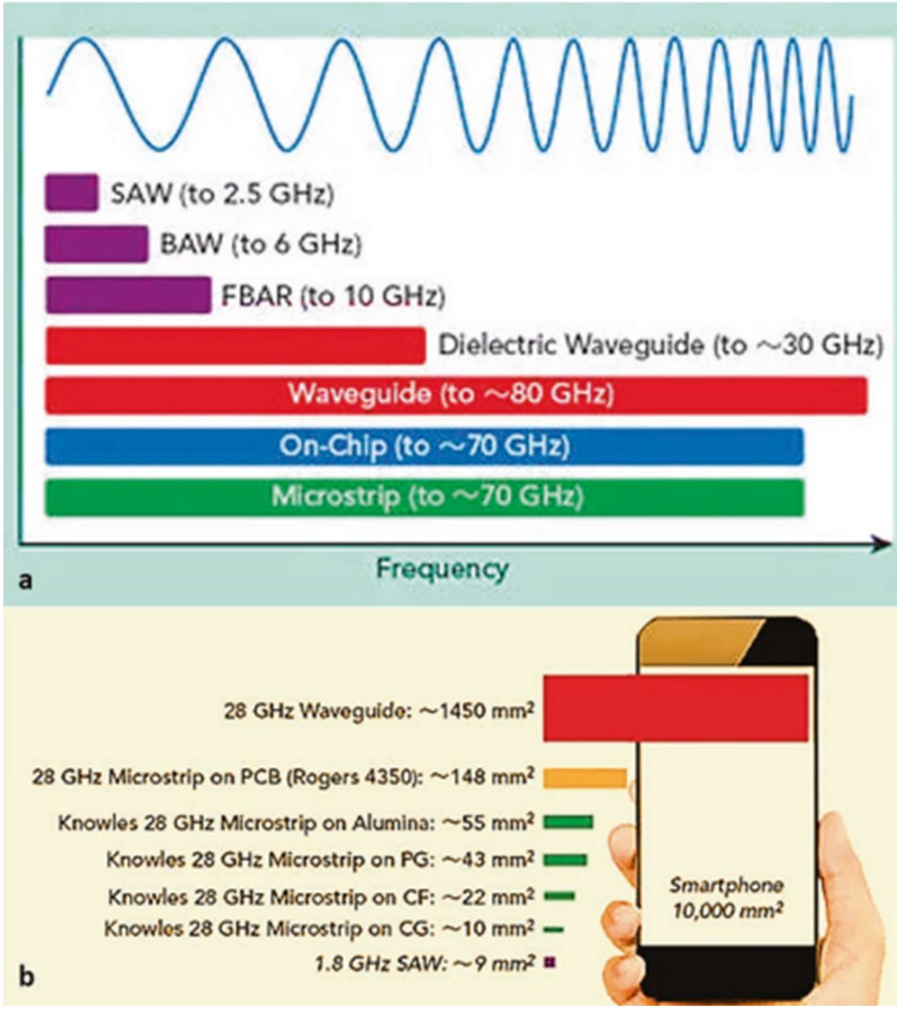


Fig. 4.2 Common standard filter technologies

Source: Adapted from Adapted from Matthews (2019), Credit: Microwave Journal): (a) Frequency ranges; (b) Sizes of mm Wave filters compared to typical SAW filters

The capacity to interface with other devices in order to construct systems on chip (SoC) solutions is one of the appealing characteristics of on-chip filters that are dependent on semiconductor technology. Other desirable characteristics include the ability to prevent short circuits and tight tolerances. On the other hand, due to the tiny dimensions of on-chip devices that are created for millimeter Wave frequencies, there will be significant performance concerns with the Q-factor, losses, and noise factor (NF). There are a number of reasons that contribute to the difficulty, some of which include the physical qualities of the semiconductor material and the cost of implementation. For example, GaN circuits strive to achieve maximum heat dissipation by designing themselves to be as thin as is practically possible. However, because the filter Q is tied to the thickness of the dielectric substrate, the high-power benefit of employing GaN devices is in direct opposition to integrating filters with a high Q. Both of these advantages are in direct antagonism to one another. Another problem is that active devices could be able to make better use of the space on the wafer that is occupied by a GaN filter configuration. At this time, it is not possible to construct high-Q filter structures on a chip for use in front-end applications.

When it comes to applications using millimeter waves, microstrip filters have been taken into consideration; nevertheless, for a variety of performance reasons, they are typically neglected. Additionally, there are at least three distinct form factors from which you can choose at the time.

1. Microstrip on PCB
2. The microstrip is a multilayer, low-temperature cofired ceramic (LTCC) package
3. Microstrip is a small form factor, single-layer package.

The fact that printed microstrip filters on PCB are so simple to manufacture is one of the components that makes them so appealing. On the other hand, millimeter-wave antennas need subwavelength dimensions, which are much beyond the capabilities of high-performance printed circuit board (PCB) solutions that are centimeter-sized. Due to changes in the manufacturing process, the performance of the printed circuit board (PCB) is further restricted. This is because the insertion loss is increased and the suppression levels are decreased.

The surface mount technology, which is sometimes referred to as SMT, is still another alternative. While surface-mount technology (SMT) assembly has been utilized for quite some time in commercial systems, it is currently being utilized in millimeter Wave military technology in order to reduce costs. Standardized form factors of SMT filters lessen the amount of time required for assembly and eliminate the need for post-tuning, in contrast to alternatives such as chip and wire. In a manner analogous to that of multilayer capacitors, the SMT layout is utilized by filters that are manufactured of low-temperature co-fired ceramic (LTCC). When building multilayer capacitors, many layers of extremely thin ceramic tape that have been printed with a variety of passive components are stacked on top of one another. The substrate is protected from bending as a result of this technique. At the moment, we are engaged in the process of developing prototypes of LTCC technology for millimeter Wave applications. With the use of this technique, it might be able to incorporate antennas and filters onto a single component, which would result in an extremely compact package. Considering that the metal coatings are screen printed, it is unfortunate that the dimensional precision is worse than it is with other thin film choices. A further issue is that the substrate is not polished to a suitable degree, which may result in considerable losses. As an additional point of interest, the level of suppression is often maintained at 30 decibels or lower.

Within the category of surface-mount filter assemblies, there is also a single-layer microstrip category. To produce resonant structures that have excellent performance, it is printed with transmission lines that are distributed throughout. This particular assembly type is employed in the process of constructing resonant structures. The production of low loss, high rejection filters that are stable from -55 to 125 degrees Celsius can be accomplished by painstaking research into the topology of the filter and the dielectric materials themselves. Please refer to Table 4.1 for a comparison of the performance of these filters to that of their SAW and BAW equivalents that operate at lower frequencies. Despite the limits of 5G New Radio (NR) systems, it is still possible for these systems to accommodate devices that have high performance and tiny form factors. This might be demonstrated by a low-loss filter operating at a frequency of 26 GHz, which has the capability to be contained inside a footprint measuring 4 × 1.6 mm and has a suppression level that exceeds 50 dB. Due to the significant size difference that occurs with half a wavelength, it is possible to include it into both planar and plank designs.

Table 4.1 Comparing SAW/BAW filter performance with SMT microstrip

Performance goals	**Typical SAW/BAW performance (<6 GHz)**	**SMT microstrip performance (<6 GHz)**
Low insertion loss	<3 dB	<3 dB
Excellent rejection	>30 dB out of band	>50 dB out of band
Broad bandwidth	<100 MHz	3 MHz
Small size	-9 mm^2	<9 mm^2
Good temperature stability	~3 ppm/°C	~3 ppm/°C

4.2 MATERIALS AND DESIGN FOR ACOUSTIC FILTERS

4.2.1 Current Application and Band Allocation of Acoustic Filter Technology

Acoustic wave technology has been receiving a lot of attention as a possible substitute for conventional radio frequency (RF) filter operations. This is mostly owing to the fact that it is small in size, inexpensive, and has excellent performance. Bulk acoustic wave (BAW) filters and surface acoustic wave (SAW) filters are the types of radio frequency (RF) filters that are currently the most often used.

The form and center frequency of the passband, which is a property of SAW that propagates laterally, are determined by the pitch, line width, and thickness of the interdigital transducers (IDT), as shown in Figure 4.3. Propagation of the SAW occurs in the latter direction rather than the former. Utilizing wafers for the fabrication of SAW filters enables mass production of these filters at a cost that is more affordable. When filters and duplexers for many bands are combined into a single chip, the amount of additional fabrication that is required is either small or nonexistent. The most important advantages they offer are a low price, a port design that is adaptable, and a really high relative bandwidth. In order to fulfil certain requirements, SAW filters are able to produce low insertion loss and excellent rejection at lower frequency ranges. In contrast to conventional cavities or ceramic filters, they are able to cover a far larger range of bandwidths while occupying a significantly smaller amount of area. SAW filters, on the other hand, are not frequently utilized above approximately 2 GHz due to the fact that their selectivity decreases at higher frequencies.

The majority of applications that require low performance at these higher frequencies are the ones that make use of them. These applications include code division multiple access (CDMA) and global systems for mobile communications (GSM), as well as the third generation of wireless receiver front ends, duplexers, and filters. Additionally, these applications include an assortment of other components. In addition, SAW devices are extraordinarily sensitive to temperature. The decrease in stiffening of the substrate material that occurs as the temperature rises leads to a reduction in both the acoustic velocity and the transmission frequency performance. When it comes to mobile devices, SAW filters often operate on signal frequencies that range from 600 MHz to 2 GHz, whereas BAW filters are located somewhere in the lower 5G bands and operate between 1.5 and 6 GHz.

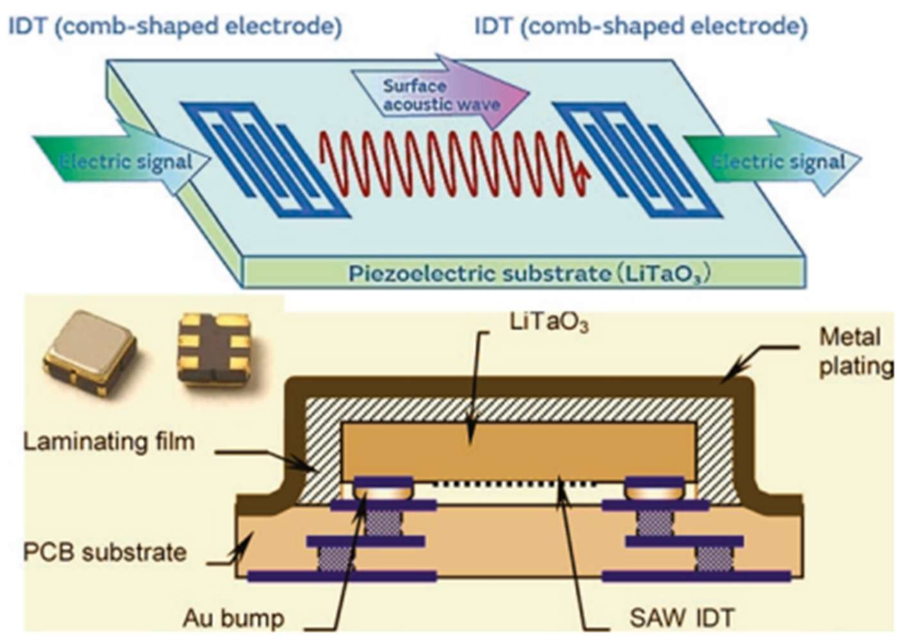


Fig. 4.3 Illustration of surface acoustic wave (SAW) filter and its electronic package (IDT Interdigital transducer)

Source: Modifed from Murata (2022), Credit: Murata Manufacturing Co., Ltd)

The utilization of a more complex and expensive layer structure during the manufacturing process of temperature-compensated SAW (TC-SAW) filters results in an increase in both the working range and the substrate stiffness of these filters. The production of TC-SAW filters is more expensive than the production of standard SAW filters. This is due to the fact that the number of mask layers required to make TC-SAW filters increases throughout the temperature-compensating process. It is important to note that this does not alter the fact that they are more affordable than BAW filters. In comparison to SAW filters, the number of processing steps required for BAWs is over ten times higher than what is required for SAW filters. Even though it produces four times as many pieces per wafer as the SAW filter technology, the BAW filter technique has a higher cost-per-filter expense than the SAW filter technology.

BAW filters are often designed using either film bulk acoustic resonators (FBARs) or solidly mounted resonators (SMRs). Both of these designs are widespread. In Figure 4.4, you can see examples of these designs. In BAW filters, an electric field is utilized to vertically guide an acoustic wave through a piezoelectric substrate. This technique is utilized in order to achieve the desired effect. It is required to precisely tailor the thickness of the piezoelectric layer to very tight tolerances in order to get the resonance frequency that is needed. Control over its thickness must be exercised with great precision. When it comes to selectivity, power management, quality factor (Q), and loss, the device ultimately performs better than SAW filters that are operating at the same higher frequencies.

In FBAR filters, the resonator is completely enclosed within an air interface that is either etched or micromachined. In this particular interaction, the hoover serves as the setting. FBAR filters, on the other hand, have greater technical challenges due to the placement of acoustic reflectors below the bottom electrode. BAW-SMR filters, on the other hand, can be set for wideband performance in frequency areas where FBAR filters are not implemented. In spite of the fact that BAW-SMR and FBAR filters are more expensive to produce, the performance advantages they offer make them more suitable for the majority of LTE frequency bands. Furthermore, due to the short transition range of around 20 MHz that exists between the transmit and receive routes, these features are most effectively used in the PCS frequency band during transmission.

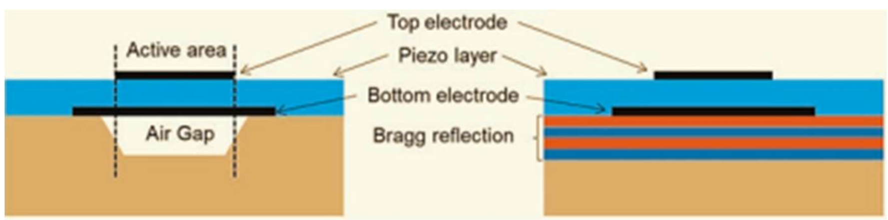


Fig. 4.4 Structure of (a) the film bulk acoustic resonator (FBAR) formed over an air gap on a substrate and (b) the solid mounted resonator (SMR) formed over a Bragg refection layer on a substrate.

Source: Adapted from Liu et al. (2020) with open access (MDPI))

BAW filters, in terms of their capacity to handle RF power, are superior to SAW filters in terms of performance. This is because BAW filters are manufactured using improved processes. In spite of the fact that they are not as effective as a TC-SAW device, they exhibit a lower degree of performance variation with temperature in comparison to SAW devices. The utilization of SiO2 within the reactor results in a significant reduction in the total temperature drift of the BAW, which surpasses the capabilities of both conventional SAW filters and FBAR filters. When compared to an FBAR device, which disperses heat laterally across a far smaller edge area, the BAW-SMR resonator is superior in terms of its ability to dissipate heat. This is because the BAW-SMR resonator is positioned on a solid foundation, which causes this phenomena to take place. As a consequence of this, BAW devices are able to achieve higher power densities, which frees up space for smaller devices to handle power loads of up to 10 W. This makes them appropriate for use in applications that involve small-cell base stations.

The technique for band allocation is depicted in Figure 4.5, along with the current commercial applications of acoustic reactive frequency filter technology. When compared to surface acoustic wave (SAW) resonators, bulk acoustic wave (BAW) resonators have a greater potential to produce radio frequency (RF) filters of superior quality. Specifically, this is due to the fact that BAW resonators demonstrate superior selectivity and less insertion loss in frequency ranges that are more than 2.5 GHz. The BAW is capable of carrying a significant amount of power due to its high sonic velocity and thermal conductivity.

Additionally, the BAW have a high quality factor (Q-factor), which causes the skirts of the flats to be steeper. As a result of their low power consumption, superior isolation, and compatibility with complementary metal oxide semiconductors (CMOS), BAW filters have become a widely used technology in the field of radio frequency (RF) communications. In addition to the reduction in the number of acoustic RF filters, the introduction of 5G technology presents challenges in the areas of data speed and coverage. When taking into account the RF system as a whole, it is reasonable to assume that in the future, the size of the module will decrease while the input power will increase. When it comes to radio frequency (RF) circuits, the BAW resonator is an essential component that has a significant impact on the efficiency of BAW filters.

The coupling coefficient is likely the most challenging issue in this regard, despite the fact that AlN continues to be the material of choice for the mass manufacture of BAW devices. A new design of laterally excited bulk-wave resonators (XBAR) with approximately 25% strong piezoelectric coupling and RF integration technology has been offered with the intention of further increasing the features of BAW devices. This design was developed with the intention of achieving the goal previously mentioned. Through the implementation of this design, there has been an increase in the quantity of BAW resonators and filters that are available for use in radio frequency communication.

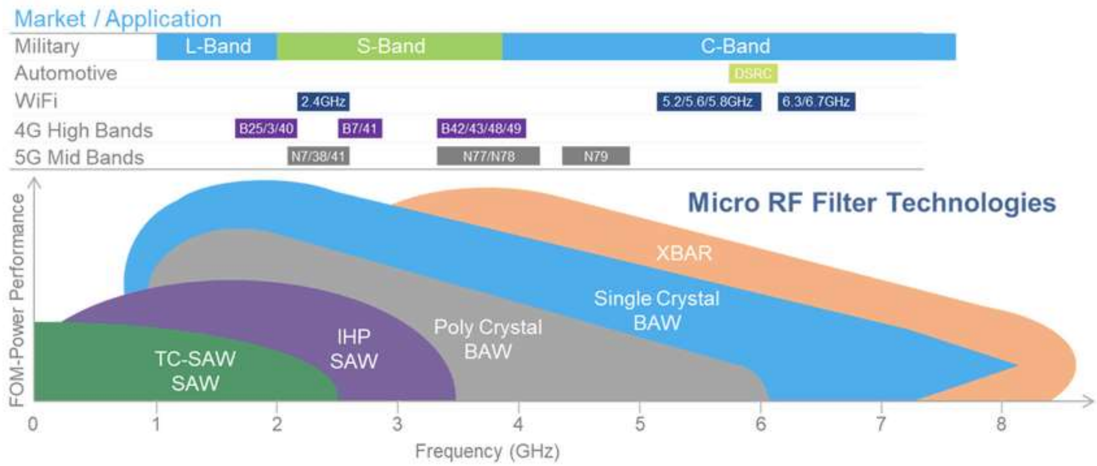


Fig. 4.5 Current market application and band allocation of acoustic radio frequency (RF) filter technologies

Source: Adapted from Liu et al. (2020) with open access (MDPI)). SAW: Surface acoustic wave, TC-SAW: Temperature-compensated surface acoustic wave, IHP SAW: Incredible high performance surface acoustic wave, BAW: Bulk acoustic wave, XBAR: Literally excited bulk-wave resonator

CHAPTER 5

THERMAL MANAGEMENT MATERIALS AND COMPONENTS FOR 5G DEVICES

5.1 INTRODUCTION

The utilization of 5G technologies, radio units, and active antenna devices in more compact locations allows for a greater number of high-power components to be packed into these devices. Because of the increased power, a greater quantity of energy is required to be utilized, which ultimately leads to an increase in the amount of heat produced. It is conceivable for 5G devices to continue to support 5G connections without experiencing any negative impact on their performance. This is made possible by the utilization of efficient heat management techniques. Because the devices are able to keep their functionalities despite their ever-increasing size, this is now feasible. There will be a significant increase in the number of devices that are capable of 5G, beginning with the base station and continuing all the way down to the edge in a variety of different form factors.

At the handset level, it is always ideal to have a longer battery life, which means that it is required to use less power than pre-5G phones. On the other hand, it is necessary to use less power than pre-5G phones. Even if there is a requirement to operate at a reduced power consumption, thermal management will continue to be a crucial concern in the design of future devices, particularly in mmWave phones. This is because thermal management is more important than ever before. A greater quantity of electricity will be required for delayed data processing activities at the base station and data center level as the amount of data that is gathered and processed in the cloud continues to increase. The reason for this is that there will be a greater demand for petrol and energy. The continued rollouts of 5G at both levels will push beyond 6 GHz, which will necessitate a greater utilization of GaN devices to power mmWave 5G networks and devices. This requires a bigger utilization of GaN devices.

Keeping the heat and temperature under control is absolutely vital in order to lengthen the amount of time that these systems may be used, especially for the components. This is especially true for the components. A wide range of issues that manifest themselves in 5G systems are caused by the presence of temperatures that are considered to be high. Among these issues are the acceleration of electromigration, the increase in noise, and the reduction in embedded antenna emission at frequencies that are wanted. It is necessary to find solutions to these problems on three distinct levels: the board level, the circuit design and operation level, and the active thermal management level. It is vital to conform to the form factor specifications of the device in order for it to be deemed a viable solution. However, this must be done without compromising the significant capabilities of the device.

5.1.1 Form Factor-Restricted Heat Management Techniques

Fixing the thermal management difficulties that have been plaguing the 5G network is absolutely necessary in order to guarantee the flexibility, uptime, and dependability of the equipment throughout a wide range of operating situations. The appropriateness and effectiveness of various heat management strategies will be determined by the form factors of the multiple components that make up the infrastructure and devices that make up 5G. This is because the solutions that are utilized in the base station and those that are utilized in the smartphone will not be consistent with one another. It is unthinkable to put enormous cooling fans into consumer mobile phones either because of the intrinsic limits of the designs of existing mobile phones or because of the designs of future mobile phones. High-level signals that are received at a base station have the potential to be impacted by electromagnetic interference (EMI) that is not desirable.

There is a possibility that this interference is caused by components of thermal management, such as fans and heat sinks. Rack-mounted computers are unquestionably responsible for a sizeable percentage of the heat that is produced by the data center building. Due to the fact that these devices are constantly monitored and are not restricted in the same way as base stations are, it is possible to implement active heat management solutions that are more innovative in this environment. Figure 7.1 (Peterson 2020) illustrates that heat management solutions can be beneficial to a variety of devices, including data centers, base stations, and handsets and the Internet of Things technologies.

5.1.2 5G Thermal Management at the Mobile Device Level

As a result of their compact size, headsets are unable to take advantage of active management approaches such as liquid cooling or cooling fans. This is also true for the majority of Internet of Things devices, such as smart home systems and wearables, despite the fact that the Internet of Things encompasses a wide range of topics and incorporates a wide variety of materials and components. The implementation of passive thermal management solutions on the board, in conjunction with circuit optimization, is essential in order to prevent these devices from malfunctioning as a result of excessive power consumption. The continuous functioning of 5G networks might be accomplished by combining passive and active cooling strategies to minimize temperatures in Internet of Things devices that are large enough to have powerful cooling fans or bulky heat sinks. This would allow for better performance of the networks. It is possible that newly developed thermal interface materials (TIMs) that have an exceptionally high thermal conductivity could assist in the dissipation of heat from crucial components and the distribution of that heat throughout the housing of the device. The regulation of temperature across rigid boards may also be possible through the utilization of innovative high-frequency laminates and different kind of substrate materials. On the other hand, given that handset boards are going towards an all-fox design, this strategy will not be effective for them.

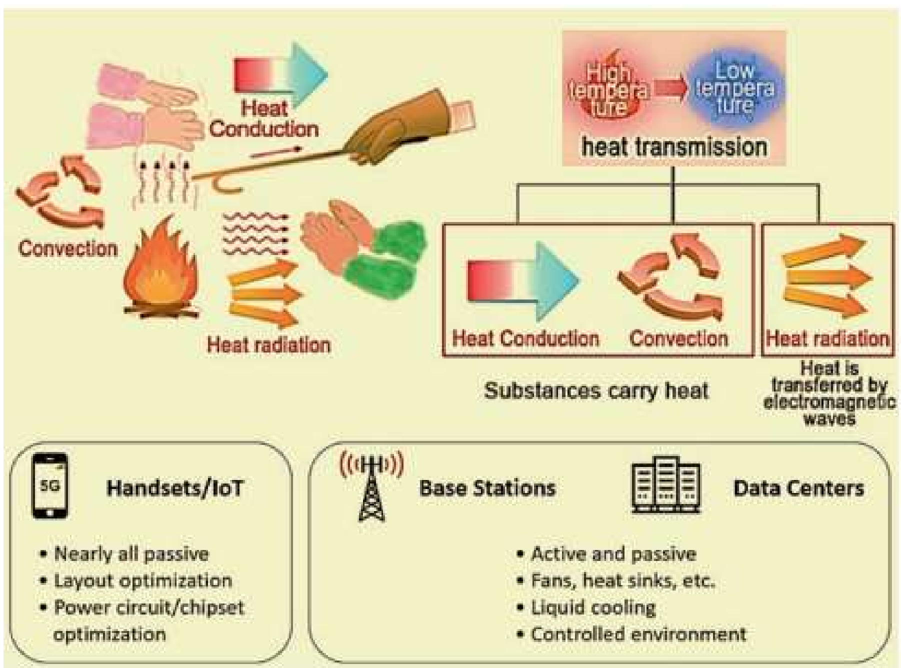


Fig. 5.1 Potential techniques for managing temperature that may be used in 5G devices at different levels

Source: Taken from Peterson, 2020; Northwest Engineering Solutions LLC is credited.

Consequently, this makes it possible for regulator circuitry to make use of a reduced number of components that are both physically and economically significant. In order to accomplish this, the MOSFETs in these regulators are forced to shut down completely at a quicker edge rate, which in turn reduces the amount of power and conduction losses that occur. When a switching regulator with a higher frequency is utilized, an additional problem with electromagnetic interference (EMI) is brought about. In order to reduce the amount of switching noise that occurs in the circuitry that is located downstream, it is necessary to optimize the switching signal, the overall topology, and the design of certain circuits. indicates that the use of smaller components helps with the reduction of designs, the enhancement of efficiency, and the control of thermal conditions all at the same time.

5.1.3 Thermal Management at the Base Station Level

Depending on the hardware of the base station, it may be possible to simulate the power regulator optimization of the phone in a millisecond. There is the potential to replicate the solutions that were implemented at the level of the data center at the level of the base station. For the goal of providing cooling, base station equipment racks typically include standard equipment such as air conditioning, internal fans, and heatsinks on important components with the intention of providing cooling. Ongoing research is being conducted to investigate the use of liquid cooling as a means of removing waste heat, and data centers are now utilizing several systems that have been successfully replicated. This technique makes it possible to do away with massive cooling fans, which are famously noisy when they are powered by electricity.

As a result, the equipment that is produced is not only lighter but also more compact. In addition, it is feasible to deploy advanced thermoelectric assemblies outside, even at base stations, provided that specific conditions are met. The fact that these assemblies are able to pump chilled air directly into rack-mounted base station equipment makes them an excellent alternative to air conditioning systems that are dependent on compressors.

A structured 5G network and subsequent networks will require enhanced heat management in order to accommodate a number of use cases, including the following:

- **(a) Central Offices:** It is anticipated that, as a result of the increased efficiency of the 5G network, there would be a transition away from the traditional headquarters of the telecom industry. When it comes to temperature control, this particular segment of the 5G network will have unique issues. This is due to the fact that the main office will function and appear more like a data center than a traditional telecoms main office. During the process of setting up the technology in these conditions, it is necessary to take into consideration a great deal of important aspects, including cooling, device compatibility, power efficiency, and uptime. Recirculating water chillers, fans, and heat sinks are the most popular types of cooling devices used for servers in today's world. In addition, there are sophisticated temperature management systems in place to control the temperatures in the racks as well as the rooms to ensure that they are at the appropriate levels. In the opposite direction, the new architecture gives us the opportunity to generate new concepts.

(b) Cell Tower Cabinets: In extremely hot or cold areas, cell tower cabinets often require active cooling and temperature control in order to function properly. This is due to the fact that they are located on the exterior of the tower, coupled with the fact that they contain highly advanced electronic components. Similar to the way that cooling has traditionally been accomplished in central offices, it has usually been accomplished through the utilization of compressor-based systems, which have been demonstrated to be inefficient in this regard. Because of these inefficiencies, thermal management systems that have a smaller form factor and the capability to work in applications that are space-constrained will prove to be valuable in the future. These systems will be able to function in applications that have limited space. In the context of this particular application scenario, cutting-edge cooling techniques, such as geothermal or evaporative cooling, might be able to provide cooling measures that are exceptionally efficient.

(c) Small Cells: Through the implementation of tiny cells into the infrastructure, we were able to improve the overall efficiency of the network as well as its capacity and the speed at which data could be sent. When dealing with a large number of users accessing data at the same time, this method is frequently utilized in regions that have a high population density. In the past, many of the places that held cellular equipment had enormous cabinets that were equipped with loud fans in order to maintain a cool environment. In locations with a high population density, the implementation of a cooling solution that is capable of providing sufficient efficiency without the bulk and noise of previous systems would be required in order to fulfil the requirements for the utilization of micro cells.

(d) **Access Points:** In the near future, Wi-Fi access points will have the capability to connect to 5G networks in a manner that is more decentralized and transparent. This opens the door to the possibility of using temperature-sensitive electronics in environments that are either extremely hot or extremely cold, depending on the design of the electronics if they are designed properly. It is possible that thermoelectric coolers (TECs) could be used as a bespoke heating and cooling solution for the component that is sensitive to temperature given the circumstances. Through the utilization of this technology, consumers are able to avoid the utilization of cumbersome heatsinks or electric heaters, hence enhancing the performance and power of their electronic devices.

5.1.4 New Developments in Thermal Management Strategies and Challenges

Because of the large quantity of innovation that will occur in the future, thermal management will be confronted with new obstacles as well as opportunities. Immediately following the establishment of the wireless standard for the fifth generation, this advancement will take place. As a result of developments in beamforming, hybrid digital/analog base station architecture, radio-over-fiber possibilities, and more modern component materials, the approaches that are utilized for heat management need to be rethought. It is imperative that these strategies be rethought promptly.

(a) **Wide-bandgap (WBG) semiconductors: moving beyond silicon:** Power amplifiers made of silicon (Si) are notoriously inefficient in the field of millimetre-wave signal processing and radio frequency (RF) transmission. Their efficiency is barely thirty percent, which is much below the industry standard. Furthermore, despite the fact that these components would have functioned perfectly with previous cell phones, more recent models require alternatives that are more power-efficient for the base stations and handsets. III-V semiconductors are now experiencing increased utilisation as a result of the anticipated jump in demand for 5G components that operate at sub-6 GHz and millimetre waves. For radio frequency (RF) components like power amplifiers, it is possible to attain efficiencies of up to sixty percent by using GaAs, GaN on Si, or GaN on SiC technologies. The price of these materials continues to be higher than that of identical all-Si components, despite the fact that they are already in production and can be purchased in large quantities. GaN on SiC is the material of choice for mm Wave components such as power amplifiers, analogue-to-digital converters, and analogue front-end arrays because, in comparison to other semiconductors, it has a wider bandwidth. It has already been demonstrated that GaN is an appropriate material through the use of electronics in passively cooled base stations and other applications; it will soon also be exploited in mobile phones. The possibility of In GaAs evolving into a III-V material is yet another exciting concept that is compelling researchers to conduct research on the subject. These devices that are based on WBG have the capability of providing the high performance that is required for 5G power applications. It is necessary to exercise control over them because of their exceptionally high-power densities, which in turn requires the investigation of new technologies for heat management.

(b) FM-over-FM: Radio-over-fiber, also known as R.O.F., is an innovative approach that can be utilised to provide broadband wireless access services that utilise microwave and millimetre-wave frequencies. Within the ROF configuration, an infrared light source, more precisely a laser diode, is modulated by a radio wave that has been upconverted from a baseband signal. Sending this modulated signal across a fibre link is the next step that needs to be taken. This optical communication system is able to transmit multigigabit data to base stations over extraordinarily long distances thanks to frequency division multiplexing (DWDM) and optical fibres that are either single-mode fibre (SMF) or multi-mode fibre (MMF). This technology has the potential to remove a significant number of lossy coaxial connections while simultaneously providing a higher bandwidth. When digital pulses are converted to RF analogue pulses and then back again, the process may be advantageous in theory; but it may also cause superfluous heat to be generated and removed from fragile optical components. This heat could be generated and removed from the components. It's possible that radio-over-fiber systems can also benefit from the cooling techniques that were previously described for base stations with similar characteristics. Additionally, it is essential to select the appropriate optical components, which should include FET laser diode drivers. Optical components that are extremely sensitive to variations in temperature include lasers, PIN detectors, and laser diode driver circuits, to name just a few examples.

It is not uncommon for the sensitivity and power output of a device to decrease as the temperature of the device increases. Within a FET driver for long-range communications, the two most critical components are a grounded heat sink and a coupled thermally conductive TIM. Both of these components are highly thermally conductive. There has been a significant increase in the utilisation of thermoelectric coolers (TECs) for the purpose of providing active cooling and accurate temperature management inside the extremely constrained space of the laser diode package over the past several years. Using thermoelectric devices, organisations calibrate the laser so that it emits at a consistent wavelength and data rate. This ensures that the data is transmitted correctly and that it is received at the proper rate. Because of the developments that have been made in the sector, fibre optic networks will be able to increase the rates at which data is transmitted while also reducing the amount of latency that occurs. In order to meet the requirements of the operational criteria, it is anticipated that future thermal energy converters (TECs) employed in these applications will require form factors that are tinier and energy consumption that is lower.

(c) Programmes: With the assistance of newly developed design software tools, innovators are now able to install and test out new devices that are capable of 5G. A total reliance on modularity is one of the defining characteristics of the development of these instrumentation. Utilising design tools like as ARM's Design Start and Gumstix's Geppetto, for example, designers are able to construct new single-board computers and systems-on-chips (SoCs) for Internet of Things products. The following paragraphs will provide you with additional information regarding both of these technologies. When it comes to activities such as designing single-board computers or system-on-a-chips, a modular approach is an excellent technique for electronics engineers to circumvent their lack of expertise in a particular field.

In order to alleviate the load that is currently being placed on matching processing at base stations, these technologies will make it possible for designers to move their processing jobs to the perimeter of the network, which is located away from data centres. The development of new embedded devices that are capable of being updated to 5G capabilities is also within reach as a result of these technologies. These devices have the potential to be useful in a wide variety of industries, including robotics, healthcare, automotive, intelligent infrastructure, and numerous others. According to Peterson 2020, prospective designers may quickly begin adding a variety of heat management strategies into their 5G-capable products and solutions by making use of these design tools.

5.2 THERMOMANAGEMENT MATERIALS AND COMPONENTS FOR MOBILE DEVICES WITH 5G SUPPORT

Heat regulation has become a more severe challenge than it was in the past as a result of the development of devices that are enabled with 5G. Because of this, we have been forced to make some significant structural adjustments and upgrade to materials of a higher quality. One of the primary objectives of thermal management materials is to lessen the occurrence of hot spots by easing the distribution of heat in a manner that is both uniform and strategic. These materials can be broken down into three primary categories: thermal interface materials (TIM), heat spreaders (which include graphitic sheets, heat pipes, and vapour cambers), and thermal insulation materials. Dissipation in the z- or through-plane direction, heat spreaders for x-y or in-plane dissipation, and thermal insulation materials to prevent surface hot spots and limit RF interference are the primary applications for thermal insulation materials (TIMs). For the purpose of carrying out their respective activities, a wide variety of various areas are dependent on certain essential components and materials.

These materials and components are primarily located in the following areas of a smartphone: the back of the phone, the central motherboard, the antenna, the lithium-ion battery, the heat spreader behind the display, and the thermal interface module (TIM) for the 5G mmWave antennas. These areas are designed to prevent surface hot spots while simultaneously minimizing RF interference. For each of them, there are a number of different considerations and requirements that must be taken into account. In addition to the most important characteristic, which is the heat conductivity in the x-y or z-plane, it is essential to take into account other aspects such as adhesiveness, viscosity, bond line thickness, rework ability, and longevity.

When it comes to heat management materials, the primary components that will undergo the most significant modifications at the device level are the motherboard and the antenna components. When it comes to the motherboard, this is somewhat clear due to the fact that 5G will make it possible to achieve substantially better download rates; the primary antenna adjustments will be for mmWave applications. It is quite improbable that the current trends in smartphone performance, battery life, and displays will change in the near future. There is also the possibility of replacing the graphitic heat spreader on the motherboard with either a heat pipe or a vapour chamber. It is a given that performance will improve, but there are a lot of benefits and drawbacks to take into consideration as well.

5.2.1 Fundamental Solutions and Design of Thermal Management for Smartphones

It is now necessary for the primary components of smartphones to have a higher power consumption because their performance is always improving. Through the process of integration, either the number of components or the size of those components has been reduced in order to maximize the amount of battery space that is available within the limits of the design of a smartphone. In circumstances where there is a demand for power, this makes it possible for the gadget to be utilized for a longer period of time. In addition to this, printed circuit boards (PCBs) possessing two or even three layers are currently being developed in order to reduce the size of smartphones while simultaneously expanding the mounting area. The problem of hot spots is made worse by the fact that components on multiple layers of the printed circuit board have complicated thermal connections. In order to establish 5G connections, additional 5G specialized antenna modules are necessary.

As a result of the significant increase in power consumption, the radio frequency (RF) components of the 5G modem as well as other components related to energy were manufactured. For the reason that there are a great number of passive devices that require power-related components, the positioning of these components on the printed circuit board (PCB) is of the utmost importance.

Additionally, memory, modems, cameras, and access points are among the most important components that consume a significant amount of power. 5G antennas are also included in this category. The power integrated circuit (PMIC) and the primary components should be located in close proximity to one another, as this is the intended configuration. A popular location for power IC and AP is roughly in the middle of the printed circuit board (PCB), which is located between the top and bottom of the board. Ultimately, it is these central processing units that are responsible for the high temperature of a smartphone. 5G base stations link with smartphones at maximum throughput, and as illustrated in Figure 5.2, the chip in the phone must be able to execute a significant number of calculations due to the high throughput. It is possible that the temperature of the gadget will rise as a result of the significant amount of heat that is produced by calculations.

Because of the tendency towards smaller form factors combined with better performance, overheating, which includes hot spots, is becoming more common and difficult to manage. This is a result of the trend. The combination of these two elements has played a role in this. Components that are based on both software and hardware are frequently incorporated into methods for regulating the heat produced by smartphones. In software, there are options for using both static and adaptive strategies to regulate the amount of power that is being used.

It is possible for these algorithms to accomplish this by dynamically limiting performance by modifying the voltage and clock frequency of the central processing unit. It is the responsibility of an adaptive system to adjust the load in accordance with its best estimation of whether or not additional power is required. When it comes to the hardware side of mobile phone thermal design, it is very important to have efficient heat dispersion within a thickness that is strictly controlled. Taking this action ensures that the device's outside shell is heated consistently and eliminates the possibility of hot spots occurring. You can prevent your skin from being overly heated and from experiencing unexpected spikes in temperature by making strategic use of the air spaces that are present throughout your skin. In addition to other functions, you should declare that the thermal inertia of the frame and the battery acts as a buffer.

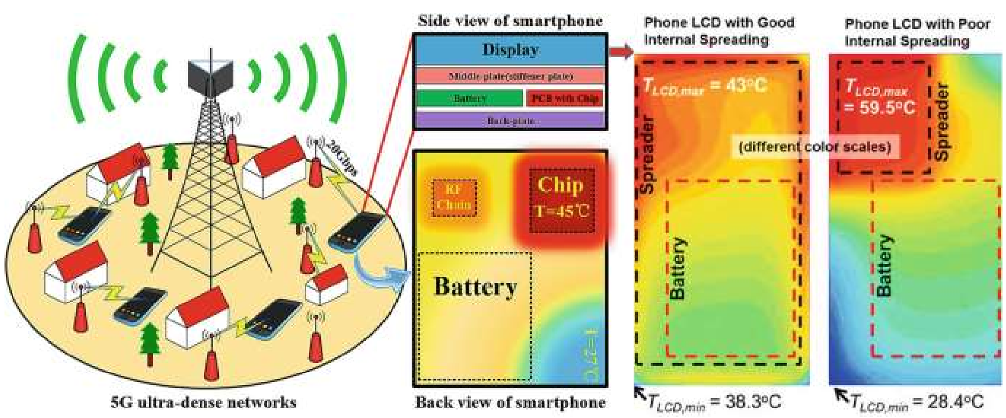


Fig. 5.2 The composition and temperature range of cellphones

Source: Extracted with permission from IEEE from Chiriac et al., 2016 and Yang et al., 2019

5.2.1.1 Guidelines for the Design of Thermal Management Programs

Finding the optimal balance between the amount of processing power and the amount of power that is consumed is turning out to be a significant element in 5G devices. The development of innovative active thermal solutions necessitates the capability of heat radiating, operation that is free of vibration and noise, scalability to smaller form factors, dependability, and cost. In order to solve concerns regarding skin temperature ergonomics and battery life, wearables need to feature low-power operation. The junction temperature is not the most important limitation for the thermal design of smartphones; rather, the skin temperature at the outside of the device is the most important limitation. The designs were developed with the intention of achieving the highest possible level of temperature uniformity, which is defined as a ratio of average to maximum temperature that is greater than 0.8, and ensuring that the temperature of the skin does not exceed 40 to 45 degrees Celsius.

Working with portable electronics presents a variety of challenges that are unique to thermal design. These challenges can be broken down significantly. The limited space available for the insertion of thermal solutions and the restrictions placed on the maximum temperature at which the surface can be considered acceptable are two examples of these features. When it comes to power dissipation, cellphones are only capable of dissipating between two and three watts at the moment.

Because of the continual contact that they have with the human body, emerging technologies such as smartwatches make the problem even worse. Whole system models are utilized in the process of designing thermal management systems for smartphones and other mobile electronic devices. The goal of these models is to provide an accurate depiction of the thermal behavior at each level, beginning with the transistor level (response times measured in microseconds) and progressing all the way up to the device level (reaction times ranging from minutes to hours). Tablets, on the other hand, have a range of minutes, whereas phones have a range of seconds. This is because different devices have distinct ergonomic ranges and use durations.

These thermally aware systems are comprised of a number of sensors, which enable performance throttling as well as a variety of temperature control options, including the ability to make predictions. A limited heat path to the outside is the result of the layout's continuous difficulty to achieve for CPU-topped memory systems. This is the outcome of the configuration. The implementation of the arrangement continues to be a challenge, despite the fact that TIM layers optimized the capacitance and the internal heat conduction. A memory, radio-frequency (RF), mixed-signal, and central processing unit (CPU) are all included within a single package that is referred to as a system-on-a-chip (SoC) in contemporary smartphones.

In light of the circumstances, there are two primary challenges that need to be conquered: the double heat flow in two-die stacks and the requirement to stack memory on top of logic processors in order to satisfy multiple thermal limitations in a single box. Both of these challenges are necessary in order to solve the issue. In terms of thermal design, portable electronics need for entirely different paradigms than those that are utilized for the purpose of cooling large-scale electronics. When it comes to this particular situation, customer-centric performance is of the utmost significance, and ergonomic considerations such as skin temperature constraints have an impact on the thermal goals. In light of the aforementioned, it is necessary for thermal engineering and electrical engineering to collaborate simultaneously in order to achieve this end of the design spectrum. This is due to the fact that performance is increasing while form factors are decreasing.

5.2.1.2 Essential Solutions for Thermal Management

A multitude of cooling methods have been developed, with hybrid, active, and passive cooling being the three primary classifications that have been established, respectively. Heat pipes, vapour chambers, thermoelectric coolers, heat spreaders, heat sinks, and thermal interface materials are all examples of passive cooling methods. Other examples include heat spreaders and heat sinks. Active cooling methods include two types of cooling technology: forced air cooling and forced liquid cooling. Both of these types of cooling technology are available. The process of hybrid cooling is achieved by the combination of active and passive cooling systems. Active cooling, on the other hand, is not a viable option for smartphones because of their very tiny size and the weight restrictions they have. It is for this reason that passive cooling is currently the standard for cooling smartphones. Particularly at the level of certain electrical components, thermal management takes use of power control that is based on software.

Not to mention all of the different approaches that can be taken to make temperature adjustments. These days, application processors (APs) are equipped with capabilities such as digital voltage feedback (DVFS) and CPU hot-plug, which enable users to control the amount of power that is consumed by devices and prevent them from operating at an excessive temperature.

Through the prevention of a central processing unit (CPU) from overheating, a thermal protection system can reduce the amount of power that is consumed by the CPU by lowering its operating voltage or frequency, or even turning off its cores. An example of this would be the thermal engine that is included in the Linux kernel. This engine is used for thermal management on Android devices, which is dependent on on-die sensors to monitor CPU temperatures. The threshold control algorithm and the dynamic control algorithm are the two algorithms that are essential to our approach. On the basis of whether or not the cores' temperatures are higher than the threshold values, the dynamic control algorithm makes a decision to either decrease the clock frequency or disable the cores simultaneously. The threshold control algorithm used by the manufacturer is responsible for determining the threshold values, which are subject to vary over time. Furthermore, the smartphone will be rendered inoperable by Android's thermal safety function in the event that the temperature of the battery increases above a predetermined level, such as 68 degrees Celsius.

5.2.1.2.1 Conduction and the Heat Transfer Process

Using Substrate Materials with High Heat Conductivity:

For usage in high-heat-flux applications such as transportation, defence, and high-end servers, innovative materials and methods for interfaces with near-junction high thermal conductivity have been developed. Additionally, these materials and methods have been developed. These materials and processes, which are utilized throughout the application process, make it possible to achieve targeted heat conduction and spreading. As a result of its exceptionally high conductivity, diamond is an excellent material for applications that involve the dispersion of heat. The use of materials that are produced from carbon is being implemented in order to develop solutions that have outstanding conductivity and low density. The only dimensions in which carbon nanotubes and graphene exhibit a high degree of conductivity are two and one.

Nevertheless, the extraordinary conductivity of diamond may be observed in all three dimensions respectively. For radar and 5G telecommunication applications, high-electron-mobility wide bandgap transistor electronics (for example, using GaN, GaAs, Al GaAs, and In GaAs) are two primary areas that require materials with high thermal conductivity. High-performance computing (HPC) is another area that requires materials with high thermal conductivity. Integrated packages that are three-dimensional and contain logic and memory chips stacked one on top of the other are what high-performance computing makes use of. It is possible for these broad bandgap electronics to experience extremely high local fluxes (above 10 kW/cm2) or overall averages (above 1 kW/cm2) throughout the implant. One method involves the fabrication of GaN in conjunction with an intermediary layer of amorphous SiN or SiC on a substrate made of diamond. We are in the process of creating these methods right now. It is required that the layer have a thickness of thirty nanometers. It is becoming increasingly common to produce substrates with lower thermal resistance by mixing different materials to variable degrees. For example, GaN on silicon, GaN on SiC, and GaN on diamond are all examples of substrates that have been created.

Dispersal and Storage of Heat in Mobile Devices:

There are a variety of approaches, both software and hardware-based, that can be utilized to maintain a consistent temperature for mobile phones. In software, there are options for using both static and adaptive strategies to regulate the amount of power that is being used. It is possible for these algorithms to accomplish this by dynamically limiting performance by modifying the voltage and clock frequency of the central processing unit. It is the responsibility of an adaptive system to adjust the load in accordance with its best estimation of whether or not additional power is required. It is essential to have efficient heat dispersion within a severely constrained thickness on the hardware side of the mobile device in order to prevent hot spots and ensure that the surface temperature is uniform across the exterior shell of the device. You can prevent your skin from being overly heated and from experiencing unexpected spikes in temperature by making strategic use of the air spaces that are present throughout your skin. Additionally, there are buffers between the two as a result of the thermal inertia of the frame and the battery.

Additionally, modest thermal models can be beneficial to operational management and design; however, these models need to take into account the thermal interaction that occurs between the CPU and the battery. In the central processor unit of a mobile phone, there are two types of power consumption: dynamic, which is produced by capacitance current, and static, which is produced by leakage and standby current. Both of these types of power consumption are necessary for the device to function properly. Phase-change materials, sometimes known as PCMs for short, are one possible solution to the problem of energy storage that arises in the dynamic response region.

Nevertheless, mechanical components and the thermal mass of batteries are the examples of the most popular types of energy storage. When fluxes are high, it is difficult to remove heat from PCMs due to their poor thermal diffusivity. This is because PCMs have a low potential for heat transfer. A partial solution to the problem could be achieved by substituting a small quantity of PCMs with heat-spreading graphite sheets. This is because the thickness of these sheets ranges from 10 to 100 micrometers. The technologies that provide heat management for mobile phones have developed over the years, and the PCMs would melt in a matter of seconds if they were used under these circumstances. During the introduction of the first iPhone in the middle of the 2000s, all that was required were features such as performance throttling and a TIM on a metal surface. Now, it is very necessary to have a thermal management unit that is equipped with a multitude of sensors and graphite heat spreaders.

It is unfortunate, but not unheard of, that mobile devices have been known to make use of vapour chambers as a passive distribution mechanism. On the other hand, mobile thermal management cannot fulfil ergonomic requirements in ultrathin form factors if it does not have an efficient heat dispersion system. To determine the performance limitations beyond which a vapour chamber, with its geometry and heat input, is more successful than a solid heat spreader of the same size, we first determined the primary heat transfer processes that affect the functioning of vapour chambers with decreasing thickness. This allowed us to determine the limit beyond which a vapour chamber is more effective than a solid heat spreader. The vapor-phase behavior of ultrathin vapour chambers that operate at low power should be taken into consideration when modifying design technique and working filament selection criteria. This behavior ultimately impacts the heat spreading resistance of the chambers; hence it is important to take this behavior into account.

Through the utilization of a graphene heat spreader for the purpose of heat dispersion, it is possible to rapidly change heat into an active cold plate through the process of thermoelectric cooling. This results in cooling that is both more efficient and more environmentally friendly. In most cases, a thermoelectric cooler (TEC) is positioned between two heat sources or sinks. This device is characterized by its thin and tiny size. Its primary function is to improve the efficiency with which heat can be disposed of. Conduction-based heat transfer is made feasible by the temperature difference that occurs between the two sides of a thermoelectric converter (TEC) when voltage is applied. This difference in temperature makes it possible for heat to be transferred.

5.2.1.2.2 Air Convection Cooling

On the other hand, radiation and natural convection are now the most effective methods for cooling many smaller platforms; but more advanced systems may soon begin to use forced air convection. It has been established through research that air cooling using fans is a realistic choice, despite the fact that the demand for thermal energy is growing consistently. The following circumstances, which are detailed below, could potentially benefit from developments in alternate approaches of air movement which are mentioned below: The first objective is to enhance the natural convection of small-scale platforms, thereby eliminating the need for spinning fans. The second objective is to optimize fans for cooling components of servers that have moderate heat fluxes, such as liquid-cooled central processing units in hybrid-cooled servers.

The use of piezoelectric fans as a cooling solution has been demonstrated to be effective when the fans are operated under conditions that are properly controlled. Their primary characteristics are their low cost, great dependability, low noise level, and low power consumption. Additionally, they have a low environmental impact. However, while operating in hostile environments, dust fouling and deposition on the blade of the oscillating fan can cause the resonance frequency of the fan to change. By altering the actuation frequency based on a measurement of the free resonance frequency, vibration frequency tracking may provide a partial solution to this problem. By doing so, the problem can be partially solved. Because of their excellent thermal performance, which has been demonstrated in the mild heat flux range of 0.1-1 W/cm2, they are ideal for low-power platforms (below 40 W) and the overlap zone (15-40 W) between forced convection heat sinks and natural convection heat sinks. This is because of the fact that they allow for a greater amount of heat to be transferred.

Artificial jets have the ability to provide solutions in the range of heat fluxes that are intermediate between forced and natural convection. Furthermore, well-built actuators and ducting have the ability to create local cooling that is comparable to steady impinging jets in the range of 0.2-5 W/cm2. Due to the fact that piezo fans and synthetic jets are essentially agitators, it is imperative that the entire air route be meticulously planned in order to avoid getting heated air back into the system. When compared to axial fans, synthetic jets have the potential to exceed them in terms of thermal performance, which would result in a quieter operation.

Additionally, the capability to adjust the jet flow angle by utilizing two adjacent phase-controlled jets that are close to one another makes active cooling management a viable option. Through the utilization of a combination of low-speed fans, synthetic jets, and piezo fans, it is possible to differentiate between the local heat transfer and the thermal advection that is brought about by masses of airflow. In the event that high pressure drop air-cooled heatsinks on server CPUs are replaced by compact liquid-cooled heat sinks, the server cooling industry will be open to the use of synthetic jets and piezo fans. It is possible to make significant adjustments to the overall budget for pumping power as a result of this. This has the potential to result in an increase in the energy efficiency of data centers while simultaneously reducing the expenses of maintenance.

In order to eliminate the tone noise that is created by the passage of blades, a number of different alternative air movers have been investigated. These include electro aerodynamics for platforms that are smaller than three millimeters, both of which are not suitable for bearings, and bladeless fans. A corona discharge could be produced in the air by the latter, which makes use of piezoelectric transformers that operate at low voltages.

It is also feasible to achieve better air conditioning through the use of other methods. Opportunities include the following:

(i). The utilisation of sophisticated tools for the purpose of optimising and modelling foams; and

(ii). The combination of numerical form optimisation with additive manufacturing of validation prototypes, which may include selective laser sintering of aluminium alloys (such as AlSi12). Given that high-end portable electronics like as 5G smartphones and tablets are straining the boundaries of radiation and natural convection as the ultimate heat sinks and spreaders, it is possible that some of these innovative air-moving technologies will find usage in a range of platforms that are yet to be developed.

5.2.1.2.3 Liquid Cooling Convective

Current efforts in the field of component-level liquid cooling are mostly focused on two primary areas: the enhancement of heat transmission and the cooling of liquid chips that are specifically targeted. Therefore, this is due to the fact that liquids, and water in particular, possess a volumetric heat capacity that is substantially higher than that of air. In single-phase cooling systems, it is essential to drastically reduce the hydraulic diameter of the cooling channels in order to achieve the desired improvement in heat transfer coefficients. The utilization of such diminutive structures would result in a significant increase in expenses, and there is a greater likelihood that they would become obstructed. These small-scale fow channels would be eliminated if a two-phase fow loop were used, despite the fact that it is more difficult and might potentially increase the overall cost of the system. Regarding this matter, there is a need for some consideration.

5.2.2 Material Choice for Heat Sinks and Spreaders

In order to cool an object using air, one of two things must take place: either the surface area of the object must be significantly expanded, or the air flow over its surface must be increased. Both of these conditions must be met. Figure 5.3 illustrates the normal dispersion of heat through the utilization of heat sinks and spreaders. The notion of push-pull air flow is employed in order to achieve high heat flux capabilities while simultaneously achieving a considerable pressure decrease. The top and bottom of the printed circuit boards (PCBs) are where the components of this design that move air are situated. Both convection and ambient radiation are viable options for effectively cooling low-power electronic devices according to their respective cooling capabilities. When natural convection is not sufficient to move air through enclosures holding electrical components, fans or blowers will use forced convection to move air through the enclosures. The transfer of heat from a heat source to a secondary heat exchanger, such as a heat sink, is accomplished through the utilization of a heat spreader, which is a primary heat exchanger that is typically constructed upon an extraordinarily thermally conductive plate (Fig. 5.3a).

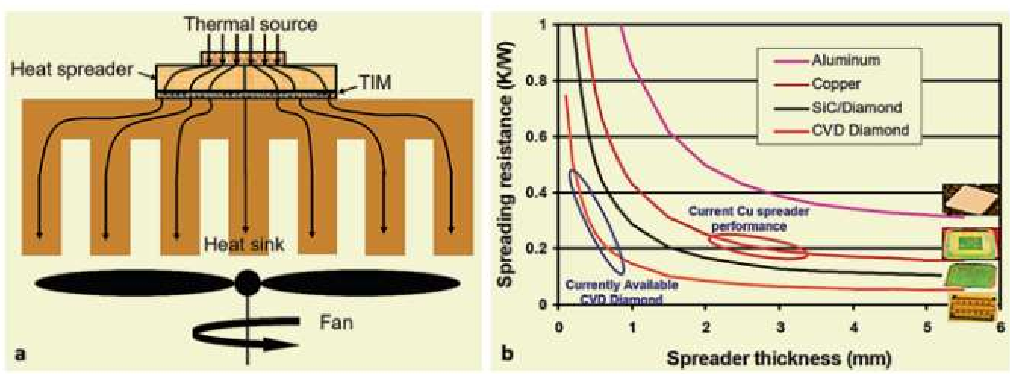


Fig. 5.3 A typical heat spreading and heat sink system (a) and a spreading model that may be used to achieve the same spreading resistance with varying spreading thickness (b) using various materials

Source: Taken from Tong (2011); Springer Nature is credited.

Passive heat sinks are considered to be entirely reliable due to the fact that they do not have any mechanical components. Aluminum radiators are examples of passive heat sinks that are created from metals that are capable of dissipating heat through convection. It is the majority of passive heat sinks. The removal of heat through the use of air cooling and the building of heat sinks of the next generation has both been made possible by the development of novel materials. Sintered heat sinks, self-fanning piezoelectric fn heat sinks, high-performance fans, and aluminum extrusions with a 25:1 ratio are all examples of heat sinks. Other examples include high-performance fans. Materials such as strand-oriented graphite, carbon-based composites, phase-change compounds, high thermal conductivity polymers, and sintered metallic powders are some examples of contemporary heat sink materials. The most desirable features include weight, thermal expansion coefficient, toxicity, machinability, controlled conductivity, and, most critically, a forecasted cost that is lower than that of metallic aluminum. As a result of the potential uses of nanotechnology, composite heat sinks are receiving an increasing amount of commercial attention.

5.2.3 Mobile Electronic Devices: Vapour Chambers and Flat Plate Heat Pipes

The trend in mobile electronics, which includes tablets and smartphones, is towards devices that are smaller and have more capabilities. This tendency results in a rise in the density of heat generated from active components. For the purpose of eliminating the thermal energy that has been created, active air cooling or enormous heat sinks are not suitable options since the client has imposed size constraints. Natural convection cooling is a passive cooling technology that depends mostly on the surface area of the mobile device and the temperature difference between the outside and ambient settings. The temperature differential between the outside and ambient settings is primarily dictated by the user's degree of comfort. Therefore, in order to maximize the efficiency of the device and the temperature of the skin, the thermal energy that is generated by the central processing unit (CPU), the battery, and any other active components must be dispersed throughout the whole surface of the device. This will increase the amount of natural convection cooling that occurs.

It is possible that ultrathin vapour chambers or flat plate heat pipes might be a realistic option for increasing thermal capacity for transient conditions and passively distributing the thermal energy that is generated inside mobile devices. A vapour chamber, also known as a fat plate heat pipe, is a two-phase cooling system that involves the passive transmission of heat from a very small source to a somewhat larger heat rejection surface. This kind of system is quite successful, as can be seen in Figure 5.4. They provide effective thermal energy conveyance by utilizing the latent heat associated with phase shift. This has special benefits in terms of temperature uniformity and the removal of local hot spots.

According to Gibbons et al. 2021, the effective thermal conductivities of these materials are orders of magnitude higher than those of their homologous materials that are composed of solid components. The performance of the heat pipe is significantly influenced by the effectiveness of the wick in recirculating the working fluid from the condenser to the evaporator. Capillarity and permeability are the performance characteristics of a wick structure that are considered to be the most important elements. Traditionally, the inner surfaces of the heat pipe chamber are etched and coated, metal foams, sintered powders, or woven wire meshes are used for the production of heat pipe wicks. Other options include fabricated wire meshes.

wire meshes are often used as wicking structures in ultrathin flexible heat pipe applications. This is due to the fact that wire meshes are also affordable, ductile, and easy to fabricate.

Three metrics are often used in order to evaluate the performance of heat pipes: (i) The capacity to transfer high temperatures $k = \frac{qL}{\Delta TA}$ [W/(mK)], where A represents the cross-sectional area that is perpendicular to the direction of heat flow, ΔT represents the temperature difference which occurs throughout the characteristic length, L represents the length that occurs between temperature measurements, and q represents the power that is delivered. The thermal resistance, which is defined as R = ΔT/q [K/W], is a measure that indicates the reduction in temperature of the device when a certain amount of thermal power is given. Neither the geometry of the gadget nor its shape is taken into account. It is a dimensionless measurement of thermal spreading that is referred to as the Coefficient of Thermal Spreading (CTS = (Tave T)/(Tmax T). Not only does it take into account the highest surface temperature Tmax, but it also takes into consideration the average surface temperature Tave, as well as the ambient far field temperature T.

When measuring the effectiveness of heat distribution, the CTS takes into consideration a number of factors, one of which is the degree to which the surface temperature is consistent across the entire region. It can be determined by the use of infrared thermography or by integrating a number of observations of the temperature of the surrounding environment. This figure is extremely important with regard to mobile electronics because natural convection is a typical approach for cooling them. If the temperature distribution is not uniform, this is an indication that the thermal spreading is not optimal. Thermal management solutions for mobile devices should strive to achieve a thermal resistance of 6250 W/m-K and a thermal design power of around 5 W. This is the ideal specification for thermal management. In light of the fact that even a temperature difference of one degree Celsius can have a detrimental impact on the screen and battery life of a device, heat pipes are the most effective method for accomplishing this objective.

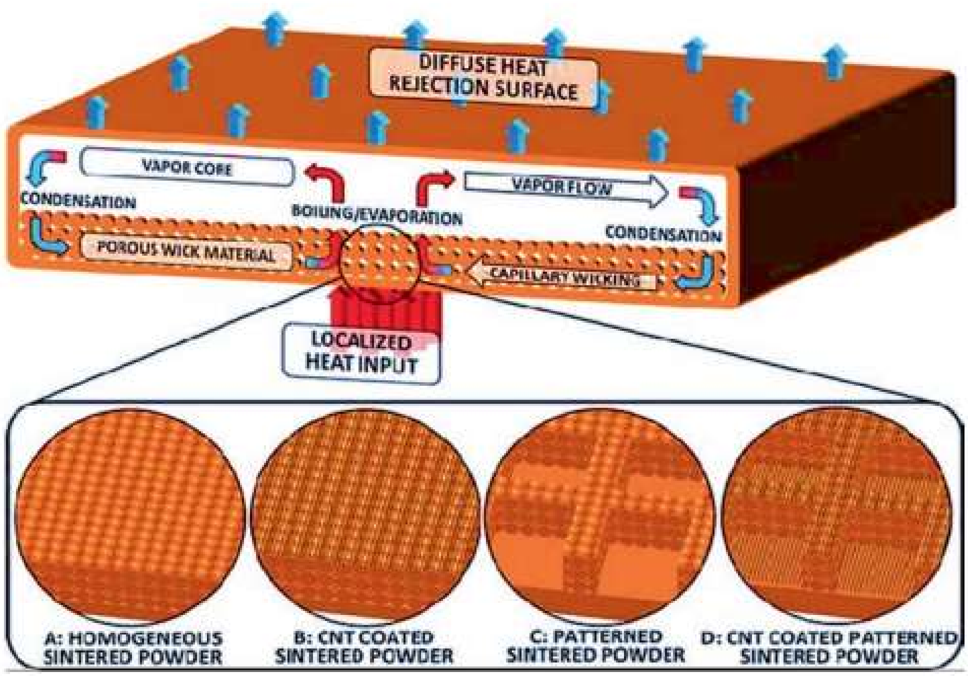


Fig. 5.4 An internal transport process schematic for a vapour chamber heat spreader with nanostructured and micropatterned surfaces

Source: Taken from Weibel et al., 2012; IEEE permission granted.

In Figure 5.5, a number of different fat heat pipes and vapour chamber methods are compared with regard to the thermal qualities the methods possess. To provide a visual representation of the fact that the entire top surface is being used as a condenser, the filled markings serve as a visual representation. The heat source of this kind of system is normally situated in the middle of the system, and the heat pipe enables the heat to be transmitted in both the horizontal and vertical directions. There is a piece of the heat pipe that is moving adiabatically at each of the empty markers. This indicates that the flow of heat is only going in one direction. In Figure 5.5a, a comparison is provided between the thermal resistance of the heat pipe and the amount of thermal power that it supplies when it is implemented. When larger thermal powers need to be conveyed, it is common practice to adopt macroscale stiff or rigid ultrathin heat pipes in order to reduce thermal resistances. One example of this is the temperature drop that occurs between the evaporator and the condenser.

However, because ultrathin and flexible macroscale heat pipes are able to convey less thermal energy, the thermal resistance of these pipes is higher. This is a consequence of the reduced thermal energy that can be transmitted. In Figure 5.5b, the relationship between the amount of power that is delivered and the thermal conductivity of the heat pipe is depicted. It is the ultrathin rigid heat pipes that have the highest heat conductivities among all of the heat pipes.

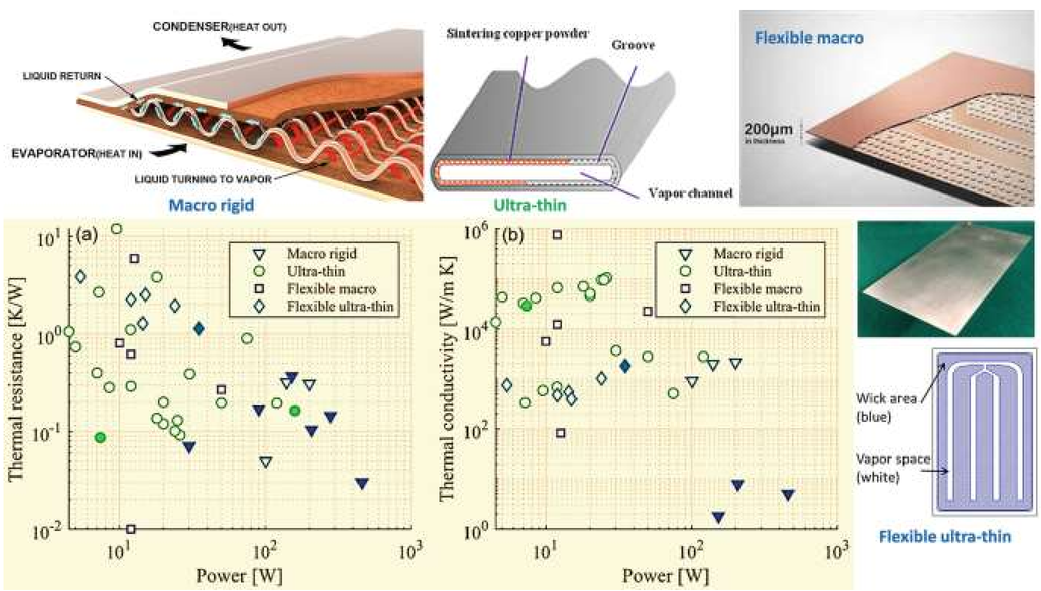


Fig. 5.5 (Adapted from Gibbons et al., 2021) with open access (Elsevier)) Thermal properties of various vapour chambers or fat heat pipes: Thermal conductivity and resistance (a and b) respectively. When the whole top surface is used as a condenser, filled markings indicate this.

The stiff heat pipes at the macroscale are the ones responsible for the largest thermal power transmission, despite the fact that they have low thermal conductivities. It has been shown that both microscale and nanoscale flexible enclosures have thermal power transfer and thermal conductivities that are intermediate. Additionally, there is evidence that thermal power is being transmitted. Given the following performance costs, it is reasonable to anticipate that flexible heat pipes will have a worse thermal performance than other thermal pipes: (i) changes to the vapour space and wicking structure during heat pipe deformation reduce working fluid recirculation; (ii) lower thermal conductivity from the flexible materials used for the heat pipe casing; and (iii) difficulty in maintaining high vacuum due to the increased permeability of the casing materials in comparison to rigid ones. All of these factors contribute to that difficulty.

It is still in its infancy that the technology for ultrathin heat pipes has been developed. When the temperature is at room temperature, the wick capillary pressure places a limit on the amount of heat that can be transmitted via a heat pipe. The following is the pressure balance that exists inside a heat pipe when it is in a state of equilibrium,

$$\Delta P_{cap} \geq \Delta P_v + \Delta P_l + \Delta P_g \quad \text{.....5.1}$$

where ΔPv represents the vapour pressure that is transferred from the evaporation section to the condensation section, ΔPl represents the liquid pressure drop that occurs from the condensation section to the evaporation section, and ΔPg represents the pressure drop that is brought about by gravity. Moreover, the capillary force that is given by the wick is denoted by the symbol ΔPcap. It is possible that the following factors will have an impact on the research that is conducted in the area of ultrathin flexible heat pipes:

1. **Shell Substance:** Creating heat pipes that are flexible and ultrathin requires the first step to be the identification of an acceptable shell material. Polymer-based casings are materials that have the potential to be useful. Polymers were selected for the purpose of fabricating an ultrathin fat plate heat pipe due to the substantial advantages they provide in terms of weight, flexibility, chemical resistance, electrical insulating characteristics, and compatibility with the majority of the processes used in the fabrication of electronics. (Liu et al. 2019; Weibel et al. 2012) The following is a list of the key challenges that face the use of polymer heat pipes, as well as possible solutions to these challenges:

(a) Heat conductivities of metals are much greater than those of polymers. This issue may be circumvented by the use of heat vias, composite materials that possess a high heat conductivity, or hybrid construction options.
(b) Metal wick structures, such as sintered copper powder and meshes, have a better temperature resistance than polymer wick structures. This is because metal wick structures cannot be sintered with the polymer casing.
(c) It is possible that heat pipes constructed of polymers may have problems with outgassing and will emit gases that are not condensable, which may impair their reliability over the long run. The use of a flexible gas barrier, such as aluminum foil, or a coating, such as indium or copper, are two potential methods that might be implemented.
(d) In general, polymers have a lower degree of wettability when using working fluids. There is a possibility that this will aid to enhance dropwise condensation in the condenser region. It is possible to circumvent this difficulty by using either a wicking structure that is sufficiently hydrophilic or an extended evaporator part.
(e) Polyimide, ultrathin copper, and hybrid materials with layers that are impermeable are examples of materials that show promise for use in flexible case construction. Additionally, it is essential to give some thought to the possibility of integrating and attaching the flexible material to a rigid component (which has a higher wall conductivity). Keeping the thickness to a minimum while maintaining the thickness of the internal vapour core is one of the obstacles that must be overcome in this situation. A strong collaboration with the material sciences will be required for the development of a potential future solution.
(f) A noticeable challenge for flexible heat pipe technology is the disruption of normal vapour flow and fume recirculation that occurs during articulation. This also results in an increase in the thermal resistance between liquid and vapour in the evaporator section of the system.

As the curvature of the heat pipe increases, articulation may result in the formation of a liquid plug in the inner chamber. This is due to the fact that the effective channel area of the wick is not very large. As a result, the permitted rate of heat dissipation is reduced. The possibility of this happening is reduced if a support structure is included at the articulation joint in order to prevent the collapse of the vapour gap. It is possible for the heat pipe to concurrently have a bend line for articulation, a decreased performance when folded, and a regular performance when open. In the event where the foldable device is simply needed to be open, it is feasible that complete functionality will be required. Increased efficiency in the heat-transfer pressure balance:

$$\Delta P_{cap} \geq \Delta P_{v} + \Delta P_{l} + \Delta P_{g} + \Delta P_{bend} \qquad \text{………..5.2}$$

2. **Wick Structure with Pores:** As seen in Equation 5.2, the capillary wick that is contained inside a heat pipe must be able to provide an enough amount of capillary force and an adequate amount of space for vapour to escape in order to achieve the required level of heat transfer performance. When calculating ΔPcap, the Young-Laplace formula may be used in the following manner:

$$P_{cap} = \frac{2\sigma cos\theta}{r_{eff}} = \frac{2\sigma}{r_{cap}} \qquad \text{……………5.3}$$

Here, recap represents the capillary radius, riff represents the effective capillary radius of the wick, □ represents the contact angle between the working fluid and the wick, and σ represents the surface tension of the working fluid used in the experiment. The effective capillary radius is equal to the distance between the wires in a mesh screen (riff = (w + dw)/2), where w is the wire spacing and dw is the wire diameter. This is the equation that describes the effective capillary radius. For channels that are located between micropillars, the capillary radius is equal to half of the channel width. The hydrophilicity and permeability of the wick that is employed should be as high as feasible; some strategies that should be taken into consideration are as follows:

(a) The formation of a micro- or nanostructured morphology on the surface via the processes of surface oxidation, chemical corrosion, and deposition.
(b) The use of nanoscale wicks, such as carbon nanotubes, in order to achieve the highest possible wicking potential.
(c) Structures like grooves and fibres, as well as pillars and grooves and meshes and fibres, are examples of hybrid structures. In order to fulfil the criteria

for sufficient capillary force and minimal vapour flow resistance, multiscale wicks are able to be used.

(d) Spiral woven meshes are made up of braided fibres, which allow for high flexibility and wicking capability. Additionally, these meshes are well suited for the production of ultrathin heat pipes.

(e) In order to ensure that the capillary pressure inside the wick is equal to twice the total pressure drop that occurs between the wick and the vapour core, it is important to choose the shortest feasible thickness of the wick. Increasing the wettability of the wick allows for a thinner wick, which in turn maximises the amount of space available for vapour. A wick's permeability is denoted by the symbol Kl. Calculating this may be done using the following formula for the screen mesh wick that is typically used:

$$K_l = \frac{d_\omega^2 \varepsilon^3}{122(1-\varepsilon)^2} \quad \text{...........5.4}$$

$$\varepsilon = 1 - \frac{1.05\pi N d_\omega}{4} \quad \text{....5.5}$$

N is the number of meshes that make up the screen mesh, whereas ε represents the porosity. For a wick with grooves, the permeability may be calculated as follows:

$$K_l = \frac{d_h^2 \varepsilon}{2(fRe_{L,h})} \quad \text{.5.6}$$

In which the hydraulic diameter is denoted by dh:

$$d_h = \frac{4A_l}{C_l} \quad \text{....5.7}$$

where Cl represents the circumference of the liquid channel that has been wetted. Determining the permeability of a nonporous structure, such as a sintered wick, may be accomplished by the following:

$$K_l = \frac{d_p^2 \varepsilon}{32} \quad \text{………5.8}$$

where the pore diameter is denoted by dp.

(f) A wick design for no pulsating flexible heat pipes that is centred on single and multiple nonporous meshes is shown here. The facilitation of a flexible wicking structure that provides a strong thermal connection to the shell material, high thermal conductivity, capillary pressure, and low permeability are some of the challenges that must be overcome. A robust partnership between surface wettability and material science will be included into future solutions. Copper chemical-plating modified polyacrylonitrile-based carbon fibre wick, flexible graphene coated polyurethane, and laser-induced graphene on polyimide substrate are some of the potential solutions that might be implemented.

3. **Design of vapor-fluid spaces:** As the thickness of the vapour chamber increases, the maximum amount of heat transmission that is permitted continues to decrease. For the purpose of maximising the effectiveness of the liquid and vapour spaces within the internal cavity through the utilisation of analytical and numerical models, rigid design optimisation is required. Optimisation of the wick and vapour domains should be the primary emphasis of the design of the space that contains the vapor-fluid component. It is recommended that intelligent geometric configurations, such as the design of blood capillaries and leaves, as well as electromagnetic channels, be examined in order to reduce the amount of pressure loss. The application of analytical and numerical modelling approaches, such as topology optimisation or metaobject evolutionary algorithms, is one way that could be taken to accomplish this purpose.

All of the components, including the design of the wick, the thickness of the outer shell, and the design of the vapor-fluid gap, are fully reliant on one another. Steam flow and condensate discharge are the means by which mass and heat can be transferred through an ultrathin heat pipe with remarkable efficiency. It is possible to utilize the Hagen-Poiseuille equation in order to determine the pressure drop (ΔPv) that is caused by the passage of vapour.

$$\Delta P_v = \frac{8\mu_v q_{\max} L_{eff}}{\pi \rho_v A_v d_h^2 h_{fg}} \quad \ldots\ldots\ldots 5.9$$

In this equation, the variable vs. represents the dynamic viscosity of the vapour, qmax represents the maximum power at the capillary limit, and Leff represents the effective length of the heat pipe. $(L_{eff} = \frac{L_e + L_c}{2} + L_a)$. ρv in this equation, dh represents the hydraulic diameter of the wick (dh = 2Al/Cl), where Al represents the cross-sectional area of the liquid channel and hfg represents the specific enthalpy of vaporization. The vapour density is denoted by Av, and the cross-sectional area of the vapour channel is denoted by Av. The permeability of the vapour space, denoted by the symbol Kv, is defined as follows:

$$K_v = \frac{t_v^2}{12} \quad \ldots\ldots\ldots\ldots 5.10$$

The thickness of the vapour space is denoted by the symbol tv. ΔPl may be determined by using Darcy's law, as stated.

$$\Delta P_l = \frac{\mu_l q_{\max} L_{eff}}{K \rho_l A_l h_{fg}} \quad \ldots .5.11$$

The density of the working fluid is denoted by the symbol ρl. In accordance with the Clausius-Clapeyron relation, the vapour pressure in the condenser section (Pcond) and the evaporator section (Pevap) may be determined using the following equation:

$$P_{cond} = P_\infty e^{\left(\frac{h_{fg}}{R_g} \frac{T_{cond} - T_\infty}{T_{cond} T_0}\right)} \quad \ldots 5.12$$

$$P_{evap} = P_{\infty} e^{\left(\frac{h_{fg}}{R_g} \frac{T_{evap} - T_{cond}}{T_{evap} T_{cond}} \right)} \quad \text{.........5.13}$$

In where P and T represent the pressure and temperature of the surrounding environment, respectively, and Rg represents the gas constant. describes the pressure decrease that occurs as a result of gravity as follows:

$$\Delta P_g = \rho_l g L_{eff} \sin \phi_{incl} \quad \text{......5.14}$$

The constant g represents the acceleration that is caused by gravity, while the angle ϕincl represents the inclination angle of the heat pipe. By combining the equations shown above and assuming that the heat pipe is oriented horizontally, the maximum power at the capillary limit may be calculated using the following formula:

$$q_{\max} = \frac{2\sigma cos\theta}{r_{eff} L_{eff}} \left[\frac{8\mu_v}{\rho_v A_v r_h^2 h_{fg}} + \frac{\mu_l}{K_l \rho_l A_l h_{fg}} \right]^{-1} \quad \text{...5.15}$$

4. **A Device that Evaporates:** The incorporation of a structural layer with a high thermal conductivity should be the primary focus of the design of an evaporator. The coupling of a flexible articulating adiabatic segment with a stiff material that has a high heat conductivity at the evaporator and condenser section is a prospective route that might be pursued. It is vital to have a good thermal connection between the shell material and the porous wick structure in the evaporator in order to eliminate as much thermal resistance as possible. Thermal vias have been shown to be an effective method for lowering through-wall thermal conductivity in a composite polymer and copper design framework. When the temperature drop through the casing is compared to the temperature drop throughout the whole system, it is possible that the temperature drop through the casing is insignificant for a sufficiently thin and low thermal conductivity casing material. The enormous latent heat and surface tension in the fat plate heat pipe prevent nucleated boiling from occurring, despite the fact that the superheat is 8 degrees Celsius. This is because of the

confining geometry of the pipe. Because of the mesh structure and the flow of the condensed work fluid, the boiling of the fluid is slowed down during the optimal performance of the nuclear boiling regime. This reduces the number of bubbles that are detached from the fluid, the size of the bubbles is restricted to equivalent to that of a single mesh pore, and the movement of fat bubbles in the vapour space occurs at regular intervals.

5. **The Condenser:** The natural convection that comes from the outside shell is the limiting factor for mobile electronics. It is necessary to accomplish excellent thermal dispersion in order to reduce the number of hot spots on these devices. In this regard, the ratio of liquid to vapour space is relevant.

   (a) It has been proposed that in order to enhance the condenser uniformity, it would be beneficial to: (i) raise the thermal conduction resistance in the centre area of the condenser, which includes the sidewall and the wick, and (ii) lower the thermal resistance of the condensation core. In order to accomplish both of these effects, it is possible to remove portions of the wicking layer that is located in the central condenser area. Because of this, the effective thermal conductivity of the wick would be reduced in this particular location. This would be accomplished by replacing the porous sintered copper with a layer of water, which has a lower conductivity (water). The effective permeability of the grooves and the sintered copper wick would rise, which would make it possible to use a thinner wick throughout the whole inner wall of the condenser side while still retaining the same pressure drop. This would result in a reduction in the vapor-core thermal resistance.

   (b) The performance of the vapour chamber may be significantly enhanced by using a combination of a superhydrophobic condenser and a super hydrophilic evaporator. There are various benefits associated with this superhydrophobic surface: (i) it has the ability to successfully avoid the creation of a liquid film, which has the potential to significantly reduce the efficiency of the condensation process; (ii) the vapour may readily condense into a droplet on this surface, which then falls straight into the wick for further condensation. This has the potential to reduce the length of the water feeding channel and prevent the centre wick from becoming dry.

6. **Making use of Fuid:** During the time when the working medium was sufficient to completely submerge the structure of the capillary mesh, a reasonable approximation of the optimal filling ratio was achieved. When designing vapour chambers of this kind, the selection of a functional fuid is of the utmost importance. The thermophysical parameters of the fluid have a significant influence on the functioning of the vapour chamber, which operates according to a two-phase thermodynamic cycle as its underlying principle of operation. Based on the findings of a distinct figure of merit has been established for the vapour (Mv) and liquid (Ml) phases in order to determine the optimal working fluid for a certain shape.

$$M_l = \frac{\sigma \rho_l h_{fg}}{\mu_l} \quad ....5.16$$

If this figure of merit has a greater value, it implies that the vapour chamber is capable of operating at a higher power before it reaches the capillary limit.

$$M_v = \frac{P_v \rho_v h_{fg}^2}{\mu_v R_g T_v^2} \quad .....5.17$$

Pv represents the pressure of the vapour, while Tv represents the temperature of the vapour. An increase in the value of this figure of merit is associated with a decrease in the thermal resistance that is present in the vapour core. The use of a working fluid with a greater Ml is necessary in order to reduce the needed wick thickness and, as a result, maximize the accessible vapour space. This is necessary in order to accomplish maximal heat transfer at minimum thicknesses. Both of these metrics should be maximized in the design of ultrathin heat pipes in order to achieve optimal performance. The selection of fuid is determined by the temperature dependency of thermophysical qualities that are specific to each fuid; it is essential that care be used in order to guarantee an acceptable vapour pressure at which the structural integrity of the vapour chamber is not compromised. Therefore, the implementation of ultrathin flexible heat pipe technology is a difficult issue that involves evaporation, condensation, two-phase flow, a flexible porous wicking structure, and the design of a flexible impermeable shell material, all of which are required for an ultrathin package size.

The creation of a shell material that is adequate in terms of flexibility, thermal conductivity, impermeability, and the ability to thermally attach to the porous wicking structure that has been established is one of the most significant obstacles that must be solved. The multiscale flexible porous wicking structure is designed to provide high capillary pressure, permeability, and flexion. Additionally, the design of the fluid-vapor space is optimized to allow for efficient two-phase flow inside the flexible heat pipe. Because of the fundamental connections that exist between these three components, there is a promising prospect for further study in the future. The study that is being suggested is of a multidisciplinary character, and it will include experimental heat transfer, fluid visualization, analytical and numerical modelling, and material sciences .

5.2.4 Constituents of the Thermal Interface

In the fields of computer, information, communication, energy harvesting, energy storage, and lighting technologies, thermal interface materials (TIMs) are an essential component in the process of thermally connecting the different components of electronic assemblies. From a microscopic perspective, the surface of any engineering component that contains electronic components is rough. There are just a few isolated areas where real contact takes place when two such rough surfaces come into touch with one another during the process of electrical assembly (Fig. 7.6). This represents around 3% of the apparent contact area. It is usual for a thermal interface material to take the shape of paste, glue, solder, or a resilient sheet.

Its primary function is to fill the space that exists between two surfaces that are next to one another. The heat transfer of this complex joint is determined by a number of factors, including contact pressure, the number, size, and shape of contact spots and voids, the types of fluid that are present in voids, the pressure of fluid in voids, the hardness and fatness of contact surfaces, the modules of elasticity of contact surfaces, the cleanliness of the surface, and the property of the thermal interface material that is used in the joint. It is necessary to have a comprehensive grasp of material and surface sciences, heat transfer at submicron scales, and the manufacturing techniques that are used in the assembly of microelectronics and other target applications in order to design thermal interface materials.

The efficiency and dependability of electronic devices are directly proportional to the selection and use of thermal interface materials in the appropriate manner. The following are some of the criteria that were chosen for technical selection: (a) low thermal impedance; (b) high thermal conductivity; (c) good conformability with minimum thermal expansion stress when joining two contact surfaces; (d) mechanically sound joint, great adhesion, and high reliability with long service life; (e) lack of sensitivity to changes in temperature and moisture; and (e) good manufacturability with ease of handling, application, and use. Instead of investing in a complex cooling system, it is often more beneficial to make an investment in the interface material. In the absence of adequate thermal contact between the components that are connecting, the use of costly materials that are thermally conducting for those components is a waste.

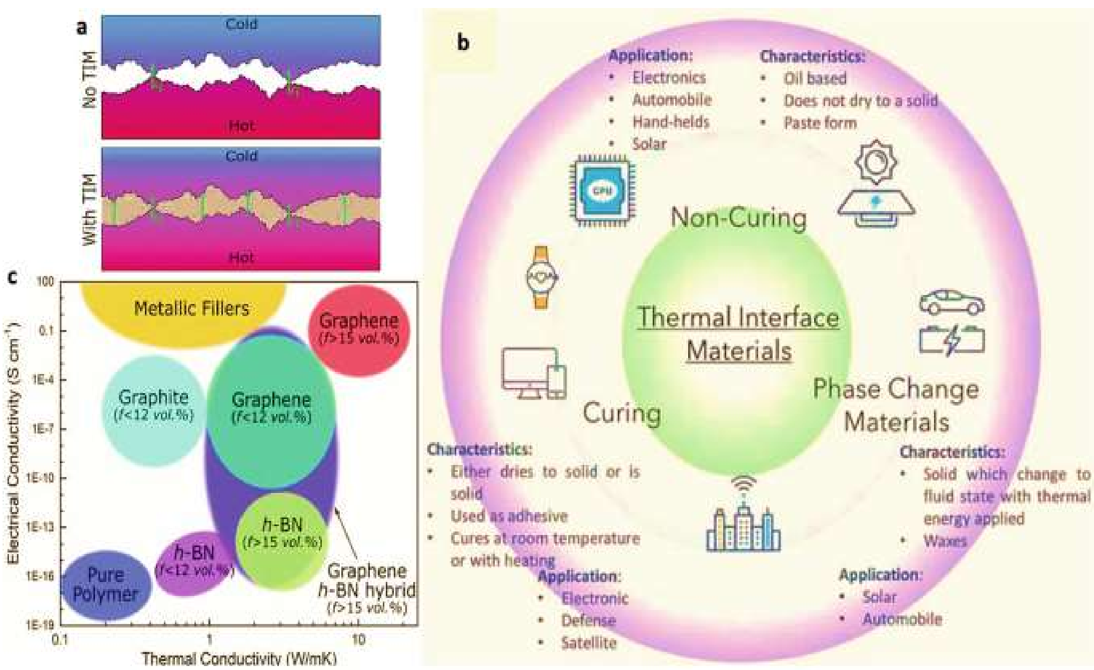


Fig. 5.6 TIM and the actual area of contact of a joint made of two rough surfaces in the real application situation are shown in the sketch

Source: Extracted from Lewis et al. (2021) with IOP Publishing permission): Principally, (a) an air-gapped, heightened physical interface without (top) and with (bottom) TIMs; (b) applications for various TIM kinds; (c) typical composite characteristics for various TIM types with randomly oriented fillers

Thermal interface materials may be classified into a variety of categories, including thermal greases, gels and pastes, adhesives, pads, tapes and films, phase transition materials, graphite, solders, liquid metals, and compressible metals. Each of these categories has a unique set of characteristics. In spite of this, developments in all of these and other fresh emerging technologies are blurring the line, making it possible to build next-generation thermal interface materials with new combinations of attributes that exceed the conventional bounds of performance. In the beginning, thermal greases were used to fill the spaces that existed between power devices and heat sinks. These greases are normally silicones that have been loaded with thermally conductive fillers.

The materials that make up thermal gels have a low modulus and resemble pastes. They are also mildly cross-linked. Resins that are chemically cross linkable, such as epoxy resin, silicone resin, or other types of resins, may be used to make them. Because of its capacity to adhere to surfaces while exhibiting less material pump-out, thermal gels operate similarly to grease in terms of their performance. Thermal gap filling pads are gel-like compounds that can be inserted and squeezed to shape with the components with thicknesses ranging from 0.05 mm to 0.5 mm or more. These pads may be used to fill thermal gaps. In most cases, they are manufactured by molding electrically conductive fillers that are reinforced with silicone elastomer. Weaved glass, metal foils, polymer films, and nanoscale conductive fillers are all examples of materials that may be used as reinforcements for thermal pads. Precut thermal pads are often available in a variety of sizes to accommodate components of varying dimensions. thermal greases are less precise, simpler to use, and cleaner than these other types of grease materials.

5.2.5 Materials for Thermal Insulation

Materials that are used for thermal insulation fulfil one or more of the following goals, with the explicit intention of minimizing heat conduction, convection, radiation, or all three in order to reduce or impede the passage of heat: Conserving energy involves reducing heat gain or loss; (2) regulating surface temperatures for the safety and comfort of personnel; (3) facilitating the condensation of water and vapour from a process; (4) improving the efficiency with which plumbing, steam, heating, ventilation, and cooling systems, electrical systems, and mechanical systems, including those found in commercial and industrial installations, operate; and (5) assisting related products in meeting standard criteria for energy storage and temperature regulation, including textiles and clothing, buildings, vehicles, vessels, and other similar products. Low-density foams and fibrous materials that restrict convective and conductive heat transfer are the most prevalent forms of thermal insulation materials. There are many different types of thermal insulation materials since there are many distinct types. Depending on the temperature environment, appropriate materials may include carbons, glasses, ceramics, or polymers that are suitable for the application.

Common examples of ceramic thermal insulation materials include natural raw materials like diatomaceous earth and pearlite, as well as heated bubble glass, which is produced by mixing glass powder with a blowing agent. Pearlite is also a kind of thermal insulation material. In comparison to foamed polystyrene and corks, they are less likely to get deformed and have a higher tolerance to rising temperatures. In order to screen and protect electronic components, a wide range of advanced insulating materials, such as aerogels, have been used in applications ranging from the military to the automobile industry to the aerospace industry to 5G device applications. These materials are often custom-designed in order to safeguard control equipment. It is expected that as more electronics are put into hostile environments such as jet engines, all-electric automobiles, and Aeroplan's, there will be a need for a variety of capabilities and even higher performance thermal insulation. This demand may necessitate the development of innovative materials and processes. The combination of reflective insulation and heat radiators, for instance, is a vital component for space applications.

The high costs of making aerogels and the related health hazards have been the primary challenges that have prevented their widespread use. Despite this, the materials have witnessed a significant increase in their utilization in thermal insulation applications, such as pipelines and structures, as well as in the thermal management of automobiles "under the hood." It is possible that this category of super insulators may become increasingly desirable for use in thermal shielding in electronic devices as new materials and composites are discovered and prices continue to decrease. Specifically, hybrid structures may be able to reap the benefits of the application of nanoparticles and the coatings that they get. Because of its convoluted thermal routes and thus poor heat transmission, an oxide nanocomposite material may be applied as a liquid coating. Its total conductivity is 0.017 W/m-K, which makes it practical to do so.

5.2.6 Heating Metamaterials

Due to the fact that their artificial structures have been carefully constructed, thermal metamaterials possess remarkable qualities in terms of heat transport that surpass those of naturally occurring materials. It is now feasible to control heat flow at whim thanks to the concept of thermal metamaterials, which has entirely subverted the design of thermal functional devices. Some examples of these devices are thermal cloaks, concentrators, rotators/inverters, and camouflages, as illustrated in Figure 5.7. These advancements indicate new potential to direct heat transport in complicated systems as well as novel packaging options connected to the thermal management of electronic devices like computers and smartphones. The trends in electronics packaging towards more power, higher density, and 2.5D/3D integration are making thermal management even more difficult . These elements are essential since they are contributing to the difficulty of thermal management.

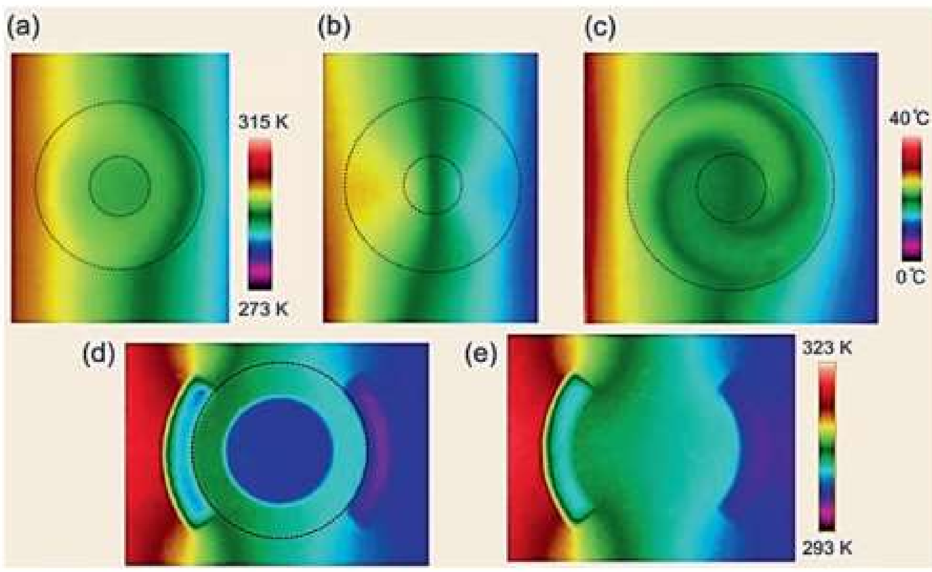


Fig. 5.7 Display of several thermal metamaterial devices experimentally

Source: Taken from Han et al., 2014; John Wiley and Sons permission granted): (A) Shield. Concentrator (b). (c) An inverter/rotator. (d) Camouflage: two insulating segments seem to be a single conducting cylinder. (e) Equivalent temperature profile of camouflage illusion, in which there is no conducting cylinder and two insulating segments in reality

While conventional cooling solutions based on large thermal-conductivity materials as well as heat pipes and heat exchangers may dissipate the heat from a source to a sink in a uniform manner, thermal metamaterials could help dissipate the heat in a deterministic manner and avoid thermal crosstalk and local hot spots, and associated packaging designs may benefit from the new capabilities of controlling heat transfer paths, underlying the scope and aim of thermal metamaterial design. (1) Anisotropic heat spreaders and diffusers, (2) heat cloaking and isolation, and (3) heat directing and bending are all examples of possible methods to thermal metamaterials. are some instances of thermal cloaks that were constructed employing multilayered composites. Figure 7.8 illustrates some of these examples. When it comes to electronic packaging, conventional thermal management systems are often limited to a single design scale, such as the device, the package, or the system. Many of the previous approaches to problem solving function independently, and they may eventually come into conflict with one another. At each of these sizes, thermal metamaterial solutions might be used for the purpose of thermal efficiency control.

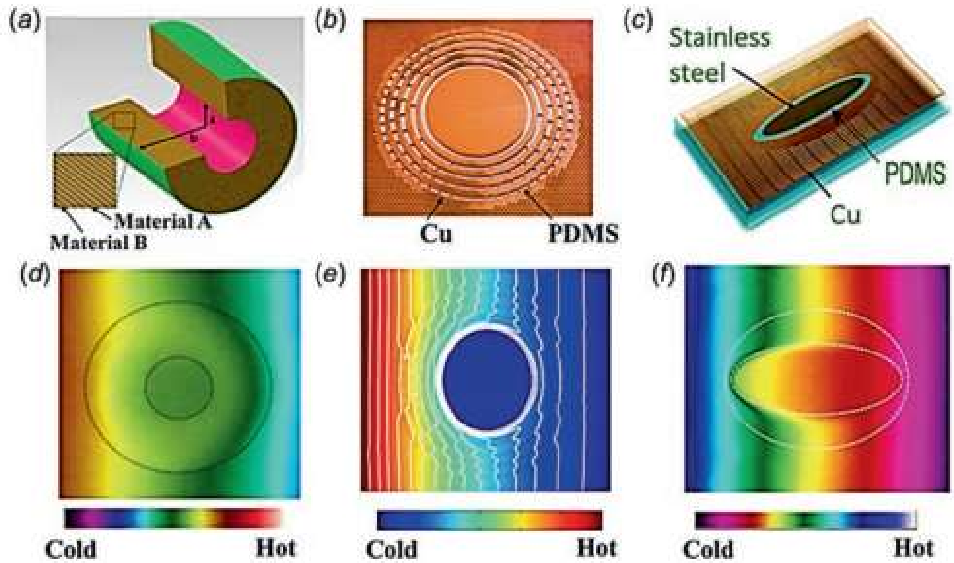


Figure 5.8: Multilayered composites used to create thermal cloaks

Source: Taken from Han et al., 2018; John Wiley and Sons permission granted): (a) The silicone elastomers (material B) and latex rubber film (material A) used in the steady-state thermal cloak, with agar-water serving as the backdrop material. 50.8 cm is equal to a and 52.7 cm to b. (b) The temporary thermal cloak made of PDMS and Cu. (c) The omnidirectional elliptical heat cloak, which is made up of an insulating layer made of PDMS, a Cu shell, and a stainless steel object. (d) The temperature profile of (a) that was measured. (e) The temperature profile of (b) at t = 5120 s as measured. (f) At t 55 minutes, the temperature profile of (c) was measured.

In spite of the fact that a thermal metamaterial design could be an effective heat transfer solution in and of itself, further research and development into related and needed interdisciplinary electronic packaging technologies might be the key to achieving synergy between thermal management strategies on numerous levels. Through the use of the metamaterial idea, it is possible that in the future, research on thermal management would naturally incorporate efforts from a variety of technological fields, such as ECAD design tools, base material systems, and active devices.

5.3 THERMODYNAMIC CONTROL OF 5G BASE STATION ANTENNA WAVELENGTHS

Traditional active electronically scanned arrays (AESA) and planar AESA are the two kinds of active electronically scanned arrays (AESA) that are often included in antenna arrays for 5G base stations. As seen in Figure 5.9a, the typical structure of AESA is shown with specific examples.

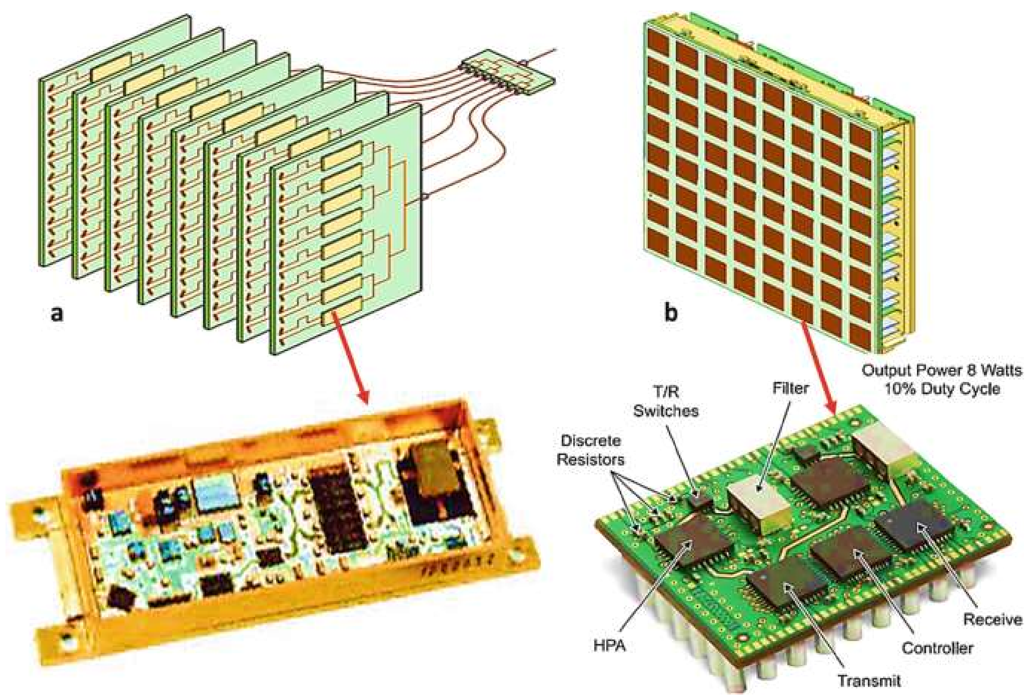


Fig. 5.9 An example of how an AESA is built

Source: Reproduced from Herd and Conway (2016) with IEEE permission): The conventional method and its T/R module, and the planar method and its T/R module

A number of circuit cards are positioned in a manner that is orthogonal to the antenna array in examples of such structures. A row of antenna elements that each have separate transmit/receive (T/R) modules installed on them are fed by each card via the antenna elements. This method has the benefit of providing a wide surface area for the thermal load as well as the T/R modules, which is a significant advantage. One of the most significant drawbacks is that they need a significant quantity of RF boards and cabling in order to route signals. In addition, if the T/R modules are of a big size, it could be challenging to create a design that is compact. An example of a planar AESA is shown in Figure 7.9b. Integration of the antenna components and RF beamformers is accomplished by the use of a single multilayer RF board in this configuration.

The addition of additional functionality into a single integrated circuit (IC) is a challenge that might be considered as a potential solution. Increasing the distance between the T/R modules in order to provide passive heat removal is a more basic method that may be implemented . This can be accomplished by creating a sparsity in the radiating components.

5.3.1 Traditional AESA Cooling

The temperature distribution of a 16x8 classic AESA array is shown in Figure 5.10a.

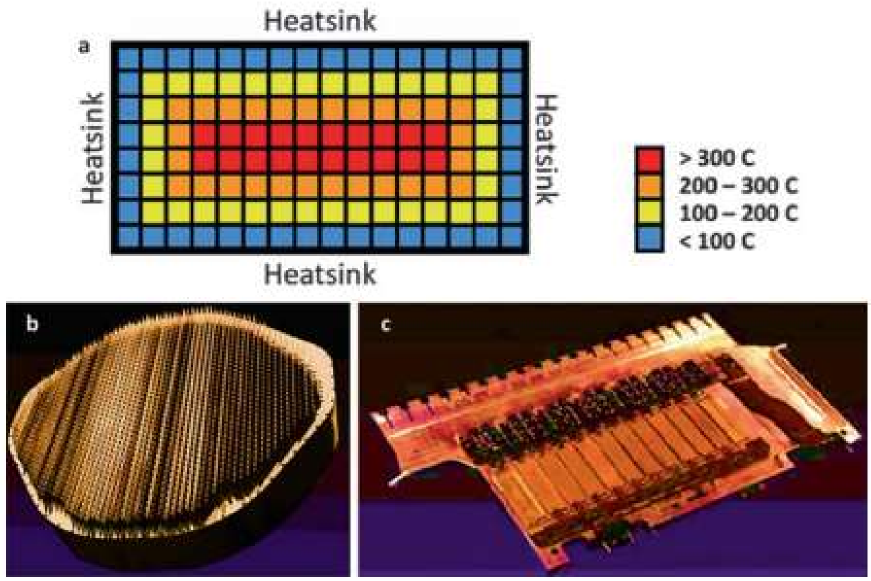


Fig. 5.10 Conventional AESA cooling

Source: Taken from Wilson (2003); acknowledgement given to Electronics Cooling; McCune (2017) with IEEE permission): (a) A conventional 8 x 16 AESA's temperature distribution; (b) an array aperture; and (c) a row of microwave modules on a slat of an aerial phased array radar antenna

It is assumed that an ideal heat sink with a temperature of 30 degrees Celsius is maintained around the array. It demonstrates quite clearly that the heat is unable to escape from the center of the array, which leads to an elevation in temperature that is not acceptable. For this reason, it is essential for such arrays to include the use of cool plates or fluid channels in order to transport the heat that is created towards the margins. An example of an X-band airborne phased array antenna is shown here in Figure 7.10b, c. This figure depicts a number of microwave modules that are equipped with transmit and receive circuitry that is comprised of GaAs power amplifiers and are mounted to liquid cooled slats.

5.3.2 Refrigeration in Planar AESAs

Because the beamforming and amplification chips are positioned differently in planar AESAs compared to their classic counterparts, thermal management in these devices is carried out differently.

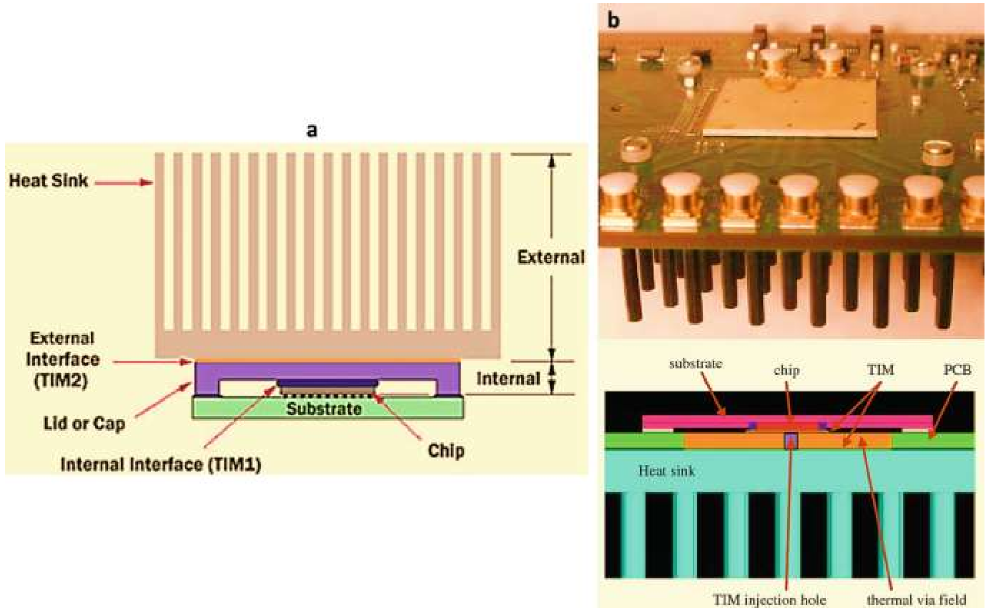


Fig. 5.11 In planar AESAs, cooling

Source: Taken from Lasance and Simons (2005), with permission from IEEE, crediting Electronics Cooling; Kam et al. (2011)): (AiP package installed on evaluation board with heat sink; (b) chip package affixed to heat sink

Due to the fact that the chips are now situated at the opposite site of the substrate in comparison to the radiating parts, it is possible to connect external heat sinks to the chips without compromising the radiation performance. In order to improve thermal conduction and heat spreading from the chip to the lid and heat sink, thermal interface materials (TIMs) or certain fat plates with high thermal conductivities are used. This can be seen in Figure 5.11a. A low-cost 60 GHz antenna-in-package (AiP) phased array is shown in Fig. 5.11b, which depicts a prototype design that includes a package that is installed on an evaluation board that also has a heat sink.

5.3.3 Cooling of Antenna Arrays at Millimeter Waves

Active cooling systems consisting of either forced air or liquid cooling are the predominant types of cooling strategies that are now in use for millimeter-wave antennas. In spite of the fact that they are efficient at removing heat, active cooling systems that include fans or pumps need the use of power. Additionally, these systems make the system more complicated and difficult to maintain. On the other hand, passive thermal management is a solution that is both more cost-effective and more efficient in terms of energy use.

However, since the heat is only dissipated passively by natural convection by using simply heat spreaders or heat sinks, it is not simple to attain thermal performance that is comparable to that of the active equivalents. This is the case unless there is a sufficient surface area that is in touch with the outside environment. CPU coolers that do not have fans may possibly be utilized in planar AESA, which allows the heat sink to be connected to the processors on the opposite side of the radiating components.

This would allow for totally passive cooling to be achieved in 5G base station arrays. Heat sink modules for central processing unit coolers are able to produce comparable results in heat removal when compared to active systems, despite the fact that they are cumbersome and heavy. In Figure 5.12, which depicts the CPU processor package of the Fujitsu Prime power 2500, there is an illustration of a sample CPU cooler. The heat pipes, which are completely passive and have low thermal resistance and efficient heat transfer, are shown in Figure 5.13 to be able to absorb heat from the chips and then transport it to the fn stack, which is cooled by natural convection. Additionally, a sample CPU cooler fn stack from FSP Windle is shown in the figure.

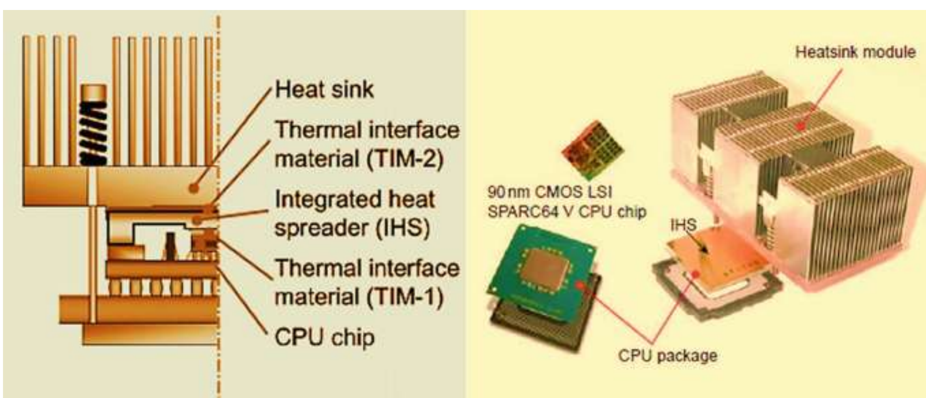


Fig. 5.12 CPU processor package of Fujitsu Prime power 2500

Source: Adapted from Wei (2007), credit: https://www.fujitsu.com/global/documents/about/resources/publications/fstj/archives/vol43-1/ paper14.pdf)

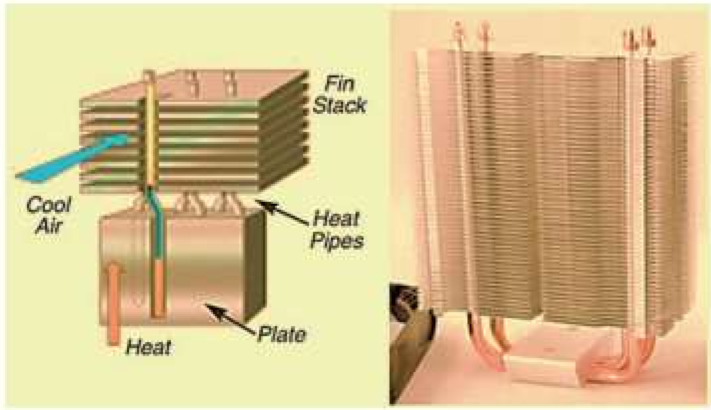


Fig. 5.13 Heat pipes and a remote fin stack are used to cool the CPU

Source: Adapted from Aslan et al., 2018 with open access (IEEE)

There are also several methods that concentrate on cooling the antenna elements rather than the array (or system) level. These methods are also worth considering. Creating a heat spreader that connects the transistor to the radiating patch is one of the ideas that might be considered. An antenna of this kind is capable of delivering radiation and thermal performance that are both within the desired range simultaneously. As a result, the use of heat spreaders in planar AESA for the purpose of cooling augmentation might be a viable solution. These heat spreaders would be placed between the chips, the ground plane, and the radiating components. Various other strategies place a greater emphasis on fractal or fanned element architectures for dual functioning, which includes both thermal and electromagnetic properties despite the fact that their structure may be optimized for a particular radiation pattern that is wanted, they are not ideal for flexible beam formation, which is a need required by AESA.

In addition, the following strategies have the potential to improve the thermal management performance of the antenna arrays :

1. Finless central processing unit coolers that have reasonably large heat sinks are able to provide enough cooling at the base station antennas, particularly in situations where the amount of power that is dissipated by each chip is quite modest.

2. In the event where the transceiver integrated circuits are very inefficient and the heat sink has a lesser capacity for heat transfer, layout sparsity may be able to give a significant decrease in the maximum junction temperature of the array.
3. Increasing the interelement distance in regular square layouts in order to generate sparsity leads to the production of grating lobes, which may result in extremely significant interference among the simultaneous co-frequency users. This is despite the fact that it is possible to achieve further cooling.
4. It is possible to adopt alternative sparse array design that are based on linear or planar irregular arrays, spiral arrays, thinning arrays, or circular ring arrays in order to get higher electromagnetic performance (in comparison to normal square arrays) while still keeping the capacity for cooling.
5. The average distance between the components is the most important factor in lowering the temperature for a variety of sparse topologies. The sunflower topology is the best one from this perspective since the interelement spacing in this topology may be the biggest among the sparsest topologies while still retaining low sidelobe levels. This makes it the best topology.

In addition, the use of heat spreaders situated in the center of the patches allows for the creation of a conduction channel between the chips, the ground plane, and the radiators, which results in extra cooling without compromising the radiation performance. When power dissipation per chip grows and the heat transfer coefficient (h.t.c.) of the heat sink that is linked to the chips decreases, spreaders are able to cool the chips more effectively.

5.4 THERMODYNAMIC CONTROL OF 5G EDGE COMPUTERS

Computing at the edge, also known as mobile edge computing (MEC), is an alternative to the conventional cloud computing format. In MEC, the majority of the data processing is moved closer to the consumer, or to the "edge" of the network, rather than being handled at a central point since that is where the majority of the processing takes place. The ultimate objective of this shift is not necessary to completely replace cloud computing; rather, the objective is to enhance the concept of cloud computing by combining its technology with the geographical dispersion of processing centers in order to address a number of the constraints that are associated with conventional cloud computing. As can be seen in Figure 5.14, edge computing is a method that moves intellect, processing power, and communication capabilities from the heart of a network to the edge of the network, as well as from an edge gateway or appliance straight into devices. Better user experiences and faster reaction times are among the benefits that will result from this.

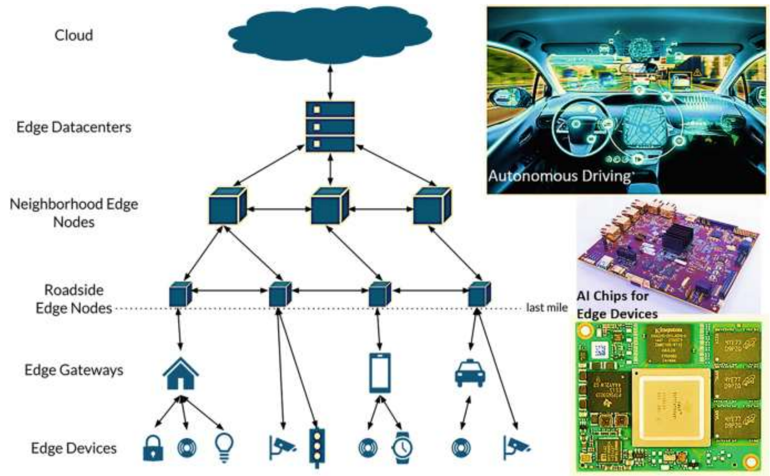


Fig. 5.14 An architectural plan for edge computing that reduces latency and processing power as it increases

Source: Adapted from Scheuermann and Bischoff (2019); credit: https://www.inovex.de/de/blog/edge-computing-introduction/)

According to O'Day and Quesnel 2020, the following layers are incorporated in the approach to the architecture of edge computing:

- **(a) Cloud:** In this layer, the amount of computing power and storage space is almost unbounded. It is possible for there to be extremely significant latencies and costs associated with the transfer of data to this tier. For an application that uses edge computing, the cloud may offer storage for the long term and manage the lower levels that are immediately accessible.
- **(b) Edge Node:** The term "downstream" refers to the location of these nodes, which are situated before the last mile of the network. In most cases, edge nodes are equipped with a significant amount of computational power and are able to route network traffic among themselves. A wide variety of devices, including base stations, routers, switches, and even small-scale data centers, are included.
- **(c) Edge Gateway:** The edge gateways are comparable to the edge nodes; except they have a lower level of power. They are able to communicate via the most widely used protocols and do calculations that do not need specialized

hardware, such as graphics processing units (GPUs). In order to translate for devices on lower levels, it is common practice to employ devices on this layer. Another option is that they may serve as a platform for lower-level devices, such as mobile phones, automobiles, and a variety of sensing systems, such as cameras and motion detectors.

(d) Edge Devices: In this tier, there are devices that are quite tiny and have very little resources. Single sensors and embedded systems are two examples of such approaches. These devices are often manufactured with the sole intention of doing a particular kind of calculation, and their capacity for communication is sometimes restricted. Smart watches, traffic lights, and environmental sensors are examples of the kinds of devices that may be found on this layer.

When time-to-result has to be reduced as much as possible, such as in smart automobiles, edge computing is becoming an essential need. Bandwidth costs and latency make it more efficient to process data near its source. This is particularly true in complicated systems such as smart and autonomous automobiles, which produce gigabytes of telemetry data. It is possible to achieve a more efficient and user-friendly flow of data traffic using MEC. On account of the fact that MEC is the most important factor in determining the success of 5G technology, this will become vital as 5G capabilities continue to expand. Despite the fact that edge computing requires fewer racks, the data nevertheless need essential cooling protection. There is a possibility that the availability of cooling resources will vary for edge computers that are situated in distant places. The various air cooling options are shown in Figure 5.15.

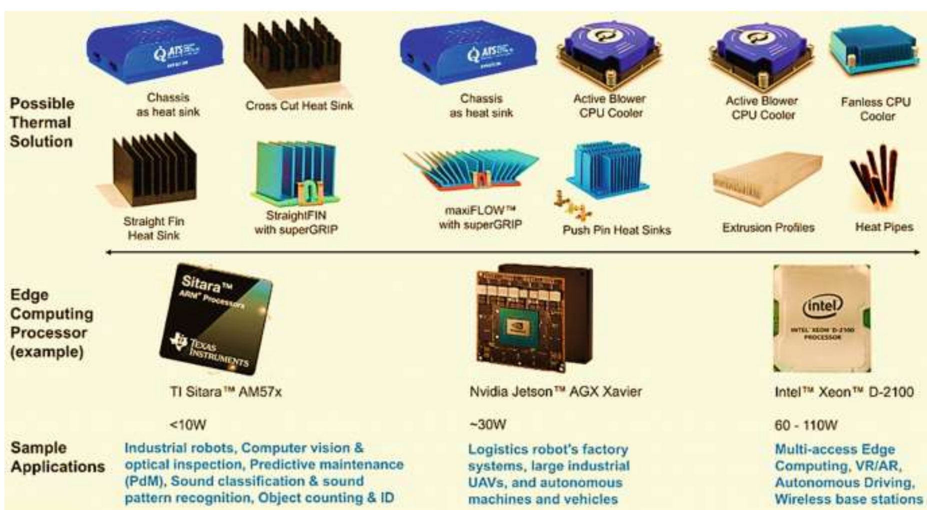


Fig. 5.15 Edge computing Air Cooling Options

Source: Adapted from O'Day and Quesnel (2020); credit: https://www.qats.com/cms/2020/01/14/edge-computing-and-thermal-management/)

As the liquid glides over the electronic components, it removes heat and evaporates, rising through the system before condensing and dropping back over the components. The result is a closed-loop, passive system that, in comparison to air-cooled systems, results in an average reduction of 22.8% in the amount of money spent on energy. In light of the fact that MEC has emerged as a more effective method of data processing and transmission, the cooling solutions that are responsible for supporting its associated technologies need to be similarly effective in lowering energy costs. As opposed to in-row air cooling, passive two-phase cooling for electronics is anticipated to be a significant increase in terms of the efficiency with which the system operates and the amount of energy that is used. Air cooling is being increasingly replaced by passive two-phase and immersion cooling systems. In truth, air cooling is being steadily replaced. These energy savings will soon pile up, especially considering that the industry is moving towards a future that will be reliant on an infinite number of edge data centers. Therefore, two-phase cooling is the most important factor in determining whether or not MEC will be a viable choice in the future.

5.5 HIGH-END MOBILE DEVICES' 5G MILLIMETRE WAVE COMPATIBLE COVERS

In order to better accommodate the ever-increasing demand for wireless bandwidth, 5G connection will make use of millimeter wave technology, which operates at frequencies higher than 20 GHz. This will allow for the allocation of more channels. mm-wave frequencies have shorter wavelengths, which results in antennas that are smaller. These antennas may be built in arrays that are tiny enough to be fitted in end-user devices like as mobile phones. Beam steering of such tiny arrays is a major enabler of the 5G vision because it enables the device to pick the optimal direction to establish connection with another device or base station while simultaneously improving the link budget by making a more targeted utilization of the radiated power. This is a fundamental enabler of the 5G vision.

The mm-wave antenna module may be built in a package that incorporates both the radio frequency front end and the tiny antenna array. This merges both of these components. This strategy, known as antenna-in-package design (AiP), helps to reduce the large losses that occur in transmission lines while operating at high frequencies. AiP modules are capable of being manufactured in large quantities and, in theory, it is possible to put them in any device, including mobile phones.

It is impossible to overlook the fact that the enclosure that is present in the majority of consumer electrical gadgets is excessively thick in the millimeter frequency band. The transmission characteristic of a dielectric slab is shown in Figure 8.1 as a function of the thickness of the slab in wavelengths. Assuming that the back of a contemporary mobile phone is constructed of glass with a permittivity of 6.84, for instance, at a frequency of 28 GHz, the optimal thickness for maximum transmission would be 2.05 millimeters. This would be nearly half a wavelength, depending on the scanning angle. In the event that the overall thickness of the gadget is roughly 8.0 millimeters, this optimal design thickness would undoubtedly be too cumbersome to be of any use in the real world. When it comes to the cover of a mobile device, then, having an ideal design is very necessary.

5.5.1 Design of Dielectric Cover

Radames that are thin and Radames that are thick are the two categories that may be used to classify the dielectric cover. Dielectric slabs that are not thicker than ten percent of the wavelength of interest are what distinguish thin Radames from other types of Radames. It is necessary for the radome of the glass-backed phone to have a maximum thickness of 0.65 millimeters in order for it to be considered thin. If the cover is too thin, it could not have the necessary mechanical strength to be utilized in the phone. On the other hand, thick Radames are built in such a way that the dielectric slab has multiples of half of the wavelength that is of interest. It is thus possible to ascertain the optimal thickness h of the thick Radames by using the following formula:

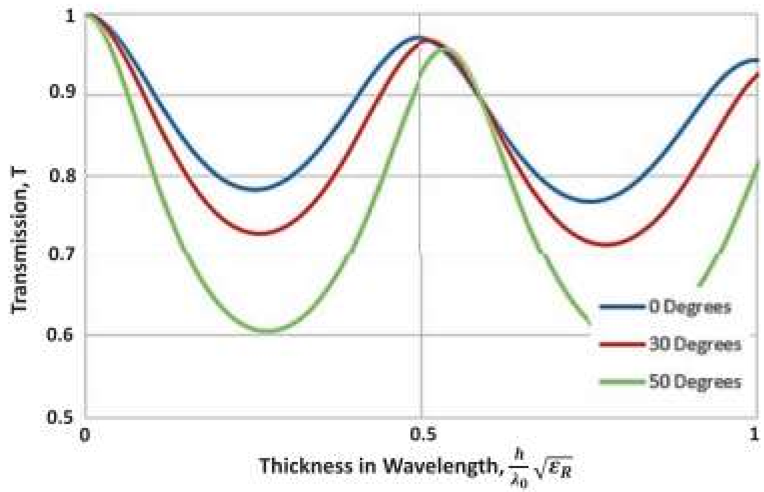


Fig. 5.16 Transmission characteristic of a dielectric slab as a function of slab thickness normalized by the wavelength inside the material, with thickness, h, and permittivity at various angles of incidence

Source: 2019 by Enjiu and Rütschlin; Dassault Systems/the 3DEXPERIENCE Company is credited.

$$h = \frac{n\lambda_0}{2}\frac{1}{\sqrt{\varepsilon_R - \sin^2\theta_i}}; n = 1,2,3,\ldots \quad \text{.....5.18}$$

Where n is the order of the radome and θ_i is the angle of incidence. In general, festered thick Radames are preferred. Equation 5.18 can also be rewritten to find the optimum εR values if the thickness h is given:

$$\varepsilon_R = \frac{\lambda_0^2}{4h^2} + \sin^2\theta_i \quad \text{.....5.19}$$

Accordingly, for the cover material that has a constant permittivity, the optimal thickness dependence on the incidence angle is reduced when the permittivity is higher.

On the other hand, for the cover material that has a constant thickness, the needed permittivity is increased when the design thickness is reduced. For the purpose of providing a clearer visual representation of these findings, Figure 5.17 illustrates the thickness and permittivity for optimal transmission qualities, taking into account three distinct materials and three distinct design thicknesses, respectively .

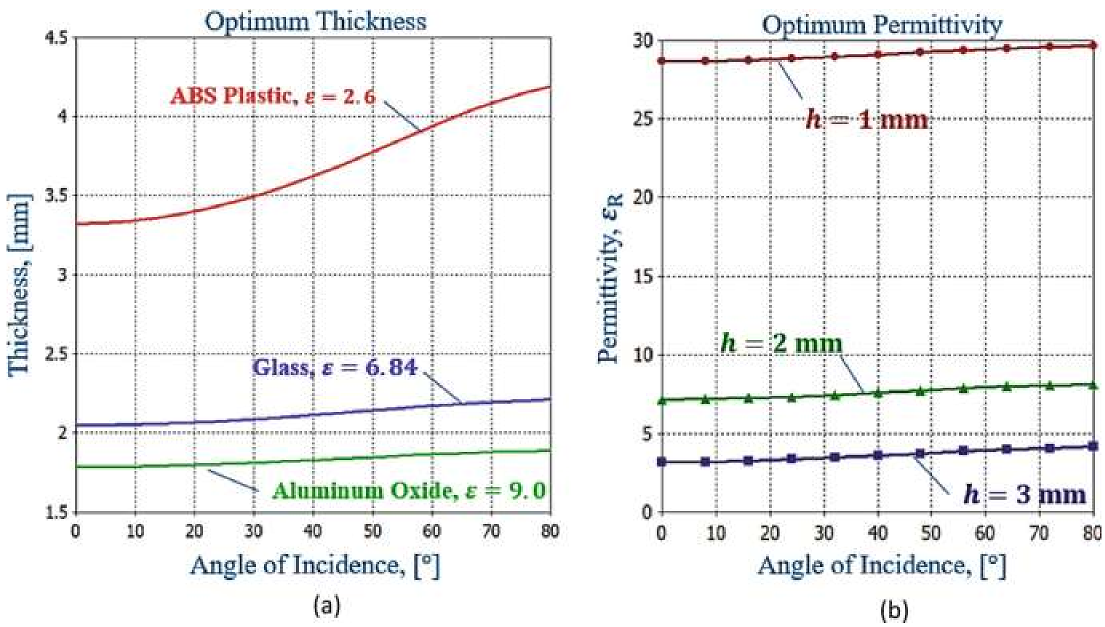


Fig. 5.17 The selection of thickness and permittivity for optimum transmission properties given three different materials and three different design thickness

Source: Adapted from Enjiu and Rütschlin 2019, credit: Dassault Systems/the 3DEXPERIENCE Company): (a) Optimum transmission thickness as a function of the incident angle for ABS plastic (red curve), glass (blue curve) and aluminum oxide ceramic (green curve). (b) Optimum transmission material permittivity as a function of the incident angle for a material with 1 mm (dark red), 2 mm (dark green) and 3 mm (dark blue) of thickness

5.5.2 Dielectric Slots are Inserted into a Metallic Cover Design

It is feasible that ceramic monolithic radome designs might provide a solution; however, the use of these designs could be hampered by mechanical limitations. Back coverings made of metal were an extremely common feature on high-end mobile phones of the previous generation. Through the use of dielectric slots that functioned as electromagnetic windows inside the enclosure, wireless communication was effectively facilitated. Additionally, certain parts of the metallic housing may be separated by the slots, which would make it possible for the enclosure to be used as an integral component of the antenna design. At frequencies in the millimeter-wave range, the electromagnetic window may be realized by the use of a frequency selective surface (FSS). In a certain frequency band and frequency range, the FSS is a two-dimensional periodic structure that may be built to be transparent in that particular space. As may be seen in Figure 5.17a and b, the FSS loaded rear cover is intended to be used with an enclosure that has a thickness of 1 millimeter. Sandwiched between two dielectric layers that have a relative permittivity of 9, the FSS layer has a thickness of 0.1 millimeters and is contained inside the sandwich.

Since antenna matching and beam steering criteria can be satisfied even for tiny gaps between the enclosure and the antenna module, the metal cover loaded with an FSS is a promising option, at least in terms of electromagnetic performance. This is because small gaps between the enclosure and the antenna module may be accommodated. The intricacy of the design and production processes is a price that must be paid for this performance.

5.5.3 Considering Integration Design

When it comes to data connections between base stations and mobile devices, 5G networks provide dependable and high-speed connections. BEAMFORM is used by base stations that are equipped with antenna arrays that have a large number of elements and operate at high millimeter wave frequencies (for example, 28 GHz). This enables base stations to provide effective and targeted communication with mobile phone handsets. Because of the very tiny size of antennas at these high frequencies, it is possible to employ several small chip-integrated arrays in each phone. This is made possible by the fact that the handsets themselves are quite small. On the other hand, the incorporation of such antenna arrays into mobile phone handsets is difficult due to the limits of space and material. Below the rear cover of the device, which, in the case of high-end phones, may be made of metal or glass, there is the possibility of accommodating arrays that are based on chips.

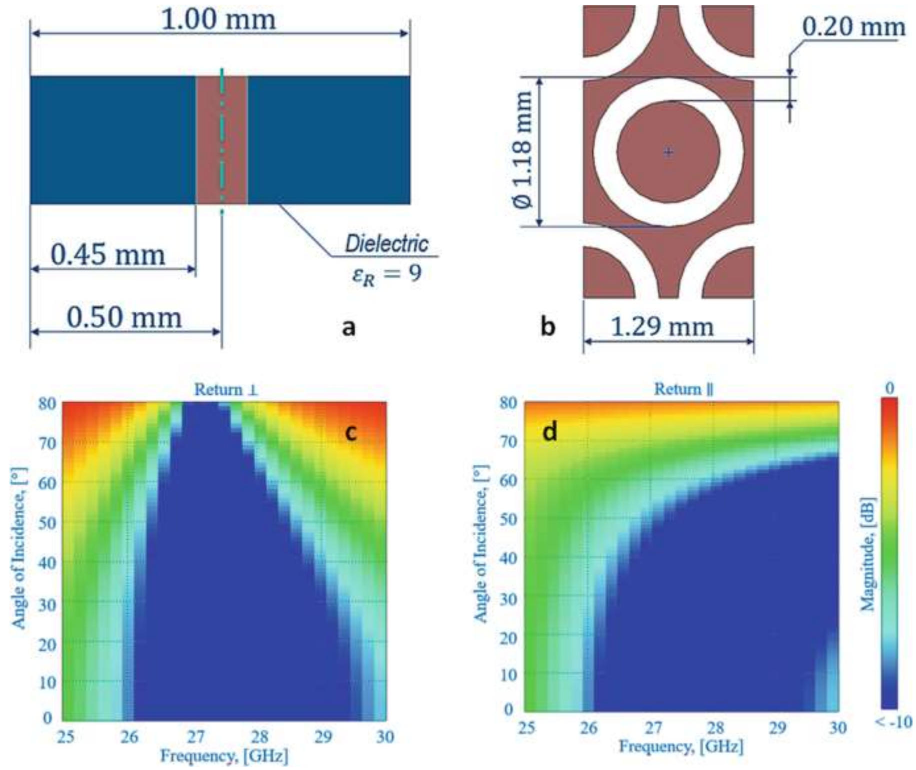


Fig. 5.18 The inserting dielectric slot design example. (Adapted from Enjiu and Rütschlin 2019, credit: Dassault Systems/the 3DEXPERIENCE Company): (a) View of the layers of the back cover; (b) view of the unit cell of the FSS design; (c) and (d) The refection coefficient of FSS unit cell analysis.

Source: In both graphs the vertical axis represents the angle of incidence and the horizontal axis represents the frequency, while the magnitude of the return loss is depicted by the color ramp. The (c) map shows the result for the perpendicular polarization, and the (d) map shows the result for the parallel polarization

If a metal cover were to be used, it would serve as a very efficient barrier, completely prohibiting communication. Glass may let electromagnetic radiation to go through it; nevertheless, the electrical thickness of the glass at high mm-wave frequencies may have a significant impact on the performance of the array. As a result, it is of the utmost importance to design the back cover of a mobile phone in such a way that it incorporates a chip-based antenna array that exhibits sufficient scanning behavior across the frequency of interest. This must be done in order to provide users with the efficient high data-rate connections that they require without compromising the aesthetics or the tactile experience that they have while handling the device.

In the case of the AiP that is used, for instance, it is comprised of a 2 × 2 array of stacked patch antennas that are meant to function within the frequency range of 26.5–29.5 GHz. Additionally, it is designed to steer the beam to +/- 30 degrees in both axes, using dual polarization. It has a length of 10.71 millimeters along one edge and is square in shape. As can be seen in Figure 8.4a, the module is positioned in the top right quadrant of the back of the phone, and there is a gap of one millimeter between the module and the cover. Within the framework of the monolithic radome design, the cover material is made of glass. As illustrated in Figure 8.4b, in order to obtain the requisite thickness of 2.05 millimeters, local thickening is applied to the section of the radome that is directly above the antenna. This is done rather than going through the process of thickening the whole rear cover, which would seem like an electromagnetic window.

The performance of this enclosure in terms of electromagnetic transmission is not excellent. The radiation pattern is fairly deformed by ripples (Fig. 5.18c, d), despite the fact that there is an increase in the antenna matching and efficiency that is seen. This is because a portion of the energy is steered within the dielectric cover or propagates over the surface of the PCB before it is emitter. In order to get satisfactory performance, the FSS ought to be at least marginally bigger than the antenna footprint when it comes to the metallic back. As can be seen in Figure 5.18, the incorporation of the FSS into the rear cover of the phone creates an ideal electromagnetic window that enables the antenna module to radiate effectively with the appropriate far field pattern, which is virtually completely devoid of ripples.

Consequently, the design of the monolithic radome seems to be an inadequate solution for this particular example application, since it resulted in the production of a rippling radiation pattern. On the other hand, if the distance between the antenna module and the enclosure can be extended, then the applicability of dielectric enclosures has the potential to give the performance that is required via their application. In comparison, the metal cover that is loaded with an FSS is a promising option. This is due to the fact that the criteria for antenna matching and beam steering can be satisfied even when there are minor gaps between the beam steering module and the enclosure. In addition, the materials that are being considered for use as front covers for mobile phones should be very lightweight, possess a high degree of transparency, have an antiglare characteristic, be highly resistant to fingerprints, and be scratch resistant. It is common practice to take into consideration wear resistance and efficient heat dissipation while designing the back cover.

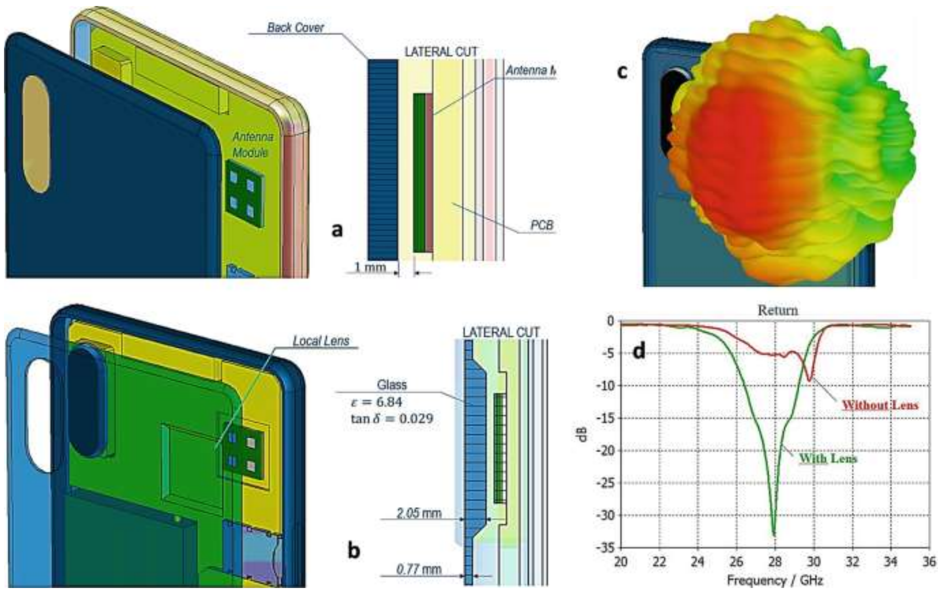


Fig. 5.19 Antenna and phone back cover integration design. (Taken from 2019 by Enjiu and Rütschlin; Dassault Systems/the 3DEXPERIENCE Company is credited): (a) A perspective picture of the phone with the back cover raised, revealing the antenna module's location (left); and (b) a lateral cut at the antenna module, revealing the phone's PCB, antenna module, and back cover stacked up (right). In (b), the phone's perspective view displays the local thickening of the glass back cover on the left, while a lateral cut at the antenna module reveals the design's stacking arrangement on the right. (c) The phone's directivity pattern at the boresight; some of the energy is steered within the dielectric and part is transmitted along the PCB surface, giving the pattern a rippling appearance. (d) The module's return loss response for a rear cover with 0.77 mm thick glass both with and without the local dielectric lens

Some of the materials that have the potential to be strong rivals in the market for mobile phone shells include tempered glass, ceramic, and polymer composites such as polycarbonate (PC)/poly(methyl methacrylate) (PMMA). In the case of the PC/PMMA composite, the scratch resistance of the product is ensured by the performance of PMMA, while the total impact strength may be ensured by PC that has sufficient toughness. Metal is still the primary material used for the central frame of a 5G mobile phone, and the trend for future development is to use aluminum alloys or stainless steel that have a high strength need.

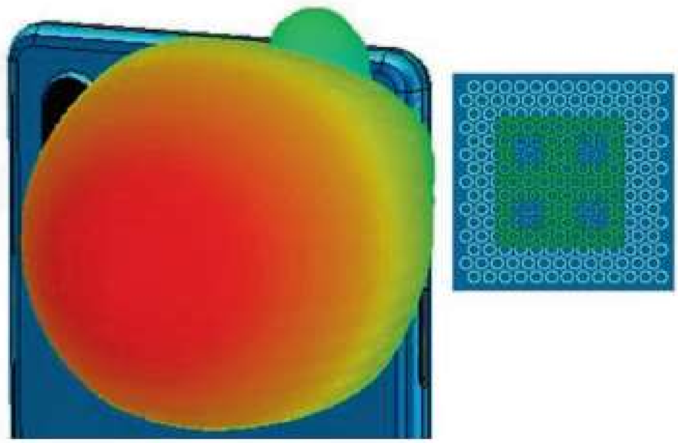


Fig. 5.20 The phone with a metallic back loaded with 13 x 15 components and FSS has a 3D directivity pattern. The radiation pattern shown is for boresight radiation; when the beam is directed, the least distortion of the 3D pattern is still visible

Source: 2019 by Enjiu and Rütschlin; Dassault Systems/the 3DEXPERIENCE Company is credited.

5.6 ENCAPSULATION OF THIN FILM IN 5G ELECTRONIC PACKAGING

A crucial enabling solution that is developing to satisfy the demanding needs of 5G passive devices and networks is the technology known as radio-frequency microelectromechanical systems (RF MEMS). These criteria include high operating frequencies, extensive tunability, decreased hardware redundancy, and low power consumption. These performance benefits, on the other hand, are nullified by the absence of low-cost packaging options that are on the market for RF MEMS. Protecting the device from structural damage and impurities, introducing low RF losses, and not degrading the performance of the switch, circuit, or complicated passive device are all requirements that must be met by the package. Wafer-level packaging methods may be used to package MEMS switches, which allows for the avoidance of high prices and the possibility of damage caused by individual handling and release. When it comes to wafer-level packaging, the most prevalent approaches are wafer-level bonding and thin film encapsulation.

During thin film encapsulation, a sacrificial layer is often used to cover the structures that are going to be sealed, and then cap film is deposited on top of that layer. Decomposition via thermal means or etching through access holes are the following methods that are used to eliminate the sacrificial layer. Last but not least, on top of the cap film, a sealing film is placed in order to completely cover the access holes. Choosing the appropriate material for the encapsulating film is of the utmost importance. It has been suggested that various materials, including organic, metals, and dielectrics, may be used with thicknesses ranging from 1 to 20 micrometers. Aluminum nitride, amorphous silicon, and silicon nitride are the materials that are used for thin encapsulating layers the most often among the varieties of materials which are available. In point of fact, they fulfil a number of essential requirements that are necessary for the packaging process to be suitable for industrial production.

It is possible to address this issue by etching holes that are spread over the cap layer or by fabricating them in the sidewall using surface micromachining methods for conventional applications. Anisotropic plasma-assisted etching methods, which are based on fluorine chemistry, are often used in reactive ion etching (RIE) or inductively coupled plasma (ICP) configurations for the purpose of patterning encapsulation caps with etch holes. It is possible for the substrate bias and physical ion bombardment to alter the stress distribution in encapsulated suspended beams as a result of device heating, which may lead to beam deformation. This is a disadvantage of the RIE and ICP approaches. Inorganic materials, such as silicon oxide and amorphous silicon, are often used as sacrificial materials. However, the etching of these materials involves the use of F-based vapours, which may provide a significant difficulty when combined with micromachining methods for structures that incorporate aluminum or silicon oxide. For the purpose of redistribution of the electrical impulses from the RF MEMS device to the outside world, organic materials, which may be readily removed using oxygen plasma etching, are intriguing candidates for sacrificial layers in manufacturing procedures that take place at low temperatures.

An investigation was conducted to determine how the removal of the sacrificial layer was affected by the percentage of cap perforated area, as well as the amount of time (t) and power (P) that was applied to the barrel etching process. Furthermore, a comprehensive three-dimensional (3D) fnite element method-based (FEM) simulation model might be used to forecast the radio frequency (RF) performance of capped and uncapped CPWs . The oxygen plasma-based release of caps, in addition to the temperatures that are quite low

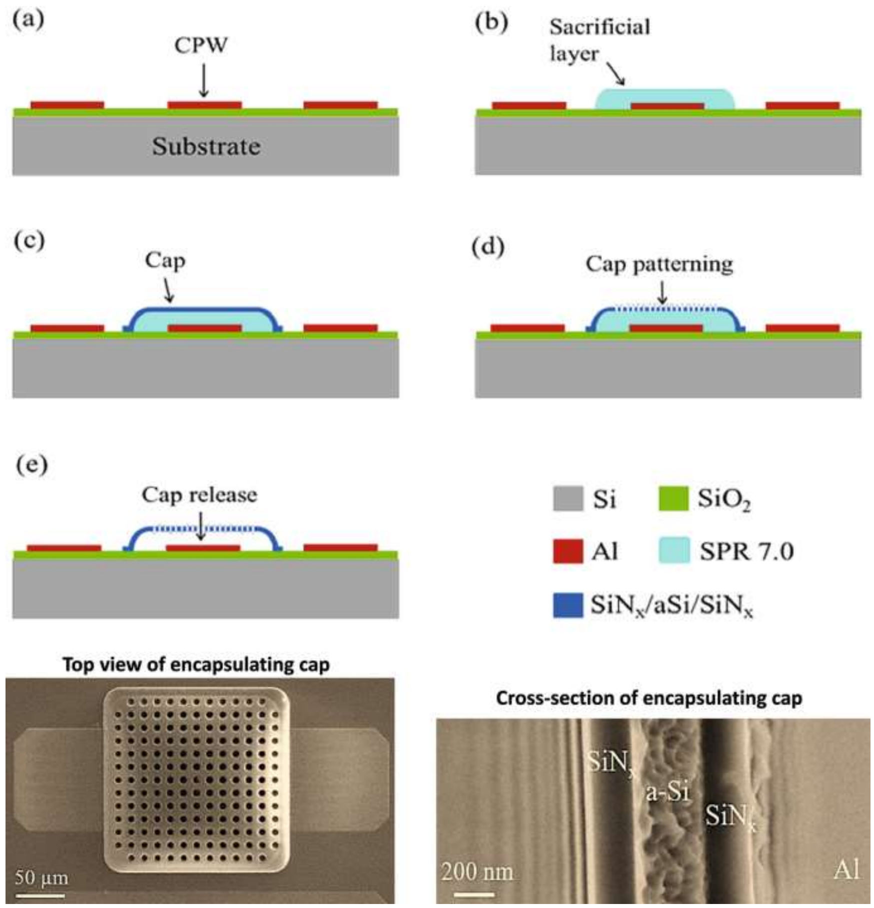


Fig. 5.21 An illustration of the steps involved in creating thin film encapsulating caps (from a to e), as well as a cross-section and top view of the finished product

Source: Taken from Persano et al. 2019; Springer Nature permission granted.

5.7 SEALANTS AND ADHESIVES FOR 5G SYSTEMS

Technologies for bonding, like as adhesives and sealants, are very important in the context of 5G systems. In the process of bonding materials in a variety of elements and components, the use of adhesive technologies is regarded to be an effective method. This is due to the many advantages that adhesive technologies have in comparison to other technologies. For instance, sealants and adhesives are helpful in overcoming the challenges that are associated with conventional techniques. These techniques include mechanical bonding, which may cause the material to become weaker due to the possibility of drilling, and thermal bonding, which involves applying heat to the materials, which then causes the properties of the materials to change.

A number of different classifications may be used to adhesives and sealants, including the following: by chemical composition (adhesive binder), natural vs synthetic, organic against inorganic, structural versus nonstructural, curing or setting process, and so on. However, the classification that is most often used is based on adhesive binders. These sticky binders include epoxy, acrylics, polyurethanes, cyanoacrylates (superglues), an aerobics, silicones, phenolics, polyimides, bismaleimides, and amino resins (such as urea-formaldehyde). In addition, various adhesive technologies can be classified according to the increasing order of load bearing capability, which typically ranges from 0.01 to 40 MPa of overlap shear strength. These technologies include pressure-sensitive adhesives, reclinable fasteners, contact and spray adhesives, acrylic foam tapes, hot melt adhesives, adhesive sealants based on polyurethane and hybrids, polyurethane adhesives (PUR), epoxy, and acrylic and urethane structural adhesives. Structural adhesives provide a number of benefits, including the ability to form strong bindings, the flexibility to design, and the efficiency of the process.

There are a variety of chemistries, curing techniques, open periods, and final bond capabilities that are included in well-established industrial adhesives. These adhesives also include low-odor and non-flammable variants of specific chemistries in order to fulfil certain regulatory and safety standards (Marques et al. 2020). The use of adhesives and sealants in 5G electronics not only makes a direct contribution to the production of electronic goods, but it also adds to the lifespan and functioning of these items over an extended period of time.

When choosing an adhesive for use in electronic applications, it is necessary to take into consideration three distinct processing phases: the uncured or liquid-resin phase, the curing (transitional) phase, and the cured or solid-material phase. Because it has an effect on the dependability of the adhesive, the performance of the cured adhesive is ultimately the most significant factor. There is also a great deal of significance to the manner in which the adhesive is applied, particularly due to the fact that it is essential to guarantee that the appropriate quantity is placed in the appropriate location. Screen printing, which involves "squeezing" the adhesive through patterns in a screen, pin transfer, which involves using multipin grids to convey patterns of adhesive drops to the board, syringe application, which involves "shots of adhesive being delivered by a pressure-regulated syringe," and various technologies for three-dimensional printing are the most common methods for applying adhesives in electronic applications. The application of syringes is still probably the most common approach, and it is often done using electropneumatic ally controlled syringes for the modest manufacturing of a wide variety of PCBs. Also included are the following kinds of adhesive:

CHAPTER 6

APPLICATIONS IN INDUSTRY VERTICALS AND THEIR NEEDS

6.1 INTRODUCTION

With its very broad band width, low latency, and network slicing (NS) capabilities, 5G networks may be able to aid service providers in the development of new application platforms that will enable the next generation of applications as well as new business models. Video streaming and apps that are based on the Internet of Things (IoT) are examples of the "killer applications" that are now available. These applications, when paired with capabilities of virtual and augmented reality, will offer up possibilities in a variety of different sectors. Table 9.1 presents a comprehensive list of new services and applications that will be significantly facilitated by the introduction of 5G.

There are a number of possible future expansions and verticals that might be taken into consideration, including aerospace, ocean, threat response, mobile platform, terrestrial, and distributed computing (cloud/IoT-fog) (IEEE 2017).The requirements for bandwidth and latency that mobility-based apps have are mapped out in Figure 6.1.

6.1.1 5G in Automotive

Connected, cooperative, and automated mobility, also known as CCAM, is one of the most important technologies for enhancing environmental and information flows, as well as increasing road safety and the efficiency of traffic flow. It is projected that the fifth generation (5G) of wireless networks in the automotive industry would significantly enhance wireless connection, make automation easier, and provide autos with a broad variety of digital services. These innovations will make it possible to serve a wide range of advanced vehicular use cases and will also pave the way for fully connected mobility and autonomous driving solutions. In light of this, the automotive vertical industry has been one of the key targets of 5G installations. With the advent of 5G technology, it is anticipated that the automotive industry would go through a profound shift. The capability of autos to create connections with other automobiles, pedestrians, roadside infrastructure, or application servers makes it possible for a variety of novel services to be developed, including the following:

(a) Vehicle platooning, as seen in Fig. 6.2, is the dynamic formation of a group of vehicles that travel together and at a very close distance from one another;
(b) Advanced driving, including autonomous and assisted driving: semi-or fully automated driving, sharing driving intentions, sensor data, and videos captured by onboard cameras with roadside infrastructure, other cars, pedestrians, and network servers for safer travel, collision avoidance, and increased traffic efficiency;
(c) Assistance with remote/cloud computing driving: a remote driver or a V2X application that drives a remote car in hazardous conditions, while carrying disabled passengers, or in public transit.
(d) Vehicle data services: these services concentrate on gathering useful data from all parties involved, such as automobiles and other users of the road, in order to provide a range of services using the current 5G infrastructure.

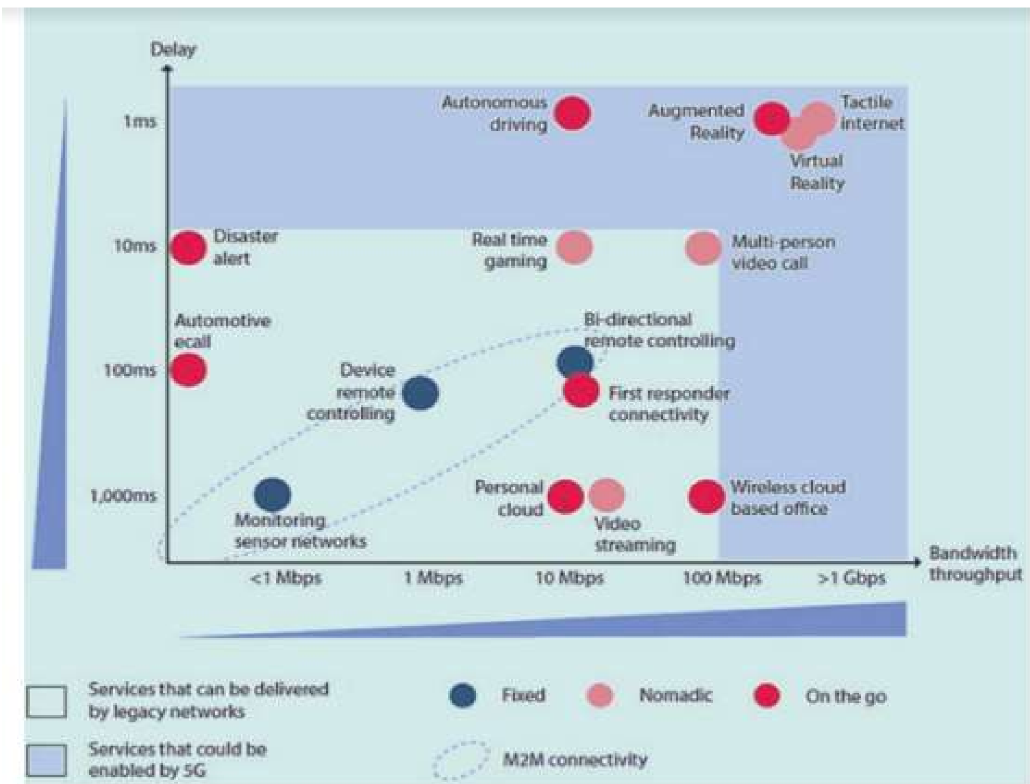


Fig. 6.1 5G application needs for bandwidth and latency

Source: Taken from Dewar and Warren.

There are a few scenarios that are covered in each of these use cases, and each of them focuses on a certain service or function. millimeter wave communications, network slicing, multiaccess edge computing (MEC), and eV2X communications are some of the technologies that are projected to be important in terms of their ability to provide 5G for vehicle services. Network slices, also known as virtual function graphs, and the resources that are necessary to construct them should have an agile lifecycle management system, be easy to setup, be reusable for one or more network services, and maybe be separated from other slices for the purpose of ensuring network security. It is possible that the MEC concept will be implemented as a consequence of network slicing, which includes resources located at the edge of the network. The MEC methodology is a helpful method that has the potential to give low latency, low bandwidth utilization, low energy consumption, and high resilience.

All of these characteristics are very crucial for automotive services applications. At that point, the idea of resource exploitation at the network's edge might be expanded to the most extreme situation, which would include the utilization of resources that are given by the equipment that is used by walkers and cars. Therefore, it is vital to achieve effective V2X communications in order to enable user equipment to share resources such as the network, storage capabilities, and processing capabilities.

6.1.2 Big Data Analytics in 5G

The management and utilization of Big Data, which refers to enormous databases that are growing in size at an exponential rate, is becoming an increasingly difficult task. The administration of large amounts of data is becoming an increasingly difficult task. Taking into consideration the following factors—collectively known as the "five Vs"—this method outperforms traditional data analysis: the amount and variety of data; the speed and significance of results derived from data; and lastly, the quality of data, which encompasses trust, credibility, and integrity (veracity). Taking into consideration the fact that the number of devices that are connected to the internet is anticipated to expand by ten to one hundred percent as a result of broad adoption of 5G for business applications.

Therefore, big data approaches will be essential for future 5G networks, as all of the use cases that have been explored rely on the extraction of data from the enormous amounts of heterogeneous data that are generated by connected devices in order to assist with decision-making and other tasks.

Creating new communication systems in the future should have as its primary objective the fulfilment of the ever-increasing demand for communication while adhering to the limits of limited resources. With the further development of big data technologies, the following is anticipated to take place:

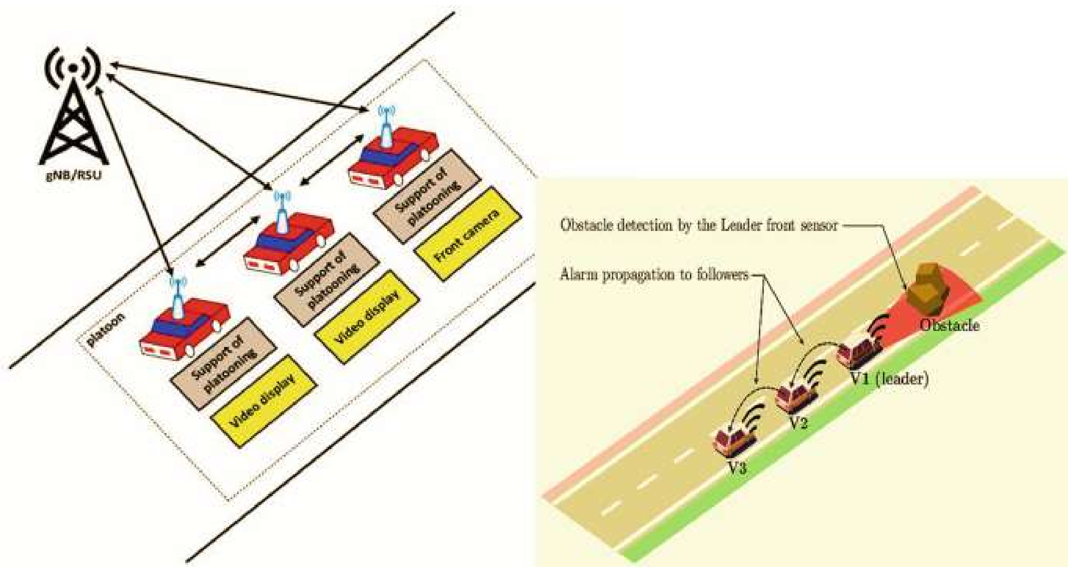


Fig. 6.2 Permeable to platooning

Source: Reproduced with permission from Bouali et al., 2021; open access (IEEE)

(a) **5G Networks Facilitate Big Data:** Networks that support 5G technology serve as the transmission infrastructure for the big data processing chain. In their capacity as a transmission medium, they transport data from a wide variety of sources to their respective data centres. Mobile user equipment, closed-circuit video, and other sensors that contribute to the big data chain are some of the numerous devices that are included in the spectrum of devices that comprise 5G networks. These networks are likewise comprised of a vast number of devices that are connected to one another. The results of this would be an even more significant increase in the number of data sources as well as the variety of those sources. Before transferring data, communication networks must first gather, store, and pre-process the data. This is an essential step beforehand. The process of achieving data redundancy in preparation for further transmission involves the utilisation of data aggregation and compression software. In addition, the movement of data to datacentres or the cloud requires the use of wireless networks, which include both core networks and radio access networks. This indicates that 5G networks are required to take into consideration the requirements of big data applications, which include end-to-end latency that is as low as possible, guaranteed data integrity, and reliability.

(b) Big Data Assist 5G Networks: In the same way that 5G wireless networks have the potential to simplify the processing of enormous amounts of data, big data in the telecommunications industry has a great deal of promise for improving both the performance of networks and the user experience. Attempts have been made by telecom operators to incorporate data analytics into their operations in order to improve the efficiency with which they manage their network resources, investments, construction processes, and user experience. Both data analytics and big data have the potential to bring about a revolution in the telecommunications business by enhancing the construction, administration, and optimisation of communications networks. Additionally, there is the possibility of gaining significant information through the analysis of user behaviour, locations, trajectories, and preferences regarding service. Using this information, we have the potential to make the network significantly faster and to enhance the experience of the users.

(c) Synergy of 5G and Big Data Enables Vertical Industries: With the commercialization, installation, and connecting of 5G networks with systemized data operations, vertical organisations will have the potential to benefit from the numerous intelligent applications that will emerge as a result of these developments. Figure 9.3 illustrates a conceptual relationship between vertical enterprises, big data, and fifth-generation wireless networks. This will ensure that individuals will always have access to internet services because the mobile communication networks that will be accessible with 5G communications will be dependable, quick, and have minimal latency. Because of their capacity to sense and collect data for business clients, networks that are part of the Internet of Things (IoT) continuously deliver high-value information that assists businesses in making decisions. Through the use of distributed computing, the delivery of intelligent services in real time becomes a feasible prospect. The combination of artificial intelligence (AI), cloud computing, and big data is what drives the intelligence-based platform that consists of computing power, large data, and service-based connectivity. The combination of these two factors results in the creation of one-of-a-kind services and the encouragement of vertical industries to expand. In the process of transitioning to digital technology, nearly all of the eight important industries that are illustrated in Figure 6.3 are encountering challenges. Through the implementation of specialised or universal solutions derived from 5G and big data technologies, vertical industries may be able to find support in satisfying their requirements for data platforms and connection. Additionally, the use of these technologies will pave the way for the development of revolutionary applications such as autonomous vehicles, aerial vehicles, visual and augmented reality, and many others. Due to the fact that data quantities, resilience, end-to-end latency, transmission capacity, real-time/on-time processing needs, and data security are all situationally dependent, design techniques for communications and data platforms should take into full consideration the desired applications.

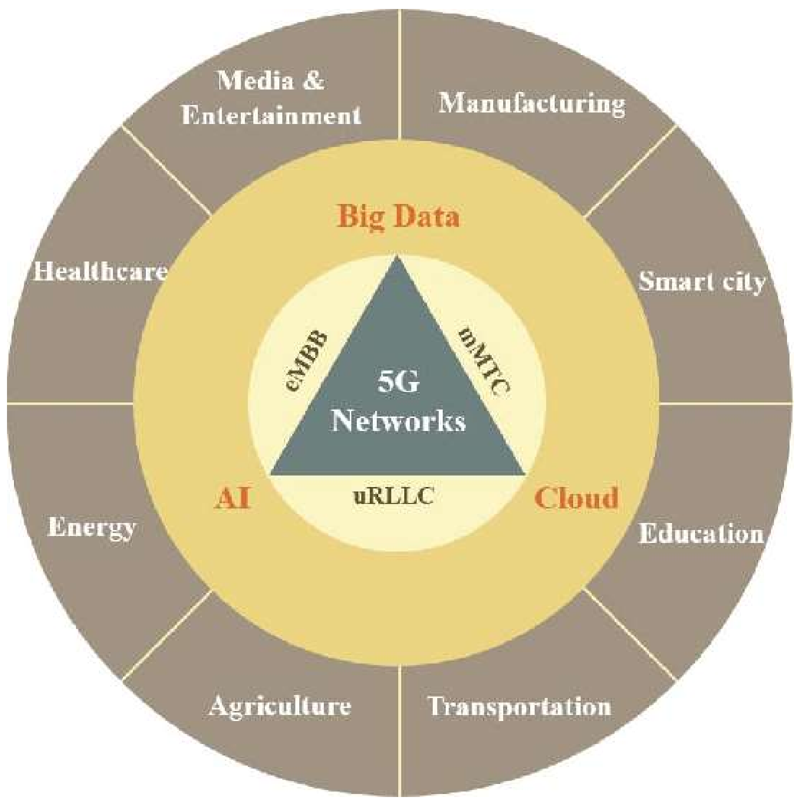


Fig. 6.3 Vertical Industries are enabled by "5G + Big Data"

Source: Reproduced from Zhu et al., 2019 with IEEE permission.

On the other hand, the following challenges may be encountered by the next-generation of telecommunications networks, which may limit the expansion of big data:

(a) Due to the complexity of usage patterns and the unstoppable development in network complexity, manual configuration is not a feasible choice for network design, maintenance, or optimisation. This is because of the fact that the complexity of networks is on the rise. The typical reactive maintenance method is not only inefficient but also unable to provide quality and service that is constant over time. Intelligent, self-aware, and able to adjust to its surroundings are the characteristics that would define the perfect operating system. Because of this, it is essential to have analytics that are both autonomous and intelligent, as well as technology that assists network operators in managing their networks in real time and doing so at a cost-effective level.

(b) There will be an increase in the amount of data that receives integration and interchange. The combined data originates from a variety of sources, and different individuals have access to a variety of formats and semantics within each of these sources. This is done in order to get the maximum amount of value from the data. On the other hand, due to the fact that a significant portion of the data is associated with particular individuals or business procedures, it is imperative that data security and privacy be properly considered. The importance of having a conversation on the usage of data in accordance with the laws governing data protection cannot be overstated.

9.1.3 5G Emergency Communications

The transition of telecommunications companies to this technology is directly responsible for the need for emergency systems to modify their emergency communication platforms so that they are compatible with broadband Internet Protocol infrastructures. This is necessary in order for emergency services to meet the regulatory standards for the next generation of emergency services. This direction is being driven by a wide variety of applications and services, each of which has its own different set of requirements regarding application performance. These include, but are not limited to, communication with huge machinery, mobile connections capable of gigabits, and Internet of Things devices that are essential to the operation of the business. Last but not least, in order for emergency systems to be considered compliant with the regulatory criteria for next-generation networks (NGNs), an upgrade will be required. The potential of next-generation mobile and wireless networks to improve efficiency, security, and device-to-device communication are having a significant impact on the public safety sector as a whole, as well as emergency communications companies and the services they provide. The legacy systems are particularly susceptible to the effects of these characteristics.

In order to keep the stringent operational and managerial requirements of emergency services, it would be necessary to integrate the anticipated features of 5G in a close manner. We have great hopes that it will increase resilience against data pollution and security threats, both of which have the potential to slow down the reaction times of first responders, and that it will provide real-time, high-priority total conversation services (voice, video, and real-time text) for emergency communication. The real development of the 5G network will determine the fate of this event so keep that in mind. In addition, the capacity of uplinks and the availability of communication channels will both be improved. This will be accomplished through the establishment of connections between devices. Because of this, emergency personnel will be able to gain access to high-quality multimedia information, which will not only improve their awareness but also make it easier for them to provide an "always connected" experience.

Network operators will have the capacity to make real-time adjustments to the transmission speed and latency of their networks thanks to the network-slicing feature of 5G. Consequently, this will ensure that the communications of those who respond to emergencies are given the utmost attention. It is required of network operators that they would deliver ultralow latency, ultrahigh availability, and dependability for these services in order to meet the severe criteria that have been established by emergency communications and PPDR services in general. The future generation of emergency services, which will be built on the enormous Internet of Things and device-to-device connections, will be distinguished by increased throughput, improved Quality of Service (QoS), and reduced buffer requirements for user devices. The low latency of the underlying access network will make it possible for these developments to be taken into consideration. On the other side, mobile edge computing, also known as MEC, has been proposed as a potential alternative to cloud-based systems.

It is important to note that MEC lays the way for the potential of a high-performance virtual environment at the edge of the network. Due to the fact that they are located in close proximity to the Internet of Things environment, the applications and services that are supplied by MEC offer increased bandwidth, decreased latency, and improved quality of service. Identifying potential medical scenarios that could be helped by remote support is something that is within the realm of possibility at the MEC level. At this level, we are able to recognize events such as sending the user a call back in the event of an emergency, remotely altering the dosage of medication, and other similar scenarios. In the event of an emergency, the healthcare provider will contact the emergency service operator so that they may obtain the patient's current location, insurance information, medical history, and data from their current health sensor.

When the situation requires first responders to take immediate action, this is the course of action that is taken. In the event that the building management system becomes aware of a potential danger, the security provider who is in charge of the MEC platform may examine the footage captured by the cameras located at the location in order to verify that the alert is not a false positive. On the other hand, the construction of a framework for the action plan for first responders might begin at the same time that an emergency call is placed to the fire brigade. Using the MEC's remote provisioning mechanism, it is possible to perform data transfer in a manner that is very close to real time. This is accomplished by reducing the delay that occurs between sensor reports.

In contrast to the current state of GSM and LTE-based emergency service networks, which means that specific use cases are assigned to individual physical networks (for example, GSM for voice and LTE for mobile data), the architecture that is being proposed for 5G emergency service networks has the ability to establish and manage multiple virtual access networks. In this way, it is ensured that all emergency services, such as the police, the fire department, and the ambulance, have access to dedicated network resources within the event site. The elimination of interference in communication between the various networks is made possible by network slicing, which ultimately results in incredibly low latency and exceptionally high throughput. Through the utilization of programmable infrastructures and software-defined features, we are able to achieve this flexible orchestration of network slices. The MEC is in charge of managing the NFV infrastructure, which is responsible for regulating the backhaul of RANs.

In specifically, the NFV controller in the MEC is responsible for organizing the allocation of bandwidth to each wearable device that is utilized by first responders for the purpose of health monitoring or communication. Delays, losses, active bearers, and other features of traffic that are related with it are also included in its scope of responsibility regarding traffic. As a result of this, emergency service providers have the option of upgrading their systems in order to expand their capacity to manage a wide range of services and applications that have variable requirements for performance. This category includes communications that include huge machines, communications that involve the Internet of Things (IoT), and communications that are mission-critical because they involve gigabit mobile broadband.

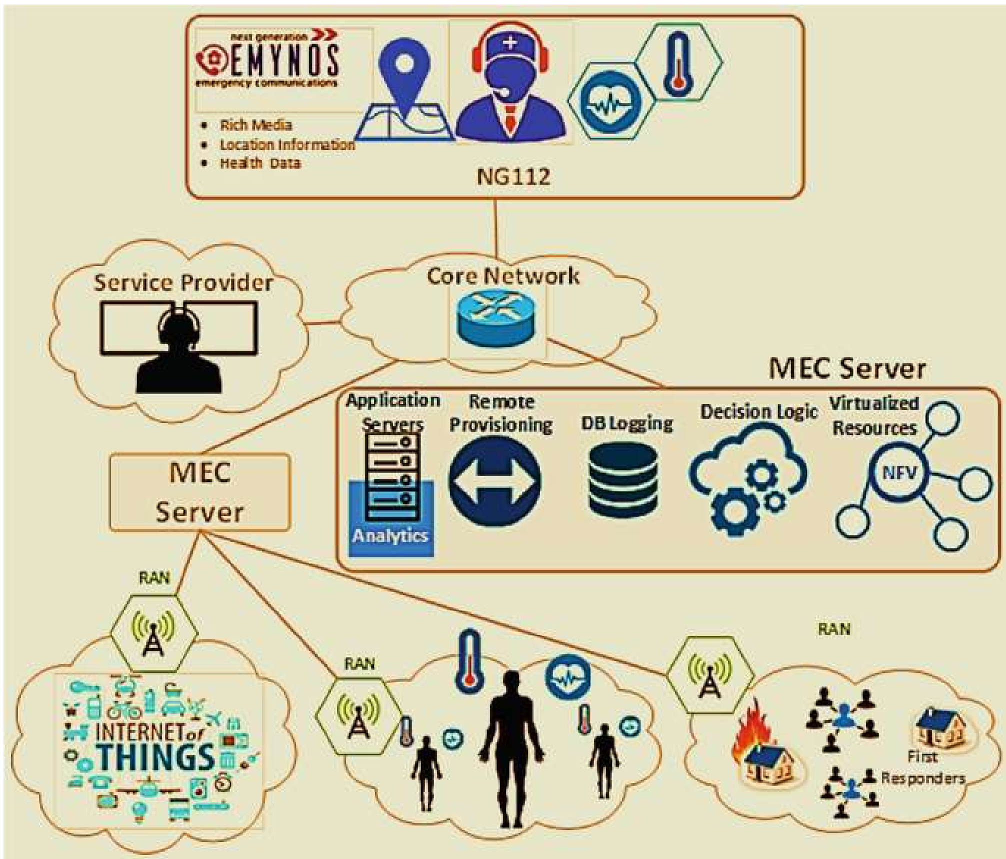


Fig. 6.4 An example of emergency communications with 5G

Source: Politis and Markakis, 2017

6.1.3 Future Factories Enabled by 5G Technology

When it comes to the industrial sector, the total digitization of the industry requires connections that are both extensive and granular. Industrial communication networks are now undergoing the process of incorporating new technologies such as the Industrial Internet of Things (IIoT), time-sensitive networking (TSN), and 5G wireless networks in order to acquire the necessary connection spectrum. The core and access networks of 5G are currently being expanded to the point where it will be able to act as a transparent or independent TSN carrier in demanding OT application situations. According to the plans for the future, TSN will be a wired networking solution that will unify data streams related to operational technology and operational technology. The optimization of cellular infrastructure has made it possible for mobile Internet networks to achieve faster data rates and wider coverage. This is now conceivable.

The connectivity requirements of vertical industries such as energy and transportation, manufacturing, and building automation are considerably different from those of horizontal enterprises. This is because horizontal businesses are more horizontal than vertical businesses. 5G networks, as a consequence, function better than their predecessors in terms of their capabilities and their adaptability. In addition, the architecture of the 5G network goes beyond the mobile broadband services that are directed at consumers and include aspects that are specialized to the Internet of Things (IoT), Industrial Internet of Things (IIoT), and linked cyber-physical systems (CPS).

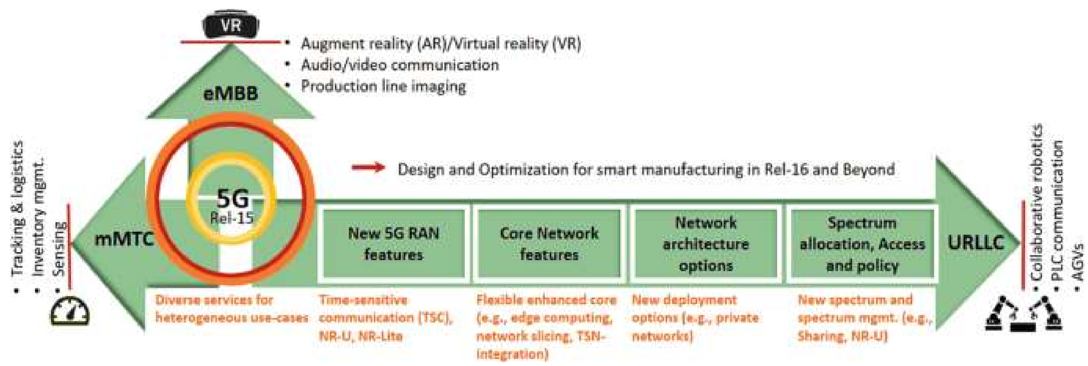


Fig. 6.5 The design and optimization of 5G for a range of industrial uses. Rel-16 concentrated on IIoT-related enhancements, such as URLLC for TSC, 5G-TSN integration, localization services, private networks (PN), network slicing, and NR on unlicensed bands (NR-U) for URLLC; Rel-17 and beyond will cover further improvements in PNs, convergence with industrial networks, network automation, and NR-Light for new IIoT use-cases. 3GPP Release 15 (Rel-15) defined eMBB, mMTC, and URLLC services for 5G New Radio (NR)

Source: Taken from open access (IEEE) by Mahmood et al., 2021)

The Industrial Internet of Things (IIoT) is a framework that utilizes industrial objects (machines, equipment, and processes) as a communication network. Its purpose is to facilitate the reliable interchange of control and monitoring data. On the other hand, the CPS makes use of the IIoT in order to provide a digital description of the items that is interactive, synchronized, and consistent. Improvements to mobile broadband (eMBB), massive machine-type communications (mMTC), and ultrareliable low-latency communications (URLLC) are the three primary connection services that 5G makes available to its users. Figure 6.5 illustrates the apparent industrial use cases of the services; nevertheless, thanks to the integration of these services, 5G has the ability to serve a wide variety of Internet of Things applications.

While it is true that 5G will eventually become a uniform connectivity fabric, there are still a lot of questions and concerns regarding how it will be able to fill connectivity gaps, integrate with the TSN, and meet time-sensitive communications (TSC) targets while maintaining local access to business-critical data when it is seamlessly integrated into the factory floor. In spite of the fact that 5G will eventually become a fabric of uniform integration, these doubts and concerns are still being voiced. A number of companies and mobile network operators (MNOs) are attempting to get knowledge regarding the design and optimization of the 5G architecture in order to expand their business models beyond mobile broadband internet.

The growing ripple in Figure 6.5 is a visual representation of this phenomenon. It is necessary to conduct extensive research on private (nonpublic) 5G networking models and the spectrum licensing options that are connected with them in order to guarantee the delivery of services that are highly individualized, flexible, and cost-effective for industrial use cases. In addition to this, 5G provides a number of autonomous deployment options that could make it possible to provide key privacy and security features, in addition to providing specialized support for important use cases.

On the other hand, it is the role of solution providers to make the management and operations of 5G networks more straightforward for public-private networks. In addition to this, they should provide stand-alone solutions for a variety of network-related problems, including optimization, functional splits both on-site and remotely, the deployment of new services, and quality of service guarantees. There is a significant increase in the number of security vulnerabilities that are present on the manufacturing floor when private 5G wireless networks share resources with public networks. The Third Generation Partnership Project (3GPP) provides a variety of strategies that can be utilized to enhance the resilience of private networks. These strategies include the replication of network operations and resources, the authorization and isolation of services and slices, and other similar strategies. The employment of beamforming techniques in conjunction with the dynamic spectrum monitoring and allocations offered by 5G is one potential strategy for providing protection against attacks that involve radio jamming. Nevertheless, a practical suitability assessment is necessary for the development of approaches such as federated learning and physical layer security. These methods are intended to provide real-time control and analytics while simultaneously decreasing the additional communication overhead that is associated with security measures.

6.1.4 Smart Health-Care Network Based on 5G

It is possible that intelligent medical devices may improve the quality of treatment that patients receive since they will enable the constant monitoring of vital signals. Intelligent diagnostic tools also have the potential to make it possible for patients to receive therapy that is at the cutting edge of medical technology. To make people's lives easier, the goal of smart health care is to provide them with information about medical disorders and the treatment options available for them. As a result of advances in medical technology, individuals now have the ability to take the appropriate actions in the event of an emergency. In turn, it reduces the expenses of healthcare by enabling medical professionals to communicate with a greater number of patients, regardless of where they reside. Additionally, it enables remote checkups, which is a significant advantage. It is essential to build a robust smart health care infrastructure in order to guarantee that individuals will have access to medical treatment even as smart cities continue to expand.

In addition to enhancing one's health, one of the most significant advantages of speedy diagnosis is the reduction in the costs associated with medical care. In the future, the Internet of Things will bring about a paradigm shift in healthcare and a decrease in the cost of medical devices. 5G networks are absolutely necessary in order for the Internet of Things to achieve widespread adoption. One of the most important applications for 5G networks is the introduction of intelligent healthcare. An example of a smart health care network that is built on 5G technology and its essential components are presented in Figure 9.6, along with the general design of the network.

It is possible that the proliferation of the Internet of Things (IoT) would be beneficial to a great number of intelligent health care applications. Telephony in the healthcare industry, assisted living, tracking of behavioral changes, remote monitoring, tracking of treatment adherence, and asset management in residential care facilities are all examples of applications that fall under this category. The machine-to-machine (M2M) link that 5G networks provide and the Internet of Things (IoT) are currently being considered as potential foundations for the development of intelligent health care. As a result of the ongoing advancements in sensing equipment, technology, and telemedicine, the healthcare business is on the verge of experiencing a paradigm shift on a worldwide scale.

When it comes to the health care industry, the following are the most essential technical considerations:

(a) **IoMT Devices:** Handling the vast volumes of data generated by IoMT devices (smart sensing remote/in-home monitoring systems, wearables, and implanted medical devices)
(b) **Smart Wearables:** these devices can track vital signs including breathing rate, heart rate, and baseline blood pressure.
(c) **Tactile Internet:** millisecond latency and fast reaction times are especially important in remote healthcare including two-way communication and robots.
(d) **Critical Communications:** using QoS methods in conjunction with remote blood sugar, ECG, and temperature monitoring to conserve capacity for critical communication.
(e) **Emergency Medical Services:** ambulance services promptly transmit pertinent patient data, including high-definition photos and videos, back to a nodal hospital.
(f) **Security and Privacy:** Maintaining the confidentiality and integrity of patient data is crucial and calls for the adoption of certain safeguards. Ensuring the necessary levels of privacy, preferred network resource management, and assured Quality of Service (QoS) would be anticipated from data handling.
(g) **Analytics:** The volume of data processing that has to be done is increasing rapidly due to the surge in UEs and widespread wireless connections in the healthcare industry. A composite large dataset including several attributes, including user location, services used, signalling information, and apps in use,

must be managed and analysed. As a result, new use cases and models are created for data extraction, formatting, storage, and display.

It is not possible for the health care business to depend only on a single technology or solution. The shift from the technologies that are now in use to those that will be developed in the future will be a methodical, gradual, and need-driven process. A hospital-based, specialist-focused strategy is rapidly being replaced by a dispersed patient-centric care model as a result of demographic and socioeconomic developments. This shift indicates that healthcare paradigms are rapidly altering.

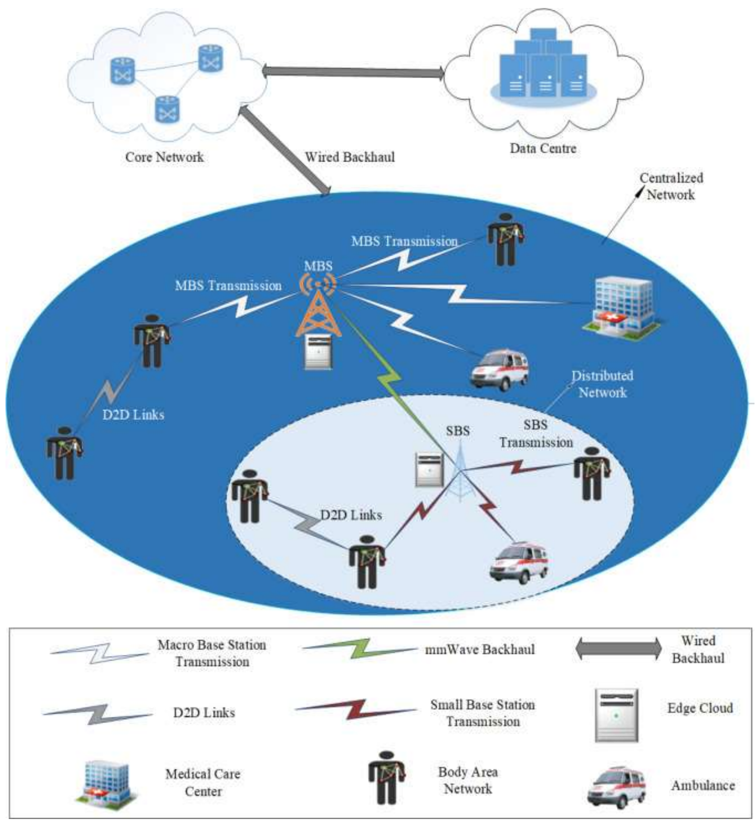


Fig. 6.6 An overview of a 5G-based smart heath care network's general architecture

Source: Adapted from open access (MDPI)

6.1.5 5G Technology for Smart Energy Management and Smart Cities

The formation of 5G technology occurred as a result of the convergence of a number of factors, including cutting-edge mobile devices and cutting-edge communication networks. Having said that, technologically speaking, it has the potential to initiate a wide variety of new economic and industrial endeavors. In addition to this, it has the potential to facilitate the collaboration of a multitude of devices with one another. 5G cellular networks will, without a doubt, bring about a global industry transformation in the years to come, which will have an effect not only on businesses but also on consumers. Internet of Things (IoT) apps, which generate enormous volumes of data, are among the most essential components of smart city applications. The most effective methods for analyzing huge volumes of data are those that are currently in use. These methods include deep reinforcement learning (DRL), machine learning (ML), and artificial intelligence (AI). When it comes to determining their long-term objectives, this could be of assistance to firms in making the appropriate decisions.

A further addition of training data might perhaps assist make these procedures even more suited, which would be like putting salt to the wound. It is possible that they will discover that this improves their capacity for learning, their ability to think critically, and their ability to make decisions. It was also at this time that there was a discernible increase in the utilization of advanced big data analytic tools in the process of developing smart cities. 5G features several exciting innovations, such as the Internet of Things (IoT), blockchain, unmanned aerial vehicles (UAVs), artificial intelligence (AI), machine learning (ML), and projects based on DRL.

6.1.5.1 5G Technology for Smart Cities

It should come as no surprise that communication is the primary focus of the communication network that 5G will provide. As an additional benefit, 5G may be dynamically adjusted to use the appropriate control layer depending on the application. This is a significant advantage. There would be a great deal of advantages to doing this. The lightning-fast data transfer speeds offered by 5G technology will make it possible for a wide variety of businesses to take advantage of its capabilities. Table 6.2 illustrates how the most advanced 5G technology is progressing towards the development of technological applications that are inherently more complex and intelligent.

For time-sensitive applications, the latency requirements for 5G technology are extremely stringent, specifically less than one millisecond. On the other hand, the requirements for applications that are not time-sensitive are less stringent. In a surprising turn of events, 5G could find applications in a wide variety of domains that call for extremely dependable networks due to stringent process controls. The dependability of the network, on the other hand, may permit applications that are not process-sensitive to be more forgiving than the average. The fifth-generation wireless technology has the potential to be utilized in applications that need the processing of enormous amounts of data in real time.

In situations where the processing of extremely small data sets is necessary, the system might be instructed to come into play. Over the past few years, the Internet of Things (IoT) has emerged as a crucial use case for 6G wireless networks due to its ubiquitous application across a wide range of diverse businesses. As the Internet of Things continues to advance, an increasing number of websites will be required in order to access and share data across the network that is the Internet. There is no doubt that the exponential growth of the Internet of Things has been facilitated by the widespread adoption of cloud computing and various other forms of centralized data storage facilities. In spite of this, customers frequently are unaware of how the system makes use of the information that they have provided, which contributes to an atmosphere of mystery. It is possible that the data's transparency could be jeopardized in certain circumstances if the system is centralized. Blockchain technology is one option that might be taken to improve the privacy and security of computers and networks that are based on the internet.

As the number of Internet of Things devices continues to increase, the change of city infrastructure is becoming increasingly apparent. The introduction of 5G networks is essential to this transformation. In this approach, the Internet of Things would be able to serve a wide variety of commercial goals across a number of vertical industrial sectors. These sectors include, but are not limited to, the manufacturing and raw material production industries. The implementation of 5G will make it possible to develop innovative applications for smart cities in the future. Additionally, there will be an ever-increasing number of pieces of electronic equipment that are capable of connecting to the internet. Due to this, integrating a multitude of vertical applications would be a simple and straightforward process.

6.1.5.2 Applications of 5G Technology in the Construction Industry and Infrastructures

There are an increasing number of technologies that are having an impact, such as smart grids, smart houses, and improvement in energy use. Intelligent buildings, in addition to their essential role in reinforcing metropolitan areas and the infrastructures that maintain them, have the potential to significantly improve the levels of comfort that are experienced by the people who live in them. The reason for this is that it improves energy efficiency, quality of life, accessibility to services, safety controls, and general comfort. Many definitions of "smart building" have been presented, although the majority of them focus on "smart grid" concepts and the optimization of energy efficiency. However, there are many more definitions that have been proposed. Storage and analysis of enormous amounts of data are both made possible by intelligent management systems, which offer features that make this possible.

Consequently, the incorporation of these systems into buildings has the potential to result in a significant improvement in energy management. The reason for this is because electrical components that are connected to a grid have the ability to learn and adapt to new behaviors, which makes them "intelligent" building systems due to their functional capabilities. In light of this, this continues to be the case. Intelligent services are those that make use of technology such as big data and the internet of things. This is due to the fact that they make use of extraordinarily large data sets for the purposes of analytics and self-directed learning.

While this is going on, the most important parts of the expansion of smart cities are the unique capacity of 5G networks to fulfil a variety of smart city requirements, smart edge systems that are based on 5G, and other technologies that facilitate collaborative intelligence. 5G will be the driving force behind the creation of pervasive sensor networks, which will serve as the basis for the development of smart cities.

6.1.5.3 Smart Building System Integrated with 5G Communication Technology

According to a number of different definitions, an intelligent building is someone or something that includes cutting-edge communication and technology systems (such autonomous monitoring equipment), effectively organizes information resources, and provides a good return on investment. The architecture of these buildings is not only ecologically sound but also aesthetically pleasing and efficient in terms of energy use. In order for the construction industry to achieve intelligent building, it is necessary for them to have both traditional industrial components and contemporary technological components. Intelligent buildings are equipped with cutting-edge communication and control systems, which ensure that their occupants are provided with comfort, convenience, energy efficiency, and environmental friendliness.

In order to achieve the functionality of a smart building, it is necessary to have a multiplicity of actuators, sensors, systems, and devices that are all interconnected and work together. As an essential component of smart buildings, a communication network that can be managed remotely from within the building itself is an essential component. The use of sensors to automatically close all of the windows prior to turning on the air conditioner is one example of a system that functions in this manner. The activation of this sensor could be triggered by a change in temperature. In layman's words, the sensor is responsible for initiating the air conditioning system, which then proceeds to exchange data with the windows. Which manufacturer is able to give the necessary technology to operate the windows and air conditioner will be determined by the specific technical parameters of the case? The importance of implementing an integration strategy to provide automated building observation and maintenance is brought into focus by this fact. An ecosystem in and of itself, a 5G network will pave the way for a society that is totally mobile and connected to the internet to become a reality for the general public. Not only that, but it will also make it feasible to design new business concepts that are viable and profitable.

The use of sensors to automatically close all of the windows prior to turning on the air conditioner is one example of a system that functions in this manner. The activation of this sensor could be triggered by a change in temperature. In layman's words, the sensor is responsible for initiating the air conditioning system, which then proceeds to exchange data with the windows. Which manufacturer is able to give the necessary technology to operate the windows and air conditioner will be determined by the specific technical parameters of the case? The importance of implementing an integration strategy to provide automated building observation and maintenance is brought into focus by this fact. An explanation of this essential characteristic is provided in further detail in Table 6.3, which also provides a summary of the most significant objectives and technical criteria for smart buildings.

As a result, environmentally responsible and intelligent building practices are gaining popularity, which indicates that they are climbing higher on the priority lists of educational institutions and municipal governments all over the world. There is a possibility that the application of sustainable development goals in the design, operation, and administration of buildings could be significantly improved by the addition of 5G technology. This objective can only be accomplished if we are able to enable services of the highest possible quality and deliver functionalities that are efficient.

6.2 PERSPECTIVES ON 6G WIRELESS COMMUNICATIONS

It is currently being deployed all over the world that 5G, which stands for the fifth generation of wireless communication networks, is being implemented. In addition, unique characteristics such as guaranteed low latency, mass connectivity, and ultra-reliability are currently going through the process of being standardized. When it comes to time and phase synchronization, 6G will require a higher level of precision than what 5G is capable of providing. Additionally, in order to fulfil the requirements of those use cases, 6G will need to have a geolocation precision of less than one centimeter and a geolocation update rate of one millisecond. This is in addition to the fact that it will provide coverage that is nearly worldwide.

There are certain applications that cannot be implemented in distant regions, such as highways and villages, due to the fact that 5G networks are currently constrained to specific common scenarios and they do not have sufficient coverage in those areas. One example of this is driverless vehicles. The incorporation of nonterritorial and specialized satellite communication networks into terrestrial networks is necessary in order to accomplish the goals of cost-effectiveness, seamlessness, and ubiquitous service availability. We need to equip unmanned aerial vehicles (UAVs) with communication networks if we want them to be able to react rapidly in potentially hazardous and challenging circumstances.

Up until the arrival of 5G, the majority of network traffic will be comprised of video and streaming applications. Cellular technology is finding more and more applications in the wireless networked control of robotic products (such as autonomous driving or manufacturing logistics), despite the additional challenges that it presents. It is required for numerous mobile objects to exchange control and sensor data in order to evaluate the amount of network traffic that is generated by these applications. On the other hand, this places an unacceptable level of pressure on a control system that is administered centrally. Meanwhile, distributed control systems that are powered by artificial intelligence are becoming an increasingly important focus of research and development. The concept of federated learning is one potential response.

In this strategy, learning is aggregated in the cloud through the distribution of dataset correlation algorithms among mobile robotic devices. It is fascinating to contemplate the manner in which this results in a new kind of network traffic that comprises a substantial amount of bandwidth and a range of latency needs. The traffic demands of the 6G network may be more than fulfilled, if not regulated, by these and other applications that utilize artificial intelligence. In addition to this, there is a possibility that is extremely genuine.

There will be a significant rise in the overall coverage range of wireless communication networks as a result of contributions from satellite communication, communication from unmanned aerial vehicles, and communication from maritime sources. Only by doing a comprehensive investigation of all of the available spectra, which includes the sub-6 GHz, mmWave, THz, and optical frequency ranges, is it possible to reach a data rate that is more substantial. The seamless integration of artificial intelligence and machine learning with 6G wireless communication networks will make it possible to improve network management and automation, which will in turn pave the way for the development of entirely new applications.

In addition, the dynamic orchestration of networking, caching, and computing resources that is made available by technologies that utilize artificial intelligence may be able to improve the performance of the next-generation network. The adoption of resilient or endogenous security for both the physical and network layers is the final trend in the expansion of networks, although it is by no means the least important development. 6G wireless communication networks will be significantly accelerated by industry verticals such as digital twin body area networks, cloud virtual reality (VR), Internet of Things (IoT) industry automation, cellular vehicle to everything (C-V2X), energy-efficient wireless network control, and federated learning systems. These verticals will have a positive impact on the development of 6G networks.

Figure 6.8 presents a high-level overview of wireless networks that offer 6G connectivity. The following are some of the components that are included in this synopsis: key performance measures, industrial segments, enabling technology, paradigm shifts, and prospective use cases. To summaries, it is anticipated that sixth-generation (6G) wireless communication networks will bring about improvements in intelligence and security, as well as worldwide coverage and increased efficiency in terms of energy, spectral density, and cost. In order for 6G networks to succeed in meeting these expectations, it will be important for them to include innovative network architecture as well as new enabling technologies, such as air interface and transmission technologies.

The design of waveforms, cell-free architecture, network slicing, multiantenna technologies, channel coding techniques, multiple access, and cloud, fog, and edge computing will all be components of this architecture. There is a possibility that upgrading from 5G to 6G will result in four significant shifts in thinking. To begin, the sixth-generation wireless network will not be limited to communication networks on land; rather, it will be able to fulfil the requirement for coverage all over the world. To construct an integrated communication network that encompasses space, air, ground, and sea, it will be necessary to link it with nonterritorial networks such as satellite and unmanned aerial vehicle (UAV) networks. An integrated communication network will cover all of these areas. During the second phase, a comprehensive investigation of all frequency bands is carried out with the objective of significantly enhancing data transmission speeds and connection counts. This investigation encompasses sub-6 GHz, terahertz (THz), millimeter wave (mmWave), and optical frequency bands.

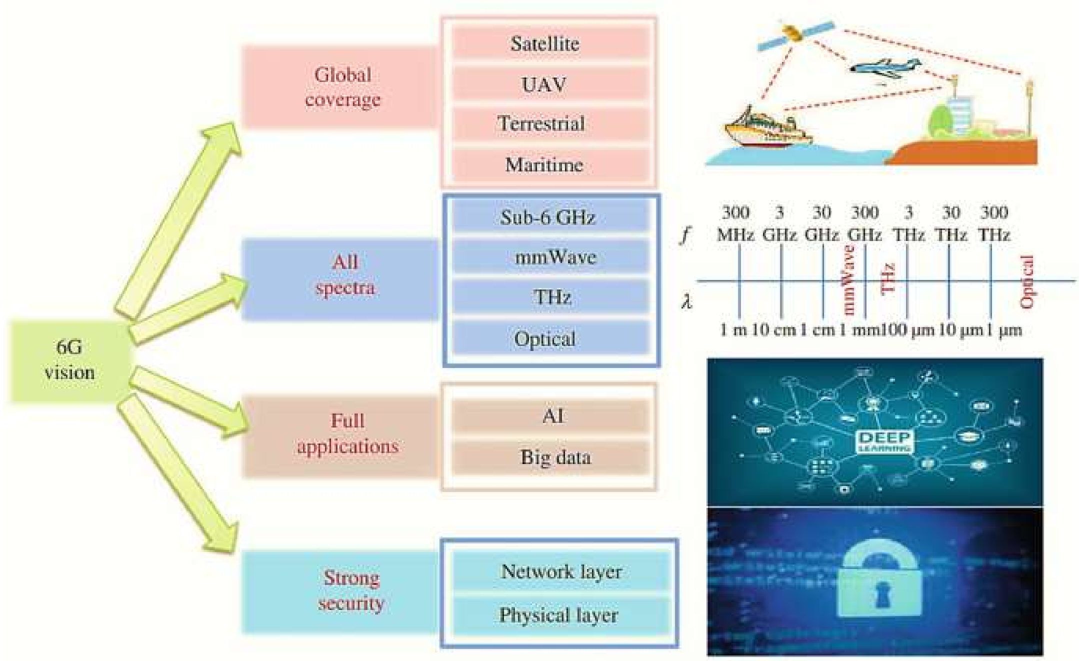


Fig. 6.7 A 6G wireless communication network schematic

Thirdly, the implementation of artificial intelligence and big data technologies will make it possible for a multitude of intelligent applications to be developed using 6G networks. The utilization of extremely diverse networks, a wide range of communication conditions, a large number of antennas, broad bandwidths, and new service requirements are all factors that contribute to the vast datasets that these applications will assist in managing the management of. In the fourth place, in the process of constructing 6G networks, strengthening network security will be an essential component. Given that it has higher technological needs than 5G, it should not come as a surprise that 6G will make it possible to build connections that are both wider and quicker. This will, in the long run, cause the line that separates the real world from the virtual world to become blurrier.

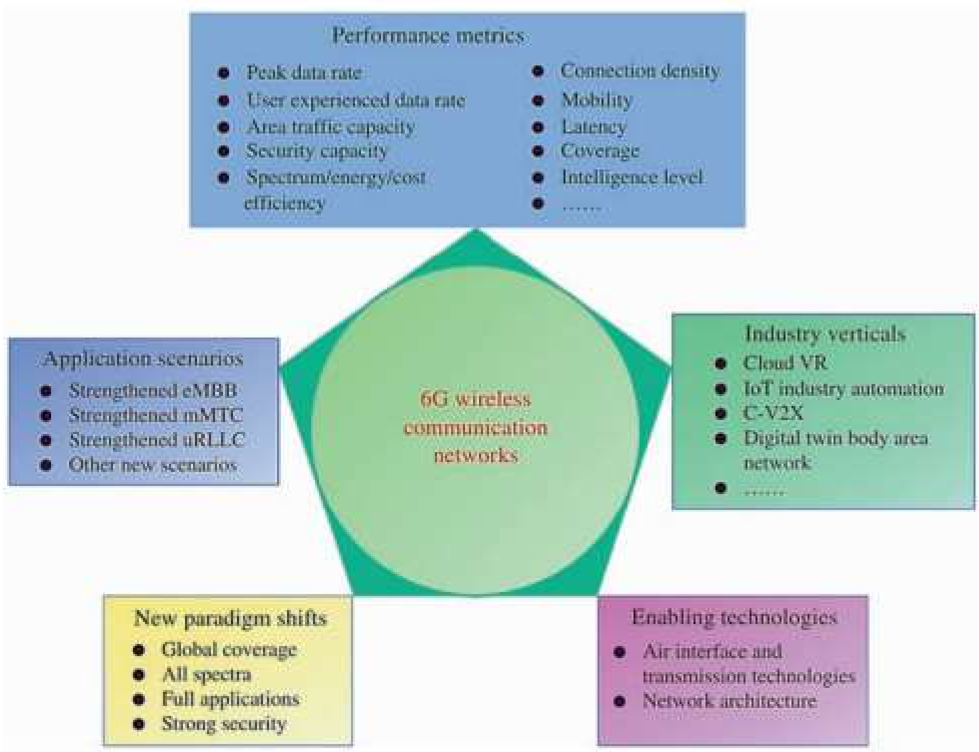


Fig. 6.8 A synopsis of wireless networks supporting 5G

6.3 CHALLENGES AND PROSPECTS OF CORE MATERIALS AND COMPONENTS FOR 5G AND BEYON

5G has evolved as a unique driving force for pioneering innovation and kindling new forms of information consumption. In addition to being an emerging engine for promoting industrial upgrading and maintaining economic growth, 5G has also emerged as a unique driving force. With the construction of 5G network systems and the increasing prevalence of 5G terminals, there has been an increase in the need for components such as base station antennas, filters, and communication PCB boards. 5G low-loss magnetoelectric functions, 5G metamaterials, and ultralow-loss high-reliability copper clad laminate materials have all been developed from a technical standpoint. These materials are used in 5G high-speed communications (56 Gbps and 112 Gbps).

As an additional point of interest, the manufacturing of optoelectronic integrated cables for 5G communication, multimodule integrated printed circuits for 5G communication base stations, and multichannel high linearity and large dynamic range RF optical transceiver integrated modules for 5G mobile communication will all prove to be essential in a wide range of applications that extend beyond 5G. Despite the fact that the core standard for 5G has been developed, there is currently continuing effort to further improve and extend it. Without a doubt, the core components, techniques, and technologies that make up 5G wireless networks have not changed.

6.3.1 Ultralow-Loss High-Reliability Copper-Clad Laminates

It is necessary that the PCB substrate material be designed for high-speed transmission in order to support future 5G applications. Copper-clad laminates, which have losses that are very low and great dependability, are the most important substrate material for 5G printed circuit boards. It is essential to exercise control over the dielectric properties of the copper-clad laminate, as well as the glass transition temperature, the thermal breakdown temperature of the functionalized polyphenylene ether resin, and other performance variables:

- ≥180 °C is the glass transition temperature (Tg) and ≤3.50/0.0025 is the dielectric performance (RC56%, 10GH).
- Z-axis expansion coefficient Z-CTE < 2.5%
- Copper foil peel strength (1OZ HVLP3) > 0.7 N/mm

The selection of glass fiber cloth, the selection of flame retardant, the creation of innovative fiber types and content, and the technique of free radical curing are some of the major technologies that are used. Controlling the amount of formula resin allows the created prepreg to demonstrate excellent processability throughout the PCB processing stage. Additionally, it is capable of meeting the requirements for a 5G high-speed ultralow-loss electronic circuit substrate needed for wireless communication. In addition to this, it demonstrates low volatility and good manufacturability, which allows it to attain the optimal mix of dielectric characteristics and heat resistance often seen in high-speed substrates.

6.3.2 5G Metamaterials and Low-Loss High-Performance RF Technology

Electromagnetic metamaterials that have been artificially processed or synthesized exhibit distinctive electromagnetic properties and have a structure that is either periodic or quasiperiodic. The electromagnetic properties of supernormality, the physical attributes of supernormality, and a one-of-a-kind artificial structure are the three most important characteristics that they possess. Because of the one-of-a-kind physical properties that they possess, metamaterials have the potential to be used in a wide variety of environments.

Electronic countermeasure radars, meta material communication antennas, absorbing materials, and meta material radar antennas are some examples of such devices. In the context of fifth-generation (5G) and potential sixth-generation (6G) large-scale satellite communications, meta material communication antennas provide advantages that cannot be matched by standard antennas. Consequently, the development of metamaterials that are of cheap cost and can be produced in large quantities is of crucial importance for 5G and other future technologies. When compared to other materials that are known to be equivalent, the meta material is rather lightweight. Additionally, the raw components are easily available, the process is relatively simple, and it can be made completely autonomously. Metamaterial with essential technical parameters for 5G radio frequency:

- The volume is less than 10%;
- The yield rate approaches 95%;
- The density is 45 kg/m3,
- and the dielectric constant ranges from 1.04 to 1.85.

The present 141 coaxial cable, for instance, is unable to deliver a more stable phase and has an unacceptably high loss of around 1 dB/m in the high frequency band of 3.3–5.0

GHz for the 5G antenna array. This loss is unacceptable. Because of this, the wireless antenna array for 5G makes use of waveguide feed technology in order to reduce losses and provide a consistent phase. This meta material, on the other hand, has ideal dielectric properties and may be built with heat-dissipating channels for use in applications that involve high electromagnetic wave power.

Additionally, the meta material displays a level of minimum loss, with its insertion loss being around 10^-4. With the usage of a meta material lens antenna, the principal equipment of the base station may use up to fifty percent less power than it would otherwise. Additionally, there will be a broader range of devices available for the principal equipment of the base station, chips and component devices that are more affordable, and a significant increase in the number of source channels.

6.3.3 5G Low-Loss Magnetoelectric Functional Materials and Devices

Low-temperature covered RF ferrite ceramics, also known as LTCCs, are seeing a surge in demand as a result of the rapid development of 5G communication technology and the increasing need for various high-performance, miniaturized, and integrated RF electronic components alike. It is proposed that the "seawater lava method model" be used to optimize the simulation of its doping, nanocrystal implantation, domain-domain transfer pinning, and other processes in order to achieve the low temperature of broadband low-loss LTCC materials (900 °C) co-fringing. This is done in order to produce the optimal formula for broadband low-loss LTCC materials. In order to expand the frequency range of the LTCC material to the microwave/millimeter wave area, the grain boundary resistance widening band model and the electric domain pinning model, which are both components of the loss model, are used. The establishment of unique effects of multiphase recombination and cooping is necessary in order to construct glass-dielectric recombination or magnetic-dielectric composite material systems. The realization of multiperformance low-loss LTCC material technology improvements in several frequency bands is made possible as a result of this.

In addition to the basic ceramic strip technology, devices such as radio frequency (RF) and microwave filters, antennas, switches, waveguides, circulators/isolators, and high frequency inductors have also been developed. Since this is the case, low-loss magnetoelectric functional material technology can be utilized to develop LTCC gyromagnetic (30–100 GHz) and millimeter-wave low-loss microwave ceramic material systems. This will allow for the support of materials in the forthcoming 6G communications and the realization of multi-performance LTCC composite materials in the radio frequency band, microwave band, and millimeter wave band. It is envisaged that the updated LTCC approach will result in the simplification of structures, the improvement of integration, and the reduction of prices for components such as filters and other combination modules. Furthermore, passive and silicon-based active integrated systems or components may be developed on miniaturized multiple chips or components for the purpose of utilization in radar and microwave communication applications.

6.3.4 Multimodule Integrated Printed Circuit Boards

The low-frequency medium mixed voltage approach is employed in order to achieve a multiband transmission architecture that is both cost-effective and efficient. In addition, the technology of low-profile copper processing is being developed in order to enhance the integrity of high-frequency signal transmission, and the technology of line etching is being optimized in order to manage single-line impedance faults. Integrated printed circuit boards that include many modules are subject to the following primary technical specifications:

- Board warpage ≤0.8%; through-hole thickness-diameter ratio ≥ 10:1; layers 8–18; thickness 1.2–4.0 mm
- Error in back drill depth of < 5%; single wire impedance fluctuation ± 10%
- ~75 μm flatness on the board and copper block surfaces, ≥300 N (30.6 Kgf) copper block push
- 288 degrees 10 s/3 thermal shock or lead-free reflow soldering 3 times, the product will not delaminate and alter colour.
- Signal transmission insertion loss > −45 dB/m@20 GHz

The creation of high-performance multimodule integrated printed circuit boards can be accomplished through the development of electrodeposition formulations, the improvement of system integration of functions, the development of in-board resistance, the improvement of the uniformity of high thickness-to-diameter ratio through-hole plating, the locally embedded copper blocks into high thermal conductivity composite materials to achieve efficient thermal control in the board, and the creation of product systems, performance testing procedures, and evaluation mechanisms.

6.3.5 Manufacturing Technology of Photoelectric Integrated Cables

In order to establish a long-distance connection between the signal and the power source in 5G communication, a large bandwidth and a high power are required. The photoelectric integrated cable is the most suitable choice for establishing a wireless connection at 5G speeds. However, the manufacture of photoelectric integrated cables is challenging because to the complex application scenarios of 5G, which also come with a number of new requirements. This makes the procedure tough. Despite the fact that there is a global unification of 5G communication standards, several suppliers of communication equipment have developed their own unique specialist solutions. By way of illustration, butterfly optical cables are not only simple to manufacture but also have the potential to be made smaller. The application of precise symmetry and tight cable forming technique allows for the creation of a product structure that is not only stable but also has high tensile and compressive strengths, is difficult to bend or twist, and is able to survive a wide range of construction conditions.

As a consequence of the fact that the fiber fusion splicing approach is not practical in many complex application scenarios, the development of critical technologies is important to the achievement of success in on-site quick splices. With the help of multilayer tight cladding technology and precise size control, it is possible to join the optical unit and the fast connection in a way that is both highly reliable and well matched. In addition, the high power and miniaturization of the 5G equipment have resulted in the production of a considerable quantity of heat. Consequently, the use of materials that are resistant to high temperatures is necessary in order to develop a cable communication system that is trustworthy and can be utilized for a long length of time while being subjected to high temperatures.

6.3.6 Photonics-Assisted Ultrabroadband RF Transceiver Integrated Modules

In order to fulfil the most important criteria of ultrabroadband dynamic analogue transmission and low power wireless access in 5G mobile communications, a multichannel High Linearity and Large Dynamic Range RF Optical Transceiver Integrated Module has been developed. It takes use of a wide range of services that are both variable and diversified in order to concentrate on the new requirements and difficulties that the microwave and light wave fusion access system presents.

For the purpose of accomplishing multiple services, large dynamic, long-distance fiber distribution, and multidomain resource centralized management and control of a new generation of microwave optical wave fusion access and ubiquitous spectrum sensing systems, it is necessary to have support in the 20 GHz frequency band that surpasses four multiband, multistranded wireless signal multichannel (≥4) integrated optical transceiver modules. The MZI-PLC silicon-based integrated coupling structure, the 42-degree inclined array waveguide coupling method, and the 3D integrated microwave package design method for electrical design, thermal management, and reliability analysis are all methods that can be utilized in the process of fabricating high linearity and large dynamic range RF optical transceiver integrated chips and modules.

An integrated radio frequency (RF) optical transceiver module that operates in the 1310 nm band was developed by using the high-saturation-power photodetector array chip and the built multichannel (‖4) laser array chip. A 3 dB bandwidth that is more than 20 GHz and an SFDR that is greater than 95 dBHZ2/3 are both characteristics of each channel, which contribute to an increase in the number of users who have interior coverage. The realization of an integrated optical transceiver module that is multiband, multistranded, and multichannel (≥4) for wireless communications has been shown to be feasible. This module offers more than four services inside the 20 GHz frequency band. The next generation of microwave optical wave fusion into spectrum sensing systems that are extensively employed, with a massive dynamic long-distance fiber distribution and multidomain resources that are controlled centrally.

The primary chips and components of high-performance RF optical modules are responsible for the realization of broadband, high linear, high power, multichannel analogue direct modulation laser technology. These modules also realize broadband, high linear, low insertion loss modulation, as well as the overall improvement and practicability of optical link performance indexes. The multimode integrated radio frequency (RF) optical module, the low noise analogue optical amplifier, the high-power broadband photoelectric detecting array reception, and the array technologies have all been achieved.

6.3.7 All-Optical Network and Super Large-Core Fiber Optic Cables

The traditional structure of optical fiber extraction, fiber drop, and connection work is greatly simplified by the skeleton optical cable due to its many advantages. These advantages include high density, simple installation of optical fiber ribbons, convenient offline or branching, and a structure that is completely dry, water-blocking, and moisture-proof. Given the characteristics of the 5G technology's own communication architecture and large data transmission capacity, the most pressing product problem that needs to be solved in the construction application of the skeleton optical cable is the large core number expansion of the skeleton optical cable. This is in addition to the optical fiber expansion problem that exists in the last one kilometer of the FTTH entrance of the all-optical network. Significant technological advancements have been made in the areas of fiber coloring technology, skeleton tension control technology, optical fiber in-groove technology, water blocking tape wrapping technology, and optical fiber and tape technology. These advancements have helped to address issues pertaining to fatness quality, production bottlenecks, and the cost of auxiliary materials (such as resin).

It is difficult to monitor the operations of the production line, including the data equipment, product quality, energy consumption, and other activities. It is required to determine the connection between the initial design and transfer learning in order to construct a simulation solution interface for modelling and data fusion. This is a prerequisite for the creation of the interface. This is accomplished by merging the various techniques to the simulation solution. By doing so, we will be able to facilitate the creation of production line design simulation systems that are both extremely dependable and efficient. Multicore high-density banded fiber optic cables will serve as the focal point of the commercial carrying optical networks that will be used for 5G. It is for this reason that 5G is entirely in line with the patterns of future social development, economic growth, and human progress, and it is certain that its uses on a vast scale will be achieved. Within the context of the 5G era, the Internet of Things will be expanded and made available to all sectors and industries, as well as to every aspect of the economy and society. This will result in a change in the manner in which individuals produce and consume products and services, as well as the provision of a powerful tool for the enhanced development and administration of society.

CHAPTER 7

GREEN IOT NETWORKS USING MACHINE LEARNING FOR 5G NETWORKS

7.1 INTRODUCTION

One example of the Internet of Things (IoT) is a prototype that aims to accomplish everyday objectives and upload them to the internet by using a system that is based on a microcontroller (µC) and consists of a transmitter and receiver component. In addition to encouraging commercial enterprises and academic institutions, the prototype places a priority on the computerization of a number of applications that have the largest potential practical effect. There is a delicate function that the Internet of Things plays in the workouts that we do on a daily basis. The technology is engaged in a variety of activities, including intelligent transportation, which offers substantial breakthroughs in transportation and administrative organization, and smart homes, which simplify the activities of occupants. The Internet of Things is used in a broad variety of sectors, such as industry 4.0, agriculture, intelligent traffic, medical services, and even everyday things such as clothing and pens. As the society has seen, the quantity of these gadgets has therefore expanded by a significant amount.

According to IHS Markit, by the year 2030, it is anticipated that over 125 billion Internet of Things devices would be linked to one another. These estimates are significantly impacted by the introduction of additional range frequencies as well as the growth of 5G networks. On top of that, Internet of Things devices are often rather tiny and have limited battery life, and the enormous volume of information that is sent between them results in substantial energy needs. In many cases, Internet of Things devices are unable to fulfil these criteria, which may rapidly result in increased energy consumption and network disruptions.

In addition to lowering the amount of energy that is used by Internet of Things networks and the carbon imprint that they leave behind, what are some energy-efficient solutions that will also increase the lifetime of the network? In a similar vein, we are seeing the rise of the Green Internet of Things, which is mainly concerned with the regulation of energy inside Internet of Things networks in order to lessen the amount of power that is used and to lengthen the lifetime of Internet of Things networks. The future will bring forth substantial breakthroughs in our day-to-day lives, and it will also assist us in comprehending the notion of "green ambient intelligence."

A wide variety of energy collecting and conservation methods will be discussed in this chapter, as well as the development of a new Green Internet of Things trend. Some examples of this trend include social green Internet of Things, 5G green Internet of Things, digitalization of green networks, backscatter communication, and many more. In this chapter, the format is as follows: The second part provides a description of the literature that pertains to power management in the Internet of Things ecosystem. In the third part, we discuss the difficulties that arise from the use of electricity in networks. There are a few methods of energy conservation and energy farming that have been documented in the study literature, and these are the topics that are covered in the fourth part. The management of electricity for the ecosystem of the Internet of Things is the topic of discussion in the fifth part. Our discussion of developing research ideas and practical applications for energy management in the Internet of Things is presented in the sixth part of this research paper. We will conclude this chapter by offering an outline of the future of environmentally friendly Internet of Things technology.

7.1.1 New Developments in the 5G IoT Ecosystem

An exhaustive investigation into the issue of energy management for the Internet of Things has been carried out, and several audits and studies have been published over the course of time. Wireless sensor networks (WSNs) that are efficient in terms of energy consumption are of significant interest to academics because of their widespread use in the Internet of Things. approaches for power farming and battery-driven power preservation were provided by the researchers. These approaches are used in the construction of algorithms for wireless sensor networks (WSNs).

The primary focus of the study was on analyzing the methods that are used to regulate the power of WSNs. The issue of sensors and other devices that are linked to networks that are part of the Internet of Things is discussed in this chapter. An investigation of a variety of power management, power efficient, and power saving approaches for wireless sensor networks was carried out by Singh et al. (2020). The researchers focused on regulating energy transfer and a variety of techniques specifically. An investigation of the power farming strategies that are used in wireless sensor networks was carried out by the researcher with the objective of extending the lifetime of these networks. This chapter provides an overview of a number of power farming and power saving approaches that may be used for expensive assets such as sensors and electronic devices.

7.1.2 Green IoT Enabling Technologies

In our day-to-day lives, we make use of Internet of Things facilities, and the number of IoT devices that we possess contributes to the carbon footprint that these uses leave behind. Devices that had both ample and restricted power resources were responsible for the creation of Internet of Things networks. One of the key objectives of the Green Internet of Things is to lessen the amount of power that both devices use while simultaneously guaranteeing that their features and capabilities are operating as they were designed to. Important and pervasive Internet of Things technologies, such as wireless sensor networks (WSNs), are made up of tiny sensor devices that are powered by batteries and collect data in order to construct reliable and effective data and communication devices.

In spite of this, these devices are often used in Internet of Things applications for the purpose of carrying out continuous as well as long-term monitoring and control duties. In addition, sensors may be located in regions that are difficult to access (for example, subterranean networks, military sites, dangerous conditions, and so on), and they may soon run out of power. It is thus more expensive and time-consuming to replace batteries on a regular basis as a consequence of this. The ongoing availability and functioning of Internet of Things devices should be ensured by power capacity and energy management. This is because, in extreme circumstances, neither the replacement of batteries nor the replacement of cables is realistic solutions.

7.2 IOT ECOSYSTEM ENERGY MANAGEMENT TECHNIQUES

The collection of electricity and the conservation of power are two methods that are often used in Internet of Things businesses to manage electrical power challenges. The development of hardware solutions, protocols, and algorithms is the primary emphasis of energy conservation plans. The goal of these plans is to decrease the amount of power that is used by networks and to extend the lifetime of systems. On the other hand, power collection may be described as a method that is used to generate energy from the surroundings of the network in order to provide network devices with a consistent source of power. Figure 7.1 shows unique techniques to power management that optimize the lifetime of Internet of Things networks and lower the carbon footprint of that network. These approaches are based on these two categories.

There are several ways to classify power-saving measures according to the technology that they use. In order to lessen the quantity of power that is used by Internet of Things networks, two primary technologies are often utilized: deep learning and machine learning. The methods for sleep and wake are the two key technologies that are utilized for the purpose of power saving. The other technology is data collection techniques that are power efficient.

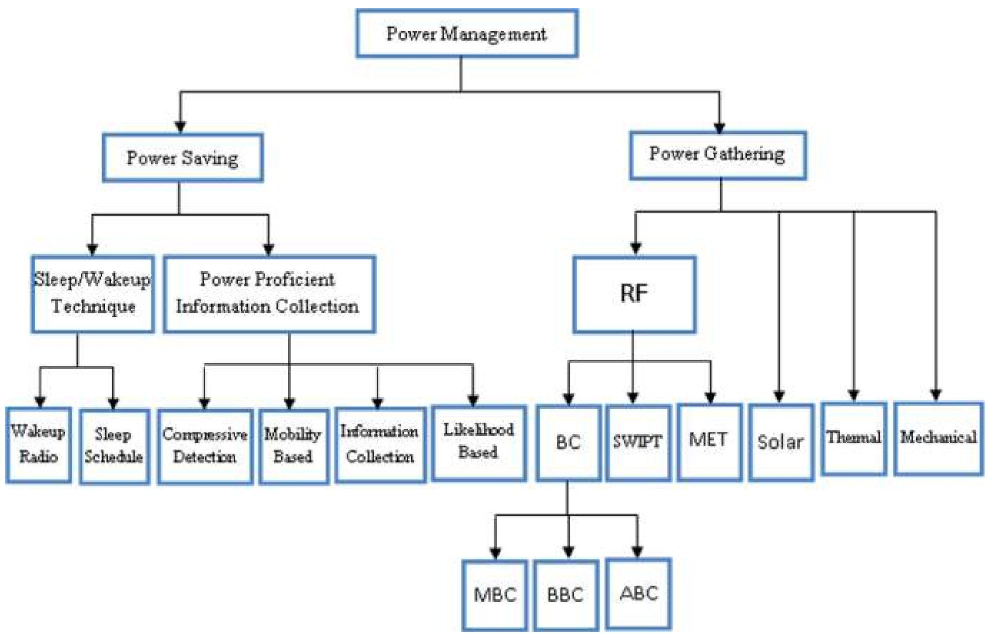


Fig. 7.1 Innovative techniques for electricity management

Mechanical power, thermal power, solar power, and radiofrequency power collecting are the four categories that fall under the umbrella of power collection processes.

7.2.1 Power Saving Techniques

Wakeup Radio:

The Wakeup Radio hardware technology is an innovative approach that aims to decrease the amount of power that devices use and to increase the lifetime of networks. There is a low-power radio that is connected to the device that makes up this component. It receives a notification from the primary radio whenever it detects an incoming transmission. An Internet of Things Wake-up Radio, which is a µC module, is connected to a wake-up receiver and a main radio via a connection. The primary radio and the µC are often found to be in a deep sleep the most of the time. As can be shown in Figure 7.2, when the receiver receives a call via a wakeup warning that is sent by the external hub, it creates an interference wave that triggers the µC to carry out the prescribed job. Additionally, the receiver transmits the data through its primary radio transmitter.

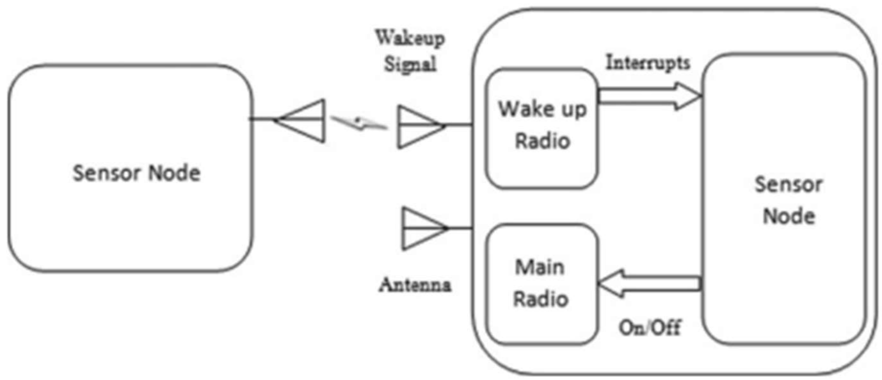


Fig.7.2 Wakeup radio

Semi-active wakeup radios, passive wakeup radios, and active wakeup radios were described by Piyare et al. (2017) as the three distinct forms of awaken radios that were classified depending on the amount of power that was used. There is a need for active wakeup radios to have batteries or some other kind of continuous external power. This means that passive wakeup radios do not need an external power source to function. The power that they possess originates from the signal that is being sent. Semi-active wakeup radio is characterized by the fact that the radio frequency (RF) front-end is not active, and a few components on the receiving end need a constant external energy source.

Among the many potential applications of wake-up radios, industry warehouses, smart homes, smart cities, and smart workplaces are all possible. On the other hand, the wakeup radio's waking and sleep wake-up capabilities have led to an increase in the amount of idle time spent gathering information while listening to the radio. When building a wake-up radio for use in a vast climate, several design views are taken into account when designing the radio:

(a) In order to maintain the balance between the power that is saved and the power that is used by the device, the center's energy consumption cannot be more than the sleep energy of the primary radio.
(b) In order to minimize the amount of time spent in idleness and ultimately maximize the number of applications for wakeup radios, it is recommended that the time required to wake up the device radio be as short as possible.
(c) A hub or many hubs may be unintentionally initialized, which results in a significant amount of power being wasted. This may occur as a result of interruptions and the improper transmission of signals that were intended for various devices. Therefore, planners need to determine the precise difference and find a solution to the problems without increasing the amount of energy that is being used.
(d) The sensitivity of the receiver hub has a significant impact on the amount of energy that is required in the Wakeup Radio environment. This is because a high level of sensitivity necessitates a large amount of energy to be present at the receiver, while a lower level of sensitivity necessitates a high amount of radiating energy to be present at the transmitting end.
(e) The Wakeup Radio's range of application has an impact not only on the kind of context in which it is used but also on the amount of energy that is consumed by the system. Multi-hop communication is required in small-range communication, for example, in order to increase the density of the hub and the amount of energy that is effectively used.
(f) Information rate: The greatest approach for improving power efficiency and wakeup idleness is to have a high information rate in the wakeup radio. In a similar vein, the maximum bit duration is required in order to guarantee the dependability of the wakeup signal and to expand the communication range.
(g) In order for Wakeup Radio to be included into the Internet of Things and existing sensor hubs, it is necessary for it to be commercially viable. It is recommended that the cost not exceed five to ten percent of the overall sensor hub costs.
(h) Wake-up radios should be created with frequency guidelines of the ISM band in mind throughout the design process.

7.2.2 Sleep Schedule

It is believed that the duty cycle, which is also known as the sleep schedule, is the most efficient approach for prolonging the life of networks and saving energy. In order to limit the amount of power that the hub consumes from its battery, sleep scheduling algorithms are used. These algorithms switch between the awake and sleep states. Typically, the hub is in a low-energy mode and will only wake up when it is required to do so. The topology control protocol, the wakeup/sleep protocol, and the MAC protocol are the three protocols that are often used to categories these approaches.

According to Passos et al. 2019, when an asynchronous sleep schedule is used, a significant amount of transmission power is consumed while waiting for the receiver to wake up via the usage of several unicast broadcasts. This results in network congestion and an excessive amount of power consumption within the system. There is a protocol that may be used in Internet of Things networks that use Wi-Fi to increase the sleep schedule length of small energy Internet of Things devices that are equipped with power gartering circuits. This technique makes use of an ideal weighting moving-normal filter to ascertain the optimal weighting variable for the purpose of evaluating the span of off periods of Internet of Things devices in order to approximate the amount of power that has been accumulated. The gadgets will make advantage of these periods of silence in order to gather sufficient energy to protect the information.

The Internet of Things precision horticulture industry now has access to yet another cutting-edge sleep schedule expert track producing algorithm. This method is used for the purpose of information transmission, track design, and algorithm analysis on all tracks, with the smallest track being selected for inclusion in the study. It is also possible to utilize this method effectively with a variety of other devices. Nevertheless, the efficacy of this strategy is contingent upon the groups in order to lessen the total power consumption that is required to enhance the connection between the respective groups.

7.2.3 Capable and Effective Data Gathering

Some examples of smart applications that make use of Internet of Things devices include smart housing, smart transportation, and others. These gadgets generate an extraordinarily large quantity of data. A significant amount of power will be required by an Internet of Things device in order for it to communicate its data to its base station and to other devices in the vicinity. To add insult to injury, Internet of Things networks often include a large assortment of devices that have a short battery life. When batteries are dead or empty, it is tough to replace them, and this is particularly true if the devices in question are located in a particular location that is difficult to access or very remote.

Obtaining the information that is created and distributing it in an efficient manner is a fantastic method for lowering the expenses associated with gearbox power. A great number of methods for acquiring information have arisen in order to lessen the amount of power that is used, to extend the lifetime of the network, and to delay:

(i). **Compressive Detection Based Method:** Within the framework of the compressive detection-based technique, the pace at which the network transmits information is slowed down, and the quantity of information that is gathered is compressed. Through the use of compressive detection, the size of the information matrix of the sensor may be reduced, which in turn minimises the amount of data and transmission that is repeated. Conventional compression, distributed source coding, and distributed compressive detection are the three kinds of approaches that fall under the category of compressed detection-based techniques. In the realm of information compression methods, some of the qualities that distinguish them include the capacity to represent information, the difficulty of computation, memory management, and compatibility. In order to alleviate the complexity of devices and enhance the effectiveness of Internet of Things networks, information compression strategies have to be used to enhance these characteristics. According to Wang et al. 2019, the clustered routing technique is yet another strong compressive sensing-based protocol that is easily accessible. A reduction in the "hot spot issue" and an evaluation of the appropriate cluster size and dispersion of cluster heads are both achieved using this technique. By using this method, the total power consumption of the network is successfully decreased, and its lifetime is effectively extended. Given the spatial-temporal connection, the temporal link between the data implies that it could be useful to employ a predictive approach to minimise power usage. This is because of the spatial-temporal relationship.

(ii). **Mobility Based Information Collection:** The use of mobility-based information collecting techniques is becoming more common in today's world as a means of fostering power-efficient Internet of Things networks and increasing their lifespan. The use of this technology enables battery-restricted devices to save the gearbox power of the device, which ultimately results in an increase in the device's lifespan. There are three viable options for achieving this goal. In the first place, the strategy involves making use of mobile sinks and static sensor hub assemblies. In situations when they are sufficiently close to hubs, they work together to collect information from those hubs. In addition, the approaches include mobile hubs as well as a static sink. Because the hubs are mobile, they are able to further improve the network's coverage while simultaneously reducing the number of hubs that are necessary. Thirdly, the method integrates mobile sensor hubs and a portable sink into its overall structure. There is a technique that is reliant on the internet of cars that allows for the design of traffic information collection with an efficient sink hub option. In order to reduce the amount of power that is used by the network, this technique demonstrates which cars and methods should be selected. Mobility-based information gathering is an exceptionally promising way for conserving power, provided that carbon imprints are allowed by mobile sink and that it is given an upgraded orientation.

(iii). **Information Aggregation Method:** This strategy involves minimising the network transmission rate and eliminating unnecessary and repeated transmission of information. It is referred to as the information aggregation approach. The process of information aggregation involves one or more hubs coming together to collect information from a variety of hubs over a period of time, then combining and transmitting the resulting information. According to Chandnani and Khairnar 2020, the information aggregation techniques may be classified into three distinct categories: the tree-based information aggregation method, the cluster-based information aggregation method, and the centralized-based information aggregation approach. A minimal amount of power is used by the tree-based information aggregation techniques, which also have a long network lifespan and an incredible degree of adaptability. Methods of information aggregation that are based on clusters have a low traffic volume and a large capacity for adaptability to internal failure. According to Purgeable and Navimipour (2017), centralised information aggregation systems provide a general emphasis on security. Hubs are arranged in a tree-like fashion to form the tree's structure. Leaf hubs and sending hubs are both components of a tree. Compared to the other two methods, this one consumes far less energy. On the other hand, the insufficiency of the intermediate hub has an effect on the whole topology and has an impact on the functioning of the network (John and Jyotsna 2018). For the purpose of selecting the µC and operating a variety of sensors, the cluster-based information aggregation technique employs a sensor-based group foundation algorithm. Using this strategy, the user is able to extend the lifespan of the network, increase the significance of the inclusion, and stop the repetition of information at a temperature of µC. However, the network's vulnerability is increased by these tactics. These techniques also take into account the propagation of the disease and the inertness of sleep (Song et al. 2017). In the centralized-based information aggregation approach, information is gathered and then sent to the centralised gateway of the Internet of Things (IoT). For the purpose of ensuring that the information on the network is consistent, these strategies include an efficient sleep scheduling strategy that also includes information aggregation. In comparison to the other two methods, the centralized-based information aggregation approach is more secure; nevertheless, this method has a limited capacity for adapting to internal failure.

(iv). **Likelihood based method:** A reduction in power consumption is achieved by the use of estimate models in the likelihood-based technique, which is used to ascertain the link between the information. Approaches that are based on probability have as their main objective the prediction of the value of the data that is sensed. According to Razafimandimby et al. (2018), the Bayesian technique helps to minimise the amount of power that is used by Internet of Things networks and prevents the transmission of significantly correlated sensor data. By using this method, a hierarchical structure was created that included the use of an information hub as well as smart gadgets. In three distinct cases, this method minimises the quantity of information that is delivered while still maintaining a high level of data accuracy. Statistical methods that are based on probabilities need the use of complex apparatus, and they do not take into consideration the significant relationships that exist between the data collected by various sensors. As a result of these issues, this strategy is tough to implement.

7.3 POWER GATHERING METHODS

The primary categories of power collecting methods include mechanical, thermal, solar, and radiofrequency (RF) methods. These are the four primary types of power collection methods. A description of each of these several types of methods is provided below.

7.3.1 RF Power Gathering

In addition to its more common term, the RF power gathering method is also known as the wireless energy transfer technology. In light of the fact that the backscatter technique is becoming an increasingly popular way to power collection, these tactics concentrate mostly on it:

(i). **Backscatter Communications (BC):** The fact that Internet of Things devices are dependent on batteries is the fundamental problem with them. This dependency results in IoT networks having relatively limited lifespans. Due to the fact that these gadgets have long life lives, they are susceptible to power depletion. At this point, the expense of replacement begins. In addition, the sensors are not always functioning properly since the battery dies more quickly than expected. Wireless signals have the potential to serve as both a source of power and a mode of communication for electronic devices that are powered by batteries in the province of British Columbia. A gadget that does not need batteries decreases the amount of power that is used, eliminates the have to replace batteries on a regular basis, and avoids undesirable behaviour that is brought on by low power.

(ii). **Simultaneous Wireless Information and Power Transfer (SWIPT):** This is the most intriguing and up-and-coming invention in the field of power management. The simultaneous interchange of data and energy via wireless networks is made possible as a result of this (Perera et al. 2017). To put it simply, this makes it possible for an Internet of Things device to get RF data and power from the same RF signal that transmits both power and data into the device. An Internet of Things device may thus be self-controlled by receiving information. In general, however, it is not practical to do power collection and data decoding simultaneously from a signal that is comparable to the one being investigated. The data might be altered or lost if there is action involving the collecting of electricity.

Therefore, in order to achieve SWIPT, a number of different recipient models were taken into consideration, including Antenna Switching, Time Switching (TS), Separate Receiver, and Power Splitting (PS) . By concurrently transmitting the energy and data that are necessary for the Internet of Things hub, this strategy helps to keep the hub operational and extends the lifespan of the Internet of Things network. This may be accomplished by SWIPT via the use of a twin battery green power collecting architecture (Liu and Ansari 2019). It consists of a first battery that functions to provide power to the device, and a second battery that functions to collect electricity in order to reduce the possibility of a power shortage occurring.

However, this strategy is only applicable to small Internet of Things networks that include devices that are capable of supporting a dual battery. It is a difficult and costly approach to use for large Internet of Things networks. In order to meet the quality of service requirements of each and every customer, there is a method-based DL that may restrict the amount of energy that is sent by a multi-transporter NOMA SWIPT. In addition, DL was used in conjunction with SWIPT. There is a twofold mode SWIPT in this method, which makes use of a flexible mode altering that is reliant on DL in order to improve both single tone SWIPT and multi tone SWIPT. However, deep learning is effective for SWIPT-based devices; it also requires a large number of resources, which are essential for the development of an Internet of Things that is both environmentally friendly and easy to maintain.

It is hypothesized that SWIPT is an exceptionally effective approach for power collection. However, this approach presents a number of challenges. An analogous signal is used to transfer both data and power in the SWIPT transmission method. The signal has to be decoded in an efficient manner in order to recover the correct data. In the event of an interruption in the power collection process, this signal may undergo a change. There is a limitation to the range of SWIPT since the signal might get weaker while travelling over vast distances. Last but not least, signals that are sent to electronic devices may be hacked. To ensure the protection of private and personal information, it is necessary to include a certain level of security in this manner.

CHAPTER 8

5G AND BEYOND: ARTIFICIAL COGNITIVE COMPUTING FOR SMART COMMUNICATIONS

8.1 INTRODUCTION

The field of computing has been seen as a significant game-changer and a source of intrigue for academics ever since the early 1990s. Previously, computers were used to compile strategies such as calculators, which reduced the level of complexity that was necessary for human computing capabilities.

After that, the application era arrived, which was characterized by the fact that everything was managed and controlled by devices that were equipped with microprocessors. The age of cognitive computing is characterized by the discussion of data and the use of statistics to make sense of it. The extensive range of criteria for enhancing the customer experience at the same time in order to cover a broad variety of connections, huge connections via the utilization of a low-power, high-capacity hub, and improved dependability in system communications, which ultimately results in the survival and profitability of the network through wireless communication.

The study was carried out with the architectures of network and enabling management optimization, flexible configuration, and rapid reaction time as the foundations. The purpose of this study is to explore the components and architecture of cognitive computing, as well as smart communication, cognitive computing, and the influence of COVID-19 and 5G networks.

The field of study known as cognitive computing aims to accomplish the goal of imitating the cognitive processes of humans. The analysis of cognitive computing is presented in extensive detail (Fig. 8.1).

In the latter part of the nineteenth century, tabulating machines were widely used, which made it possible for significant developments to be made, such as the census report and the Social Welfare System in the United States. Programmable computing was originally introduced in the 1940s, and ever since then, it has made it possible for everything.

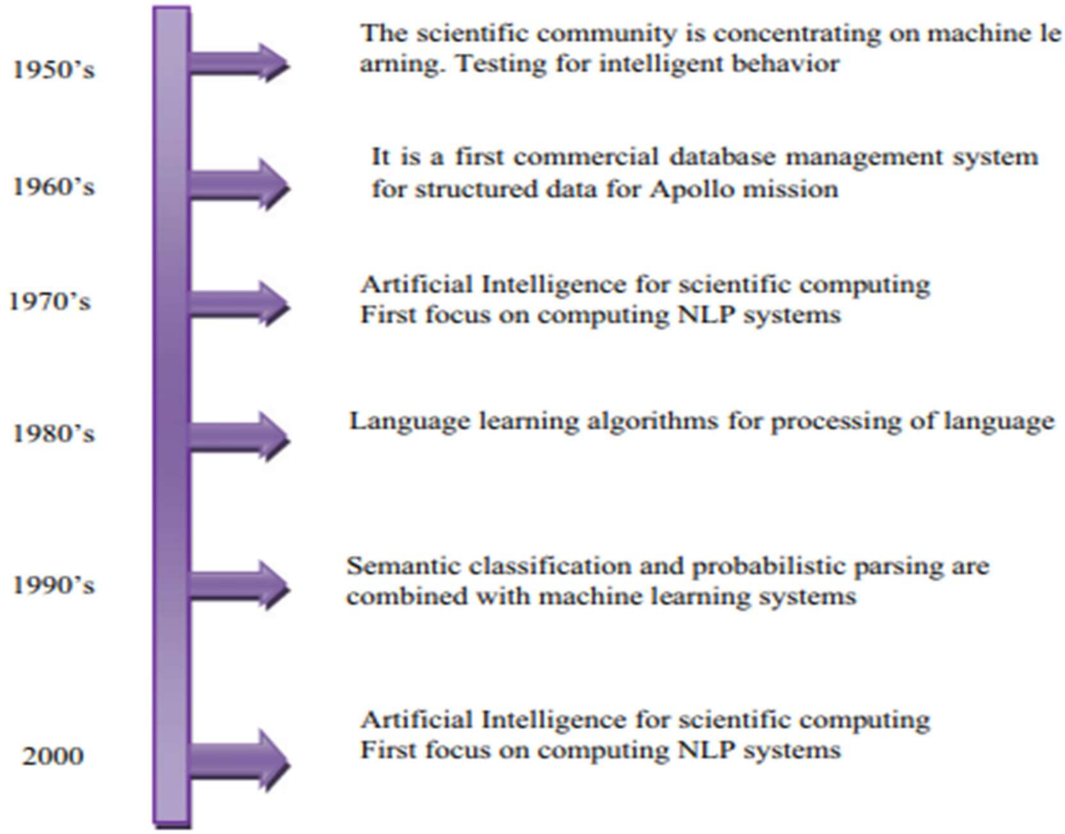


Fig. 8.1 The assessment of cognitive computing (over the last few decades)

Interplanetary exploration to the Internet and everything in between. (Fig. 8.2) illustrates the assessment of cognitive computing during the course of the last 10 years, specifically with regard to the present year.

The cognitive computing era may be considered the third most disruptive paradigm in the history of computing, behind the tabulating machines and programmable computers as the most profoundly disruptive paradigms. In order to facilitate the development of intelligent systems that are capable of self-awareness, artificial intelligence was developed. A comprehension of the systemic and semantic links was achieved by the cognitive processes via the identification of data from the text and utterances. In the field of artificial intelligence, natural language processing, machine learning, reasoning, and data mining, the 1990s provide early evidence of effective reinforcement learning. Within the realm of cognitive computing, Google Deep Mind and IBM Watson are now at the forefront of the field.

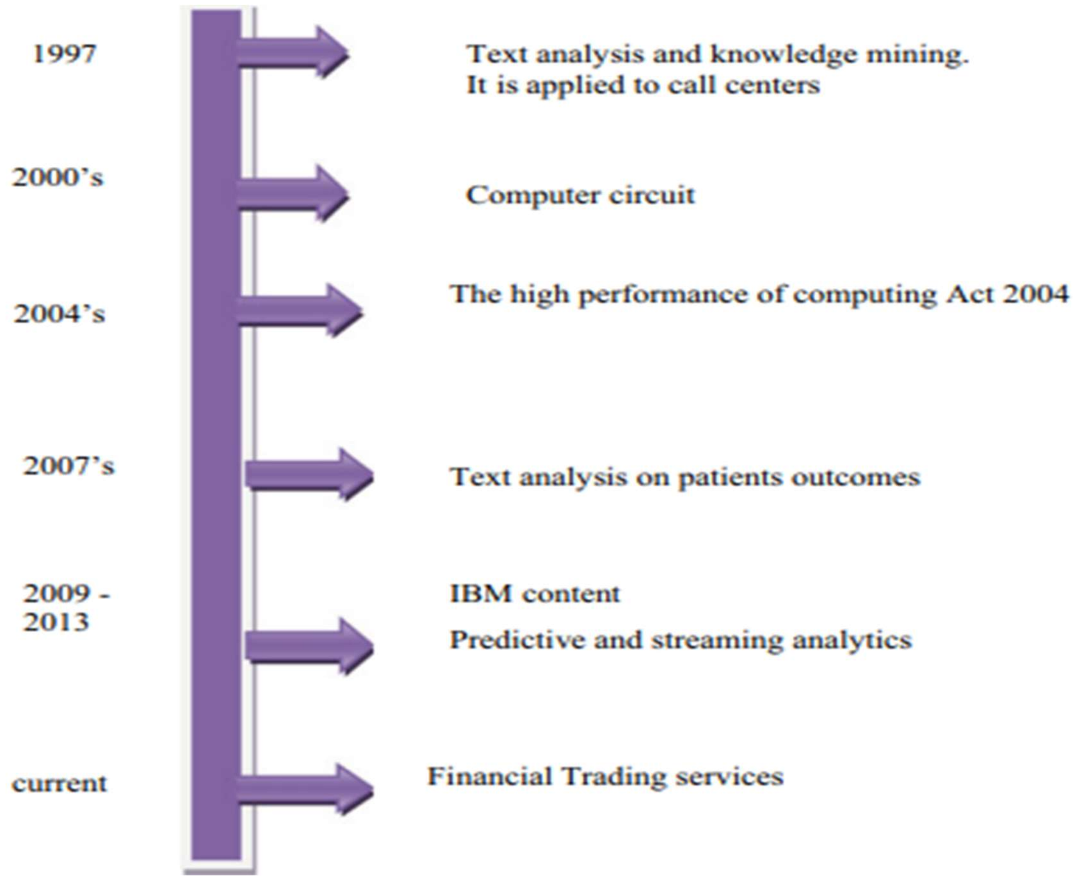


Fig. 8.2 An assessment of cognitive computing (from the previous 10 years to the present)

As a consequence of this, cognitive science spans a broad variety of subject areas and academic disciplines, such as anthropology, linguistics, artificial intelligence, philosophy, psychology, and neuroscience (Wang 2003). To a certain extent, the achievements of a scientist working in the field of cognitive science up to this moment are intrinsically tied to the use of interdisciplinary research approaches. Systems that use cognitive computing incorporate the most advanced aspects of a number of different technologies, such as natural language processing (NLP), real-time computing, artificial intelligence (AI), and machine learning.

Through the use of these technologies, a cognitive computing system is able to analyze vast quantities of semi-structured and unstructured data files. Cognitive computing offers benefits that are superior to those of data analysis. It also helps businesses to recognize possibilities and reveal obstacles in real time, and then analyze those opportunities and challenges in order to react more quickly and effectively. This, in turn, leads to an increase in the amount of interaction with customers.

8.2 THE COGNITIVE COMPUTING COMPONENTS

Cognitive computing may be defined as a set of autonomous and cognitive knowledge processing theories and technologies that imitate the mechanics of the brain in a manner that goes beyond the traditional forced data processing. There are conceptual and behavioral models that have been created for cognitive computing. The use of a powerful mathematical method known as label mathematics has been suggested as a means of addressing the design and implementation of cognitive computing systems. Applications of cognitive computing have been discovered in a broad variety of fields, including autonomous agent systems and intelligent search engines, amongst others.

The components that make up the whole cognitive figure architecture are as follows: The algorithms

- Algorithms are instructions that must be followed in order to create methods that are used for the purpose of resolving certain issues.
- Artificial intelligence (AI) is the process by which a machine can see or, more broadly, observe its own state and then figure out how to accomplish the objectives it was created to accomplish.
- The first iteration of artificial intelligence was gradually more fundamental; hence, information retrieval and cognitive registration are often considered to be basic forms of AI.

Reasoning and Decision Automation:

- This is a method that an intellectual data structure use in order to put its knowledge of where to attain its goals into practice.
- The meaning of the word "reasons" is "goal." Although it is not very similar to human cognition, it is designed to be similar to it.
- The result of a thinking method might very well be decision mechanization, which is when a program creates and executes a solution to a problem on its own without any human intervention.

Sentimental Intellectual Ability:

- For a considerable amount of time, psychological figures avoided making an effort to portray impassioned understanding; yet, interesting activities still took place.
- Affective, an invention developed at the Massachusetts Institute of Technology, is one example of an organization that is working towards the goal of processing the frameworks that were established in order to maintain a feeling of humanity via these signposts as attractiveness and then produce reactions.
- The purpose of intellectual figuring frameworks, which are outstanding in their capacity to sense emotional signals and display a remarkable similarity to a human being

8.3 THE ARCHITECTURE OF COGNITIVE

The architecture of a cognitive computing system is shown in Figure 8.3. This architecture is used to construct a variety of cognitive programs by using cloud-based solutions, tensor flow, and several database tools. When it comes to human computer interaction with emotional cognition, there are a few technical difficulties that are now accessible. The domain of computers is concerned with the consumption of information, while the communication sector is primarily concerned with the conveyance of data.

The majority of the knowledge that is conveyed in applications of cognitive computing that are used in the real world is expressed via data, which may be either unstructured or semi-structured. When it comes to the material world, the Internet of Things (IOT) gathers information that is both real-time and useful. It does this by establishing a vast network via the use of the World Wide Web and by enabling intercommunication amongst enormous wearable sensors in order to provide a bridge between the digital and physical worlds.

It was in the middle of the twentieth century when cognitive robots were developed. More than half a century has passed since the introduction of robot technology, and during that time, human production processes have undergone significant transformations as a result of these developments. Currently, this approach is a crucial marker for determining the level of methodical and technological development that a nation has reached.

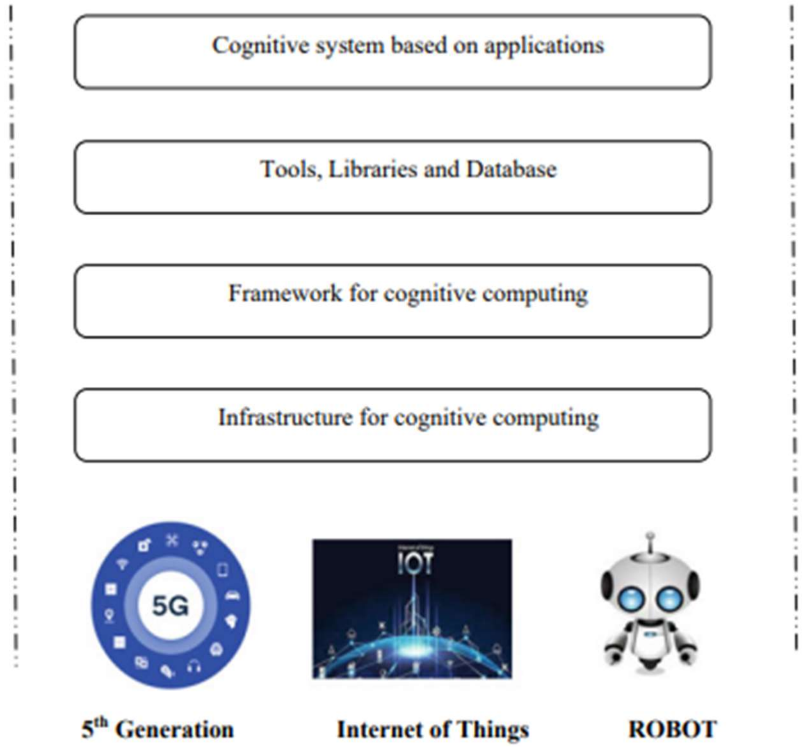


Fig. 8.3 The system architecture of cognitive

A process of upgrading and manufacturing of high-end work. This new trend in community ties is evidence that the future generation of robot systems will be able to visualize human beings in a variety of different ways. For example, in the form of a partnership relationship, in which people and robots cohabit in unity and make their respective benefits available to one another, this is an example of a partnership connection. The next generation of robots will be able to cohabit with people, which is a crucial characteristic for them to possess. The cognitive computing system architecture that is being given is built of many components, including the Internet of Things, cloud computing, and big data analysis.

The ability to comprehend one's feelings is a significant application of cognitive computing. Additionally, it is a ground-breaking method for interacting with human mechanisms. At the moment, modern human–machine interaction systems usually indicate that human–machine interaction is maintained in a setting that involves physical distance. Both the intellectual service for language offered by IBM (International Business Machines) and the cognitive computing application offered by

Google place an emphasis on the recognition of brain-like cognition in addition to judgement. This is accomplished by using a cloud-based service paradigm in order to provide help for making accurate decisions. While cloud computing and the internet of things offer the technology and software underpinning for cognitive computing, they also give methodologies for analyzing large amounts of data as well as ideas for deciding and recognizing innovative ideas and issues of valuable data statistics.

8.4 COGNITIVE COMPUTING FOR SMART COMMUNICATIONS

Applications that make use of intelligent information and a variety of services are progressively gaining popularity all over the globe as network technology and artificial intelligence methods continue to advance. The communication that is based on the cloud is making it easier to provide some strong communication features. The provision of a cloud-based infrastructure and an intense data processing system alone is inadequate, however, because of the restricted capacity of the system as well as the need for low latency, increased dependability, and a pleasant user experience. The field of research known as cognitive computing is concerned with the imitation of cognitive processes. Due to the need of resource optimization, the development and completion of 5G smart networks is a particularly challenging endeavor (Chen et al. 2017b). Cognitive computing, when paired with artificial intelligence, is responsible for the administration of cloud computing applications and wireless system services and then applications. As can be seen in Figure 8.4, there are a few characteristics that are associated with cognitive computing.

Through continuous learning and the acquisition of new information, the cognitive systems continued to update their knowledge, therefore achieving dynamic training. It is continuously interconnected and probabilistic in nature due to the fact that it links a number of modules and makes judgements based on the historical events that have occurred in the past. It is possible for the system to find the most effective solution to tools by assigning a weightage to each and every viable alternative.

8.4.1 The Cognitive Computing for the Society—Use Cases

Applications and services that provide intelligent information are gaining popularity all around the globe as a result of the progress that is being made in network technology and artificial intelligence. Another factor that has contributed to this is the fact that cloud communication provides powerful communication capabilities. On the other hand, just providing a cloud-based environment and intensive data processing methods is not enough because of limited capacity and the need for low latency, high reliability, and an improved user experience. The field of research known as cognitive computing focuses on the replication of human cognitive processes. Therefore, the expertise of cognitive scientists, when paired with communications, has the potential to enhance the performance of current systems by delivering increased precision and decreased latency.

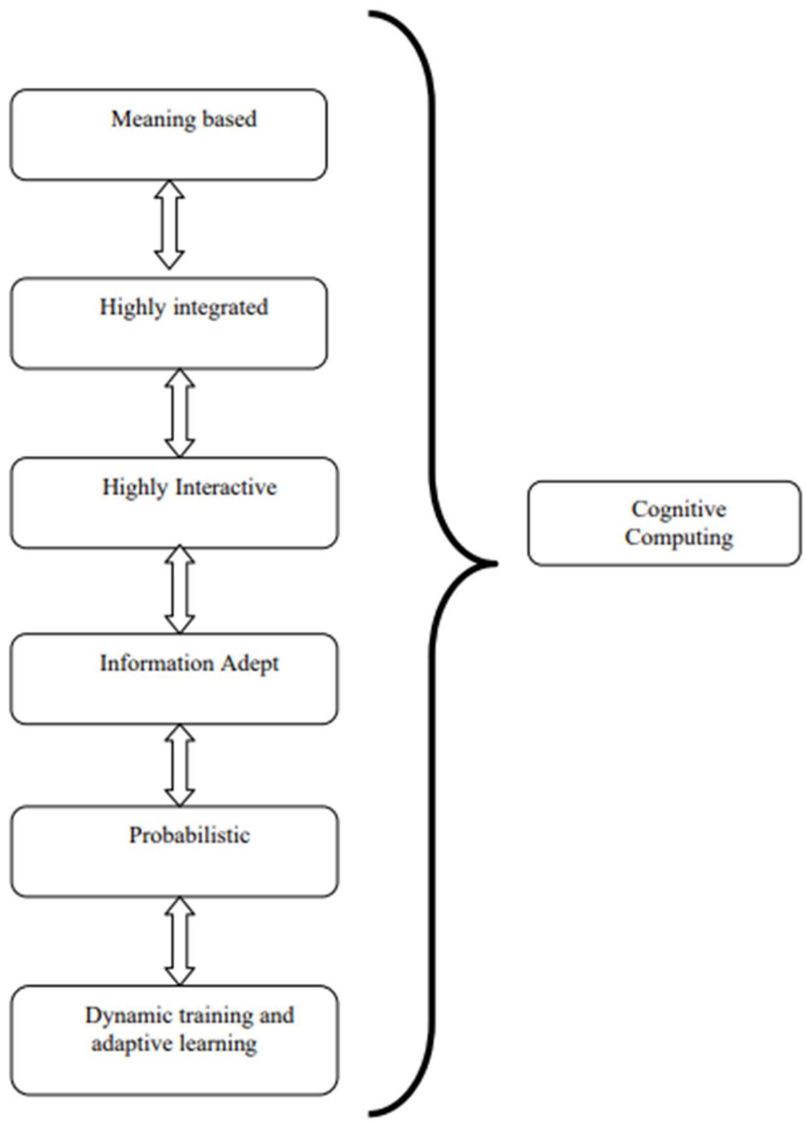


Fig. 8.4 The characteristic of cognitive computing

Computing is used in a variety of industries, including education, healthcare, finance, and retail businesses, among others. Utilizing research data, clinical data, biomedical data, and biomedical data, construct a cognitive-based system for use in the healthcare industry. Cognitive computing necessitates the use of talents such as learning, reasoning, and planning techniques. Regardless of the applications that are required for the cognitive process, they are as follows: product, process, and insight

One use of cognitive computing in the financial industry is the identification of fraudulent activity. Detecting fraudulent activity is made easier by retail-based systems, which were developed for the purpose of mining financial data. In the field of education, the learning styles of students are recorded as the date and later mined for information. The elimination of human mistake, the provision of digital support, the avoidance of incorrect judgements being made by machines, and the aid in doing repetitive tasks are all benefits of computers.

CHAPTER 9

6G: A WIRELESS NETWORK'S FUTURE PATH

9.1 CONCERNING THE MOBILE FOUNDATION

For the purpose of highlighting the relevance of mobile communications on a global scale, the introduction will supply readers with facts and numbers that are up to date. It will also be shown to them how these technologies are beneficial to individuals, businesses, the government, and the environment when used. In the second chapter, which investigates the evolution and current state of the art of older mobile communication systems, the primary motivating factors that have led to the creation of such innovative solutions in each successive generation of cellular networks, from 1G to 5G, will be brought to light. In the beginning of this book, we will talk about the challenges that were involved in developing the first commercial cellular network (1G).

The purpose of this is to provide readers with a reference for the rapid development rate, a mathematical assumption of the singularity, and to assist them in comprehending the history of unique technologies. As it makes the transition from analogue to digital communications for the very first time, it will continue to explain the second generation of mobile communications and emphasize the great technological leap that it represents. As part of its analysis, it will also investigate the ways in which the success of the third industrial revolution fueled the development of 3G cellular networks, which in turn gave rise to the generation of mobile Internet users. The topic of conversation will then transition to 4G, and we will investigate how this network made it possible for a variety of multimedia apps to be developed, such as peer-to-peer services and Fintech. We will also investigate how these applications altered the way in which people connect and trade, which ultimately led to the creation of the mobile economy.

An in-depth analysis of 5G, which stands for the fifth generation of mobile communications, will, in the end, expose the role that it played in sparking the fourth industrial revolution and linking the whole world via the growing use of robotics in every area of the economy. a summary of the most important aspects of the Connected Agenda 2030, which links the future of industry 4.0 and society 5.0. The readers will be better able to investigate and appreciate the necessity for a new network that extends beyond 5G from a social and technological point of view as a result of this. to get familiar with the challenges, possibilities, and possible downsides associated with 6G research, it is necessary to examine the prospective Key Performance Indicators (KPIs) as well as the current study topics.

9.2 THE DEFINITION OF 6G

In the year 2030, it is anticipated that the sixth generation of mobile communications, sometimes known as 6G, would be introduced. In light of the fact that discussions on the standardization of the subsequent wireless communication system have only just begun, it is now much easier to characterize 6G based on its future features. In order to promote the formation of civilization 5.0, a new civilization that combines the physical and digital worlds, the cornerstone of 6G attributes is the concept of a human-centric network that is also lightning-fast, decentralized, intelligent and cognitive, omnipresent, and completely environmentally friendly. When the characteristics of 6G are taken into consideration, it becomes abundantly evident that there is still a great deal of challenges to be conquered before this cutting-edge network can be put into operation. Now that we have established that, let us evaluate each of these qualities that constitute the 6G Network.

With an emphasis on people, the wireless network: The needs that mankind must fulfill in order to advance as a society are the focus of this notion, which outlines a wireless network that links the isolated hamlet in the highlands to the megalopolis. It is not a concept that is not well defined. The notion, on the other hand, is founded on the expectation of constructing a network that would assist humanity in accomplishing its social, economic, and environmental goals, such as those that were articulated by the United Nations (UN) in the year 2015. The idea is founded on the Goals for Societal Development (SDGs), which were established by the United Nations with the intention of compelling all nations to take prompt action in order to solve urgent problems that are affecting society. The ultimate goal is to achieve global peace and prosperity via the promotion of these two concepts. It is possible that technology will be of assistance in this regard, especially with the introduction of mobile broadband internet.

Within the next 10 years, the digital divide that now exists as a result of the fact that three billion people do not have access to the internet and are still disadvantaged by the abundance of the mobile economy must be sealed. One of the most important aspects of a human-centered network is the enhancement of educational opportunities, especially in rural and disadvantaged areas. Through the use of Brain-Computer

Interfaces (BCI), the creation of the next generation of smart limbs, and the utilization of nanotechnology to expedite the detection and repair of internal organs and tissues, it may also be possible for it to assist the progress of medicine. In addition to this, it can help with distant operations, which may improve the overall well-being of society.

Ultra-massive multiple input multiple output (UM-MIMO) technologies that are based on the new chemical component of graphene are the leading technologies in this scenario. In order to improve gain and data transmission, graphene will be used to miniaturize hundreds of thousands of antennas. However, even this decrease in size comes with its own set of physical limits, and in order to continue offering a solution, new radio technologies will be required. As a consequence of this, the holographic radio is an appropriate technology that aims to be the answer to the issues that are associated with other MIMO technologies. Networks Free of Central Authority: By continuing the path that was established by the decentralized network architecture of 5G, it is anticipated that 6G would be able to meet several use cases and deliver end-to-end service dependability and reliability.

As a result, it is envisaged that secure communications will be required in order to ensure that all nodes are able to be confident in their abilities. Wireless blockchain communication and quantum communications are two cutting-edge areas that have the potential to fulfill the promise of a highly secure network. For the purpose of completing the orchestration of a decentralized network, it is likely that both of these technologies, including self-defined wide area networks (SD-WAN), network slicing, sophisticated gossip protocols, and a hypothetical 6G QoS Flow, will be used. Managing a large variety of different service applications and Service Level Agreements (SLAs) requires the assistance of the organizations that have been discussed above.

Artificial Intelligence (AI) and Machine Learning (ML) will be included into 6G regions, to create a wireless network that is not only cognitive but also intelligent. Machine learning and artificial intelligence will be applied at the network's edge and core in order to handle the Big Data. Big Data will be generated and exchanged through the use of a variety of technologies, including Machine-to-Machine (M2M), Industrial Internet of Things (IIoT), Holographic Communications (3D Video), Internet of Nano Things (IoT), Knowledge Home (Human Bond Communications Beyond 2050), and CONASENSE (Communication, Navigation, Sensing and Services). Context awareness is thus the cornerstone of the sixth-generation wireless network. The use of machine learning and artificial intelligence is necessary in order to achieve these objectives.

Environmental Network: Lastly, it is crucial to have wireless connection that is environmentally friendly. As a means of preventing climate change and significantly lowering carbon emissions on Earth, it is important that we reduce our dependency on fossil fuels. Whoever is engaged in the engineering of 6G will need to have this notion crystal clear in their mind in order to avoid the development of a technology that is extremely power-hungry. Energy harvesting, coupled with sustainable means of saving energy or reusing energy, has to be devised in order to ensure that 1 Terabit/joule of wireless consumption will be given as a green Key Performance Index (KPI) for all end-to-end 6G communications. This is necessary in order to protect the environment.

9.3 A MOBILE SOCIETY'S DAWN

First things first: before getting into the subject of cellular networks, it is essential to have a solid understanding of the benefits that these networks provide to their customers. The establishment of a link between the services that are offered, the technology that is used, and the mobile services is necessary in order for us to do this. It is thus possible to project ideas for future networks that may surpass the ones that are now in place and result in a society that is even more innovative and inclusive.

This may be accomplished by reflecting on the positive trends that were developed as well as the challenges that were experienced. As a result of the advent of worldwide System Mobile (GSM), the second generation of mobile networks, new economic prospects were made available, and worldwide communication was made possible via the capability of transferring hypertext through user equipment (UE). Additionally, the introduction of the third generation of mobile communications, commonly referred to as Universal Mobile Telecommunication Systems (UMTS), which was created by the Third Generation Partnership Project (3GPP), was a significant event that marked the beginning of a whole new era of wireless communication. This new age was ushered in by the arrival of the Mobile Internet.

The introduction of mobile internet was a significant innovation that had a significant influence on society and technology all over the internet. The hard efforts of the 3rd Generation Partnership Project (3GPP) into the development of the mobile broadband architecture made this breakthrough feasible. This design facilitated multimedia applications throughout the range of the 3G signal and enabled users to access the Internet while they were physically moving about. According to the Measuring the Information Society Report published by the International Telecommunications Union (ITU) in 2018, there are 3.9 billion people using the Internet throughout the globe.

After then, the Mobile Economy was established as a result of the convergence of the Internet and wireless communication. the rise in the number of people who have subscribed to mobile phones over the course of the last five years, with a global study and a division of economic regions according to Developed, Developing, and Least Developed Countries (LDCs). The Mobile Economy, as it is now known, was responsible for the creation of millions of new jobs, the reimagining of the economy, and most crucially, the transition of millions of people from informal to formal work. Simply in 2019, the mobile market was responsible for the creation of 16 million direct employment and 17 million indirect jobs. On mobile apps (APPs), such as peer-to-peer (P2P) services or well-known cooperative resources, one could see occurrences that are comparable to these circumstances. Investopedia defines peer-to-peer (P2P) services as a "decentralized model whereby two individuals interact to buy or sell goods and services directly with each other or produce goods and services together."

This is the definition of P2P services. There are a few examples of peer-to-peer (P2P) services that provide an overview of the mobile service sector. Some examples include carsharing, food delivery services, and online shopping. In addition to providing an overview of the advantages that mobile communications have brought to the economy and a variety of applications, it also gives an outline of the present impediments that are preventing the digital divide from being closed. There is a broad variety of applications for mobile services, ranging from social media networks to online banking, which has contributed to the increased popularization of the mobile business. The total income generated by mobile apps in 2019 was $461 billion USA dollars. We expect a growth rate of 49% by the year 2023. When seen from this angle, it is easy to understand how the mobile application sector is dependent on mobile broadband services in order to retain its financial stability and continue to provide mobile customers with new goods.

9.4 DIFFICULTIES TO OVERCOME

As a consequence of this, investments in mobile network infrastructure are absolutely necessary. In 2019, telecommunications companies generated a revenue of $1.03

trillion, as reported by the most current statistics from the GSM Association (GSMA). It is anticipated that this figure would increase to $1.14 trillion by the year 2025.

It is anticipated that the capital expenditures (CAPEX) of telecommunications firms would amount to one trillion dollars until the year 2025, with a significant percentage of that amount going towards the development of 5G infrastructure. With regard to the distribution of the connectivity market between 4G and 5G networks, it is projected that the percentage of 5G connections will be 25%, while the percentage of 4G connections will be 56%. Actions need to be performed simultaneously in order to ensure that the Sustainable Development Goals (SDGs) set out by the United Nations in 2030 are achieved. The seventeen Sustainable Development objectives (SDGs) focus on social, environmental, and economic challenges and highlight the important role that technology should play in promoting these objectives in the future. These goals were established by the United Nations in December 2017. It is possible for technology to be built to meet any known human need while still being based on Return on Investment (ROI) when it is employed with a humanistic approach. The Sustainable Development Goals are as follows:

1. In order to eradicate poverty on a global scale
2. In order to eradicate hunger
3. To promote health and well-being for the community
4. To provide a decent educational experience
5. To advance the cause of gender equality
6. To provide access to healthy water and sanitation
7. To provide energy that is neither expensive nor polluting
8. In order to provide adequate employment and economic development
9. To provide the business sector, innovative ideas, and infrastructure
10. In order to lessen the disparities
11. To make cities and communities more environmentally friendly
12. To provide consumption and production based on ethical practices
13. To combat the effects of climate change
14. To protect the oceans and seas from pollution
15. Take measures to preserve forests and combat desertification.
16. To encourage the maintenance of peace, justice, and robust institutions
17. To reinvigorate the international relationship for the promotion of sustainable development.

The pandemic caused by the COVID-19 virus will undoubtedly have an effect on the Sustainable Development Goals (SDGs), which are already experiencing and will soon experience the most severe economic downturn since the Great Depression. The United Nations forecasts that three percent of the world's GDP will decline in the year 2020, which will have significant repercussions in the years that follow. Taking into consideration the COVID-19 response statement issued by the United Nations, which declares that "No country can overcome this pandemic alone," Having global solidarity is not only the ethically correct thing to do, but it also benefits the interests of everyone. It is possible that wireless communication networks might be of assistance in achieving the Sustainable Development Goals (SDGs) while also maintaining pace with the current status of the globe. Mobile communications, for instance, make it easier to travel less, which will have a direct impact on the amount of greenhouse gases that are released into the atmosphere.

According to the National Aeronautics and Space Agency (NASA), the term "greenhouse effect" refers to the concept that "greenhouse gases" are responsible for the retention of heat in close proximity to the surface of the Earth. Carbon dioxide and nitrous oxides are two examples of gases that contribute to global warming. As a result of human activities in the environment, which increase emissions of greenhouse gases, climate change is directly influenced by these activities. In the event that mobile broadband infrastructure expands to cover a greater number of consumers and businesses, there will be a reduced need for travel and the utilization of intelligent transportation.

There is a possibility that wireless connections may directly cut carbon emissions. This is because they eliminate the need for presenters to be present at meetings and enable video conferencing. In addition, the advent of e-commerce and the acceleration of product delivery, which eliminates the need for either a retailer or a client to travel in order to acquire anything, are examples of ways in which greenhouse gas emissions might be reduced. It is possible that the aforementioned examples indicate the manner in which mobile services and improved mobile telecommunications infrastructure help to the creation of a more pleasant atmosphere.

The government's objective was to create a more humanistic technological approach. Industry 4.0 is inferior to Society 5.0 in terms of its comprehensiveness. In Society 5.0, technology is not seen to be the exclusive means of obtaining return on investment (ROI) and running an effective and waste-free business. Instead, it makes use of technology "to balance economic advancement with the resolution of social problems," which increases its human-inclusivity and makes it more accessible to people. In order to achieve the objectives of the United Nations Sustainable Development Goals (SDGs), Society 5.0, and to make development in line with these principles, the scientific community has to develop a network that is more advanced and innovative than the 5G network that is already in place. In order to facilitate the growth of a digitally connected and human-centered society in the future, it is essential to begin investing in the process of defining the roadmap for the technologies that will be behind the 6G network. Innovative technology will be used by 6G in order to facilitate the provision of wireless connections, wireless fidelity (WI-FI), satellite constellation systems, fixed networks, and an improved society.

9.5 CONSENSUS AND UPCOMING WIRELESS TRANSACTIONS

For the purpose of gaining a comprehensive comprehension of the potential issues that may be brought about by future use cases that may be brought about by 6G, it is necessary to address and explain both a starting point and an example. As a direct result of this, the CONASENSE is shown on this particular subpage. CONASENSE is an abbreviation that is used to denote the four key research disciplines that are concerned with the integration of communication, navigation, sensing, and services. These fields get their name from the acronym. The researchers at CONASENSE society formed the presumption that it would take place somewhere around the year 2012 at the beginning of the ten years that had just passed. It has resulted in the production of a great number of books and articles that are related to the topic.

A society of the future that would experience the convergence of real and virtual locales is the foundation of the CONASENSE concept, which is based on the research that is now being carried out to identify how to integrate and enable the primary sectors that have been defined. This idea revolves on the interconnection of different systems as a means of establishing a new reality that offers new benefits to the human race. In order to achieve this goal, it is necessary to make it possible for five dimensions to coexist. These dimensions consist of the three physical dimensions that are presently in existence, in addition to time and the cyber domain. It is important to note that every single aspect of this idea has an effect on the perspectives that academics have about innovative communication networks and the interdependent services that these networks provide. This is a presentation of the CONASENSE architecture that you are welcome to view according to your preference:

Active Participation in: A cognitive network that provides better data compression, energy efficiency (nanogenerators), protocols for a decentralized, complex, and heterogeneous network, and improvements in data transmission optimization are the foundation of this field of study.

Navigation: The capability of future sensors to give every network node with a precise position in space and time down to the centimeter is crucial for a number of forthcoming wireless applications. This capacity is essential for maintaining navigation. It will facilitate the development of autonomous vehicles, drones, and nanotechnologies that are used in the medical field. As a consequence of this, the Global Navigation Satellite Systems (GNSS) in combination with an extra external sensor will guarantee a level of dependability and safety that is 99.999%.

Sensing: The function of sensing is to provide context awareness for the data that is being sent, and it is responsible for doing so on both sides of the network's nodes. The provision of geolocation-based services and the guaranteeing of data delivery are also included. The implementation of this technical solution is likely to be accomplished via the use of semantic algorithms and quantum communications.

Services: In addition to a few services, it offers benefits to a wide range of additional services over the whole spectrum. Experiments in space exploration, unmanned aerial vehicles, and airborne vehicles are all instances of innovations that have occurred in recent times. It has been determined via an analysis of the CONASENSE design that the use of 6G is essential in order to achieve the desired goal of effectively integrating the real and virtual worlds.

9.6 RESEARCH HOME IN THE AGE OF 6G

Acquiring knowledge Transmission of Human Connections Beyond the year 2050 is the focus of the Knowledge Home initiative. The concept of Knowledge Home encompasses the concept of enabling individuals to involve themselves with their surroundings and the environment via the utilization of technology through the integration of their five senses. There are five senses that humans possess: the visual, the auditory, the gustatory, the olfactory, and the tactile senses. These five senses are responsible for the construction of the human sensory organ, which is responsible for enabling the human sensorial cells to receive and convey information. As a consequence of this, the Knowledge Home idea will make it possible for all of these sensory impulses to engage with a wireless network in order to generate interactions with individuals in real time inside a personal area network (PAN). A smart home, workplace, spaceship, or even stadium may offer people with responses by assessing their five senses and interacting with them to propose appropriate reactions based on the sensations that were recorded inside the PAN area. This idea is based on the idea that a smart home, workplace, spacecraft, or stadium could provide people with responses.

After doing an analysis of a person's senses, a sensor may provide a number of responses. These responses may include controlling the temperature in response to people's reactions in the surrounding area or contacting the proper authorities in order to provide assistance to those who are in need in the immediate vicinity. In order for all of these things to become a reality, human bond communication (HBC) has to be supported by a robust and intelligent network that priorities data traffic with a high quality of service. That being the case, it is necessary to construct an intelligent network in order to accommodate the enormous amount of data traffic that is generated by sensors and humans in real time.

9.7 INDIVIDUAL BUSINESS MOBILE GENERATION

In 1947, engineers Douglas H. Ring and W.R. Young of Bell Laboratories were the first to suggest a mobile communication network. Their paper, which was named Mobile Telephony – Wide Area Coverage – Case 20564, was published in the year 1947. In this study, D.H. Ring provides an overview of the principles of a commercial cellular network, with particular attention on the most important aspects. The ability to provide service to a mobile device at any time and from any location in the country should be the ultimate objective of a mobile phone system that has undergone significant development.

In spite of the fact that they fall short of this ideal, systems may still be able to deliver essential services in the meantime as it will take a significant amount of time to accomplish this ultimate aim. A challenge is presented by D.H. Ring in this statement, which indicates how the technology that was available at the time was inadequate to provide such important telecommunications networks. In addition, he acknowledged the contributions of W.R. Young, the brilliant guy who was responsible for the conception of the cellular cell idea, which was the expected method for addressing the issue of frequency allocation.

On the other hand, it was designed to provide practically little voice security and to transmit voice only in analogue format. In addition to this, it had a hefty membership fee, a limited number of services, no multimedia service aggregation, a bad data rate, a limited capacity, and a short battery life. First-generation wireless services were only available to a restricted number of clients who were able to buy them. This was a modest luxury for such customers. A number of other obstacles, including as its limited radio coverage, its sensitivity to radio interference, and its lack of interoperability when travelling between countries or roaming, were additional factors that kept it from being extensively adopted. Another significant problem was that the systems employed by different countries did not have interoperability with one another

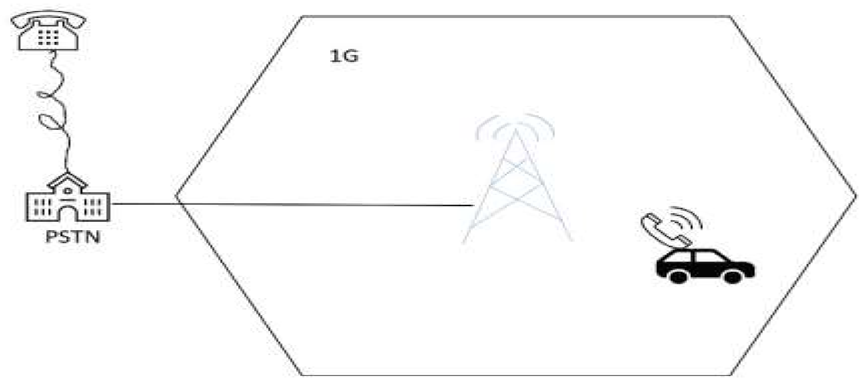


Figure 9.1 First Generation of Mobile Communications

Despite the fact that the Advanced Mobile Phone networks (AMPs) and the Total Access Communications Systems (TACs) were two successful cellular networks, they were unable to exchange services with one another. In order to circumvent this issue, the European Commission began providing assistance for a new wireless network standard that was founded on digital transmission. The outcome of this effort was the Development of the Second Generation of Mobile Networks, also known as 2G. In spite of this development, the first generation of wireless communication systems continued to dominate the market for more than ten years, beginning in the 1980s and continuing until the middle of the year 1990. In a technical sense, the first generation of cellular networks used frequency division multiple access (FDMA) and frequency division duplexing (FDD) in order to give a data capacity of 300 bits per second (bps). The many different cellular system technologies that were developed for the First Generation of Mobile Communications are shown in Figure 9.2.

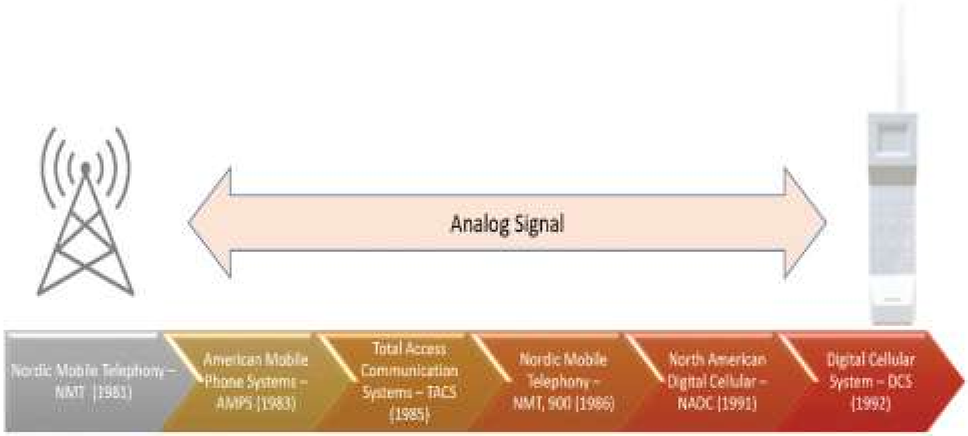


Figure 9.2 First Generation of Mobile Communications Standards

9.8 THE CURRENT INITIATIVES FOR 6G RESEARCH

In the year 2010, the first 6G themes started to appear on the internet. There has been an increase in the amount of discussion on 6G research from the year 2019, as 5G networks have become more established on a global scale. The process of planning for future cellular networks often begins ten years before the network is introduced, and this trend has continued ever since the invention of the third generation of cellular protocol (3GPP). Considering that standardization is becoming an increasingly important factor for the success of any firm, notably in the telecommunications sector, it is reasonable to anticipate that the situation would be the same with 6G. Over the course of the next ten years, various entities, including institutions, researchers, engineers, scientists, intergovernmental agencies, standard bodies, and the private sector, will collaborate in order to establish the conceptual framework and technical specifications that are required in order to make 6G, the next generation of mobile networks, economically feasible. There is a concerted effort being made by the 6G research teams to expeditiously define the 6G framework and the needs for it.

Among the worldwide initiatives that have been announced up to this point are those that are being carried out by universities, corporations, and existing organizations that are focusing their attention on this matter:

1. The CTIF Global Capsule includes components such as the 6G Knowledge Lab, which is located at Aarhus University in Denmark. Future wireless

technologies, such as 6G, are subjects of research conducted by a dedicated group at CGC.

2. The Finland-based 6G Flagship. Sixth Generation Flagship is a specialist team that investigates the specifics of fifth generation and the fundamentals of the sixth-generation industry.
3. The Chinese Ministry of Technology and Science has announced the establishment of a research and development agency for the sixth generation of mobile communications. This announcement comes as part of a national push to construct the sixth generation of mobile communications.
4. The International Telecommunication Union (ITU) has created a group called the Focus Group on Technologies for Network 2030 (FG NET-2030) with the purpose of conducting research on cutting-edge global telecommunications technologies that will emerge after 5G.
5. The Next G Alliance is an initiative currently being driven by the commercial sector in North America with the objective of advancing research and development for 6G networks. Twenty-one years from now, this organisation will be fully operational.
6. A study group was created by the Radio Communication System Technical Committee (RCS), which is a section of the Institute of Electronics, Information, and Communication Engineers (IEICE). Additionally, a symposium on 6G technology was held in Japan.
7. WPMC2020 is a conference that provides a variety of conferences on the 6G technology and is also accessible in a virtual format.

In order to construct the service architecture of future wireless technologies, technology must first conduct an analysis of the patterns that will emerge over the next period of many decades before moving forward with the development of 6G. "The Knowledge Future: Intelligent Policy Choices for Europe 2050" is the name of a plan that was developed by the European Commission. Within the framework of this plan, the European Commission investigates Europe as well as the challenges that the global society will face in the year 2050.

The objective is to provide a road map for the purpose of recommending technologies that will put knowledge into action via the utilization of applied research, technological innovation, and technological advancement. Concerns about the environment and the expansion of the human population will be the primary subjects of conversation. The

European Commission's recommendations are based on the following four concepts over the next thirty years:

1. In Europe, there is a system of open knowledge, also known as shared information.
2. Capacity for Adaptation and the Use of Trial and Err for Innovation
3. The level of cooperation that exists throughout Europe
4. The availability of resources and a tax base to foster innovation, research, and educational opportunities.

Additionally, the United Nations established the World in 2050 (TWI2050) agenda, which focuses on measures to promote the achievement of the Sustainable Development Goals (SDGs), and began discussing what to expect beyond the year 2030. Innovations for Sustainability: Pathways to an efficient and appropriate post-pandemic future is the title of a study that was recently published by the United Nations and the International Institute for Applied Systems Analysis (IIASA). This paper was created with the intention of addressing the challenges that would develop after the year 2030 and up to the year 2050.

The advice that is presented in this article is to provide an evaluation of all the potential positive benefits that innovation may bring to sustainable development for everyone, while also taking into consideration the potential adverse consequences and challenges that may arise in the future. With a particular emphasis on production and consumption, the research places an emphasis on inefficiency and sufficiency in the supply of services to individuals. This is done with the intention of leveraging innovation for the purpose of achieving sustainability. There is a connection between this statement and Society 5.0, and we believe that 6G will help the measures that are required to unleash innovation and accomplish the goals of the Sustainable Development Goals (SDGs).

9.9 6G RESEARCH DOMAINS – STRUCTURE

Nevertheless, what is it about 6G networks that makes them such a significant area of research for the foreseeable future? The answer to this question is based on each and every one of the conditions that must be satisfied in order to bring about social and environmental growth on Earth. This is the foundation for the solution. In order to give support for the foundations of 5G Networks and to enable a hyper-connected society that is capable of satisfying all social, industrial, and environmental needs while simultaneously sustaining prosperity, the world will require a strong future network. This will be necessary in order to provide a foundation for 5G Networks. Therefore, in order to take advantage of improvements and technologies that are not compatible with the ecosystem of 5G, it will be necessary to construct a complicated network that is capable of accommodating these new technologies and innovations.

Therefore, in order to achieve this goal of bringing together the most cutting-edge aspects of science and technology, it will be necessary to collaborate on a global scale and make investments in fields that are not related to one another in order to construct a network that combines fixed and mobile broadband communications. This will be necessary in order to accomplish this objective. Taking this into consideration, the subject of the research might be broken down into a number of different threads that are interrelated and unifying. This would be an important study for the construction of future wireless communication networks, including 6G networks.

9.10 SIXTH GENERATION KPIS

When it comes to the planning of future technologies, a great deal of questions arises. It is feasible to assert that science and futurology are intertwined with one another. The discussion about the next generation of wireless communication, on the other hand, will take precedence over the study of futurology. Data and trends will be the driving forces behind the objective of establishing new services via the use of telecommunications in order to bridge the gap between technology and future use cases. When it comes to key performance indicators (KPIs) and innovative applications of cutting-edge technology, 6G must exceed 5G in every aspect. Additionally, the innovative network is obligated to provide the key performance indicators (KPIs) that are listed in the table that follows. In the next ten years, each key performance indicator (KPI) will bring a challenge that must be overcome. As a result of this, 6G must provide new key performance indicators (KPIs) and a jump of 10–100% over the current 5G KPIs. Within the research community, these are the major key performance indicators (KPIs) that have been developed and intensively discussed.

When discussing the notion of 6G, which is intended to transcend the capabilities of 5G, it is predicted that a complex network would develop. There is a high probability that the new cellular environment will seem like a complex and decentralized network. In the first place, given that 6G will be required to serve a variety of use cases. Furthermore, the non-consolidated technologies that were not accessible at the time when 5G was being deployed will be very important over the course of the next 10 years. They are going to be very important in determining the continued innovation for applications that are reliant on 6G. Because of the continued growth of direct-to-device (D2D) communication and, more specifically, the Internet of Things (IoT), there is a need to improve the capacity of the UPLINK for Massive-Machine Type Communications.

In addition, the approach that should be taken to successfully address the wide variety of services. for the reason that every single service request that is made over radio network access will have a different service level agreement (SLA) and answer. To find a solution to this problem, 6G will need to be prepared to use machine learning at the network's edge in order to improve its algorithms and adjust its responses in a more expedient manner while still adhering to the service level agreement that was previously established. Since this is the case, it follows that a 6G network will provide an extraordinary quality of experience (QoE) to all users and services. Additionally, network slicing will be a component of its ecosystem. In order to offer further assistance for the management of Big Data with nearly no latency between front-backhaul, the future wireless network will need to include essential basic entities in the ecosystem. Here are some examples of these entities:

Quality of experience (QoE) must be guaranteed for every 6G user equipment (UE) and critical multimedia applications. For example, while transmitting an HTC (holographic type communications), it is necessary to take into consideration the user experience (UX). Additionally, self-adjusting mechanisms for gearbox quality need to be created in order to accommodate such essential services. It is necessary to have synchronization with the immersive audio and high-quality image in order for the holographic communication to provide the expected results that are nearly as lifelike as they seem.

Therefore, in order to prevent any possible delays in the transmission of data packets that came from the HTC sender to the HTC receiver as a consequence of network traffic congestion, it will be necessary to develop ways that will reduce the impacts of jitter. For instance, the holographic principle describes the process of lighting an object with a laser light source and then recording the three-dimensional patterns that are produced as a consequence of this illumination. Reconstructing the original object in three dimensions is made possible on the receiving end of the process. Audio, on the other hand, is a fourth component that is taken into consideration and captured in holographic communications. In the present moment, there are a number of different options available on the market for this form of communication. Those who do not use Head Mounted Displays (HMDs) or those who use them as suitable 3D glasses for the user experience are not the same as those who use them. For example, Microsoft Haloboration makes use of three-dimensional cameras to capture images of people and objects. These images may then be reconstructed with the help of a head-mounted display (HMD). Another example of this is the 5G Holographic Cloud Communication Network, which was created by two Chinese businesses, ZTE and WIMI. The system makes use of a 5G network, 4K video terminals to gather 3D pictures, and a holographic algorithm to reconstruct the images into a physical representation in order to facilitate the transport of holographic communications.

There are a lot of challenges that come up with this form of holographic communication, including the fact that:

- The process of digital packetization
- Making use of the power of computers in order to implement quality of service (QoS)
- The Capacity of the Network itself
- A Latency that is Extremely Low
- Prioritization of Traffic (Including Quality of Service and Network Slicing)
- The Development of Methods Capable of Encoding and Decoding

Under the Focus Group Technology for Network 2030 (FG NET 2030), the International Telecommunication Union (ITU) has committed to doing research into holographic communications as part of its aims beyond 5G. FG NET 2030 broadened the scope of their research on holographic communications to include the tactile HTC by providing users with the ability to touch holograms with their hands. The conclusion drawn from this is that "Ultra-low delay requirements (to provide an accurate sense of touch feedback) are imposed on underlying networks by tactile networking applications, and, in particular, as far as mission-critical applications like remote surgery are concerned, tolerate no loss." A powerful network that is cognitive and detects in real time the requirement to prioritize traffic that addresses all of the key performance indicators stated below may thus be implemented with the use of artificial intelligence, machine learning, and machine expert computing (MEC). These criteria, which highlight the enormous quantity of data that is transferred for a single application from beginning to finish, highlight the significant role that 6G plays in enabling the many usage cases that HTC has.

In order to achieve the fundamental objective of ensuring the safety of essential applications, the 6G Network Core will need the ongoing implementation of Network Slice principles in addition to other technologies that are now in existence and coming into existence. In spite of the fact that it is an outdated technology, Multiprotocol Label Switching (MPLS) will be an essential component of the 6G Core. At this moment, it will be powered by Software-Defined Wide Area Networking (SD-WAN) architecture in order to deliver the best possible user experience for this sort of communication, especially from the perspective of Cloud Services architecture. SD-WAN makes it feasible to decouple the control plane of the wide area network (WAN) from the data plane.

9.11 MACHINE LEARNING AND AI

The amount of traffic that involves different types of data will continue to rise. Utilizing Artificial Intelligence, Machine Learning, and Network Slicing in conjunction with Edge Computing will be necessary in order to fulfil this need completely. The use of data mining and analytics would be required in order to provide faster responses for the many different types of devices and services that are operating at the core of 6G networks. Because of the enormous amount of data that will be flowing through it, an intelligent network will need solutions that go beyond the traditional methods in order to react swiftly.

A flexible mode of operation is required for the cellular network that will exist in the future. The term "predict-to-prevent" refers to the process of responding proactively to potential threats and congestion on a network in order to either self-correct or reallocate resources. It has been shown that this orchestration is one of the fastest approaches for processing heterogeneous data and controlling it according to the relevance it has for users and applications. In a similar vein, the combination of machine learning and artificial intelligence with software-defined wide area networks (SD-WAN) and self-organizing networks (SON) has the potential to provide an extra level of assurance in terms of delivering quality of service and quality of experience for vital services. The incorporation of artificial intelligence into the architecture of the 6G network is required. The artificial intelligence will operate together in a cooperative manner both on the perimeter and in the core of the 6G network. As a result, the most effective method for dealing with Big Data will be to use artificial intelligence to control the fronthaul and backhaul of 6G networks. In this regard, two different AI techniques will be developed.

Generalized artificial intelligence is the most advanced kind of flexible artificial intelligence system. It is designed to solve all of the many issues that have been brought up. Quite frequently, this design is modelled by the operations of the human brain. In this composition, AI is a collection of several types of AI. It is possible for the generalized artificial intelligence to control the primary operations of the 6G network, which include network traffic, jitter, latency, cyberattack detection, and continuous route analysis. The results, together with analytics that allow real-time monitoring, will be sent to the telecom providers. It will automatically evaluate which resources are ideal to allocate to a particular microcell and then deliver the findings to them. Taking these actions will result in a reduction in both CAPEX and OPEX expenditures since they will be seen as sensible planning for future network development or cost optimization initiatives. The last component, which will involve machine learning, will be located at the network's edge. Through the use of its algorithms, machine learning will make it possible for the Network to grasp the context of the data information that is acquired from the multiple events and service requests that are taking place at the Network's border.

Author Details

ISBN: 978-81-970527-2-9

Dr. Manvi Sharma, is Ph.D in VLSI domain from Thapar Institute of Engineering and Technology, Punjab, India, M.Tech in ECE from Indira Gandhi Institute of Technology (now IGDTUW) and B. Tech in EI from APJ Abdul Kalam Technical University, India. She is also certified in Artificial Intelligence and Machine Learning from IIT Kanpur. She has several Research papers in SCI, Scopus indexed journals and conferences, also has some patents and copyrights on her name. Her area of interests are Communication systems, Mechatronics and IoT.

Xoffencer International Publication
838- Laxmi Colony. Dabra,
Gwalior, Madhya Pradesh, 475110
www.xoffencerpublication.in

MRP: ₹ 550/-