1 Introduction

Gernot Hueber and Ali M. Niknejad

1.1 5G

A lot of the focus of this book is on 5G, so you may be wondering, what exactly is 5G? And, perhaps more importantly, how does it impact me as a circuit designer? Hopefully we can answer the first question in this chapter, and leave the rest of the book to address the second one.

1.1.1 What Is 5G?

The term "5G" has been around for a while as it is really a marketing term. People were talking about 5G even before anyone knew what 5G was going to be about. Even today, if you ask five different people, "What is 5G?" you may get more than five answers! Well, the name is naturally 5G because it is the "Fifth-Generation" mobile network standard. Ultimately, 5G will be defined by standardization bodies such as 3GPP (3rd Generation Partnership Project), and even then the concept of 5G will evolve. The reason that it's so difficult to pin down a clear definition for 5G is that it's going to be a worldwide network standard for the next decade, and there's a long wishlist of new technology elements that people want to see in 5G, and then there's the reality of building and deploying a new network and keeping costs and power consumption at a reasonable level. So 5G is a compromise between our dreams for the next-generation radio versus the reality of what is technologically feasible and economically viable.

5G technology is positioned to address all of the shortcomings of 4G technology. In particular, people envision "everything in the cloud," which can offer a desktop-like experience on the go, immersive experiences (lifelike media everywhere), ubiquitous connectivity (intelligent web of connected things), and telepresence (real-time remote control of machines) [1]. To address these new application scenarios from a mobile device, the following "rainbow of requirements" shown in Figure 1.1 have been defined: (1) peak data rates up to 10 Gbps, (2) cell edge data rate approaching 1 Gbps, (3) cell spectral efficiency close to 10 bps/Hz, (4) Mobility up to 500 km/h, (5) cost efficiency that is 10 to 100 times lower than 4G, (6) a latency of 1 ms, and finally, and perhaps most importantly, (7) over 1 M simultaneous connections per km^2 [1,2].

Before we dive into the details, it's useful to have a very brief history lesson.

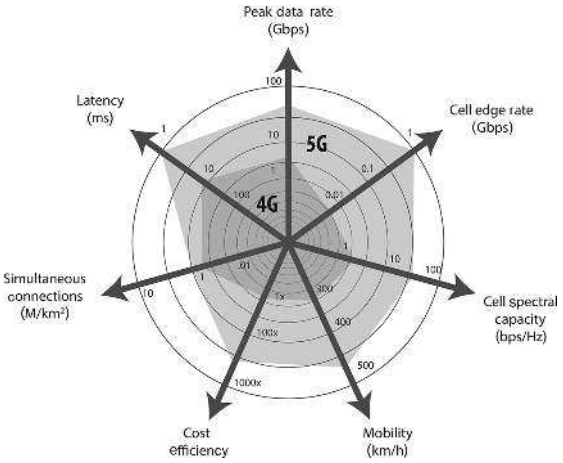

Figure 1.1 The 5G rainbow of requirements, adapted from [2].

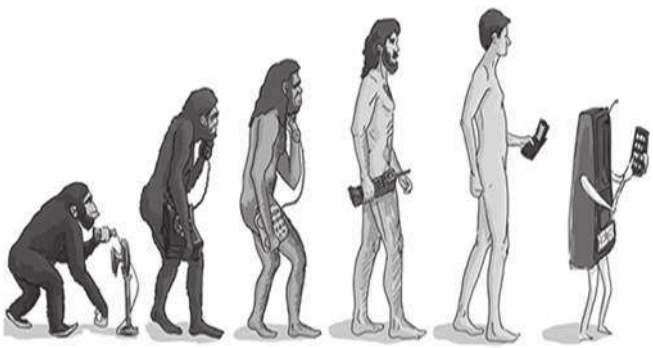

Figure 1.2 Evolution of humankind alongside wireless communication technology [3].

1.1.2 A Brief History of the Gs

Some of us are old enough to remember the days of brick-sized phones and analog mobile communication, the so-called 1G era and the Advanced Mobile Phone System (AMPS), first deployed in 1979 (see Figure 1.2). The system was analog and operated originally in the 850 MHz frequency band. The channel bandwidths were only 60 kHz and it was intended for voice communication. One important distinction to note is that 1G systems were circuit switched, so once a call was activated, the spectrum was allocated to a user, even if both sides of the link were silent.

In the early 1990s, the 2G generation took over and offered digital communications for the first time, including the ability to use Time Division Multiple Access (TDMA). In most parts of the world, 2G and the term GSM (Global System for Mobile communications) were synonymous, which used 200 kHz per channel, and Gaussian Minimum Shift Keying (GMSK) modulation (constant envelope) for power amplifier (PA) efficiency. But in addition to the GSM standard, a second-generation AMPS standard

called Digital AMPS (D-AMPS), also referred to as TDMA, was in operation (IS-54 and IS-136). At the same time, Qualcomm was actively selling a new radio access technology known as Code Division Multiple Access (CDMA), and these radios were standardized as IS-95.

There were some 2.5G systems that used packet switching, known as General Packet Radio Service (GPSR), as opposed to circuit switching, which allowed the system to offer more efficient spectral access. This meant that more time slots could be allocated on demand, and the latency and data rate depended on the number of users connected to a base station. By today's standards, 2.5G systems were dog slow, topping in at 50 Kbps. At first no one was really using mobile for data, and this didn't seem to be an issue. But the increasing popularity of mobile communication drove the need for more bandwidth and more speed. This is where the 2.75G standard evolved and offered EDGE (Enhanced Data Rates for GSM Evolution), offering theoretical speeds of 1 Mbps, by using 8-PSK encoding (three bits per symbol).

Interestingly, the first iPhone was released in 2007, 16 years after the introduction of 2G, and it was still a 2G device. For those of us lucky enough to have owned a first-generation iPhone, the experience was both amazing and also tortuous because of the slow network speeds due to 2G limitations and also due to the fact that in dense urban environments, everyone was all of a sudden trying to access the network for Internet connectivity at the same time. These early smartphones, especially the iPhone, were heavy users of data, and they really showed the world that the 2G network was not good enough. Other devices at the time were already using 3G technology, which came in many shapes and sizes.

In the late 1990s and early 2000s, the 3G networks started to operate and offered improved data rates by increasing the bandwidth of channels (up to 5 MHz) and adopting spread spectrum techniques and higher-order constellations (16- and 64-QAM) and multiple-input and multiple output (MIMO) techniques. The Universal Mobile Telecommunications Service (UMTS) radios were introduced as hybrid 2G/3G UMTS/GSM radios. Sometimes these systems were referred to as W-CDMA systems, due to the use of a wideband code division multiple access technique. Data rates increased to 384 Kbps in the original systems, and evolutions pushed the data rates higher to Mbps regions with High Speed Packet Access (HSPA) and HSPA+ offering up to 168 Mbps in downlink and 22 Mbps in the uplink. The adoption of multiple bands meant more complex front-end circuitry, wider bandwidths, and therefore more linearity to handle more complex modulation schemes. In parallel, the CDMA2000 standard (IS-2000) offered peak data rates of 14.7 Mbps using 1.23 MHz of channel bandwidth. Unfortunately, the CDMA2000 and UMTS/HSPA radios were standardized by different committees and were not interoperable, making phones not only region-specific but also carrier-specific.

Today we are living and fully immersed in the 4G world of LTE, or Long Term Evolution, the "winner" technology that is ubiquitous worldwide. One of the requirements for 4G was to offer over 100 Mbps of peak data rate for highly mobile access and approximately 1 Gbps for low mobility access. The Samsung Galaxy Indulge was the world's first LTE smartphone starting on February 10, 2011 [1]. To move toward these lofty goals in power transfer and low latency, LTE networks were all Internet Protocol (IP)

packet switching, employed very dynamic network architectures for optimum sharing of network resources, offered scalable bandwidths from 1.4 MHz up to 20 MHz, and distributed these resources on demand using Orthogonal Frequency Division Multiple Access (OFDMA) [4]. The spread spectrum techniques widely used in 3G systems were abandoned in favor of OFDMA, or the division of a wide bandwidth into smaller bands, modulation of the subcarriers at a much lower rate, and the use of a cyclic prefix in the guard band, thereby circumventing frequency-dependent fading and intersymbol interference. Using many subcarriers also allows the base station to optimize resource allocation by allocating spectrum resources in both time and frequency slots. More efficient turbo codes and MIMO techniques also improved the link quality.

One well-known pitfall with OFDMA is that the composite multicarrier signal has a very high peak-to-average power ratio (PAPR), which spells disaster for power amplifiers, requiring high back-off and linearization. These issues are well known to the power amplifier community as WiFi networks adopted OFDM (Orthogonal Frequency Division Multiplexing) as early as 1999 and the introduction of 802.11a. To circumvent this high PAPR, and single-carrier FDMA (SC-FDMA) for the uplink reduces PAPR. This slightly complicates the transmitter and requires frequency domain equalization in the receiver.

While most 4G systems converged on LTE, providing compatibility in theory, in practice the number of LTE bands exploded covering from 450 to 3600 MHz and both frequency division duplex (FDD) and time duplex (TDD) access. This meant that designing a "worldwide" LTE phone would be a formidable task due to the number of different front-end components required to cover disparate frequency bands and access modes (FDD versus TDD). LTE-Advanced (LTE-A) is an extension of LTE with new features including up to 8×8 MIMO and 128 quadrature amplitude modulation (QAM) in the downlink and carrier aggregation of contiguous and noncontiguous spectrum allocations, allowing up to 100 MHz of aggregated bandwidth. This means a device with LTE-A has a theoretical peak download data rate of 3 Gbps [5]. While this rate is impressive, in practice most users never reach these peak data rates.[1]

As evident in this brief history, each generation of mobile standards has embraced the latest advancements in communication theory and technology, in particular advances in coding, multicarrier modulation and wider bandwidths, and MIMO techniques to enable ever-increasing data rates and more efficient and dynamic networks. Each generation lasted about a decade, and it is a small miracle today that we can all enjoy watching our favorite cat videos from virtually anywhere on the planet.

1.1.3 Do We Need 5G?

So why do we need 5G? LTE and WiFi are amazing technologies that have served us well. Will the investment in a new network pay off? First, let's consider the new generation of users of wireless technology. A typical 12 year old today was born with

[1] The coeditor of this book has obsessively tested his phone all over the Bay Area and topped out at 162 Mbps downlink and 43.5 Mbps uplink.

a smartphone or tablet in her vicinity for most of her life. She may have never even experienced Internet blackout as a whole generation of parents replaced the TV with the smartphone/tablet as the de facto caregiver. The TV was limited in mobility whereas a smartphone can be carried anywhere and offer not only videos, but countless games and other forms of entertainment that only this new generation can understand.[2] This generation has a different relationship with bandwidth because they constantly stream video. Students prefer watching lectures online, especially because they can slow down and speed up the lecture and look up things while watching. To give a simple but illustrative example, the coeditor of this book was telling his daughter about paper and how it's actually a fibrous material that looks like a thin layer of spaghetti under an electron microscope. Before finishing his sentence, his daughter was watching such videos on YouTube. What surprised the coeditor was that she went directly to YouTube rather than to an Internet search engine or to Wikipedia.

Now let's try to imagine what a kid will do in 10 years when trying to understand something new, such as an internal combustion engine works. Hopefully this will be an ancient relic that arouses her curiosity since electric propulsion will completely displace such engines. She'll slip on her virtual reality or augmented reality goggles, or perhaps use a holographic projector to show the engine. She'll be able to rotate the engine, look at the different parts, and then with a simple gesture, she'll be able to blow out the engine into thousands of parts. She can then put back the engine and just look at a few components, say inside the engine block, and play with the pistons and see how they move up and down and generate a rotational motion through the crank shaft. She'll be able to learn a tremendous amount in a short period of time. Clearly, this data will have to be downloaded from the Internet and played back in real time. Maybe she's repairing a classic automobile and needs to see the 3D images again while she's in the garage. Remember that a single base station will need to serve hundreds or thousands of curious kids, all at the same time.

In certain situations, the demands on the network will explode. Imagine a classroom full of thousands of students learning anatomy. The professor will have a virtual cadaver in front of him and he'll be making incisions and demonstrations of different parts fit together. Every student will have his or her own virtual cadaver as well. In fact, there's no need to use an inanimate body, because a virtual body that is alive and moving is much more interesting, for example to understand how muscles and connective tissue work together to enable different motions. In this scenario, we have thousands of simultaneous three-dimensional (3D) high-definition (HD) connections, all in the same geographic location. This is clearly beyond the capabilities of both WiFi and 4G networks today.

At the Berkeley Wireless Research Center (BWRC), we looked at these issues and considered a blank slate to imagine what should the next generation of wireless look like. In December of 2013, we codified our vision with the xG network, shown in Figure 1.3. Our vision is for a new network that utilizes a massive number of antennas in access points to allow a high degree of spatial multiplexing to many different disparate devices,

[2] Such as watching others play video games or watching someone playing with slime.

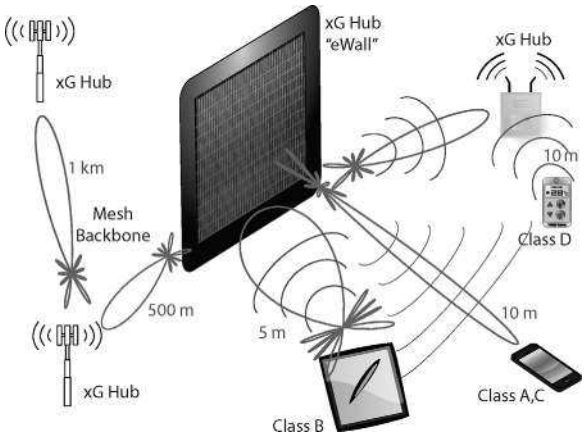

Figure 1.3 The BWRC "xG" vision for the next-generation wireless communication system (December 2013).

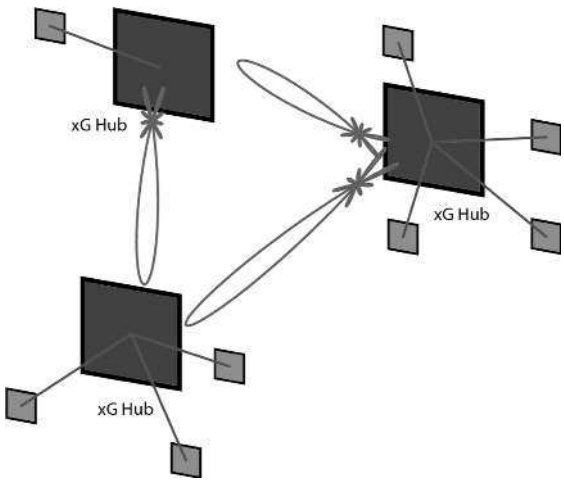

Figure 1.4 Wireless backhaul using phased arrays and mesh networking can reduce the cost of deployment of a 5G system by obviating the need for fiber connectivity.

from cell phones and tablets to Internet of Things (IoT) devices. Multiple RF and mm-wave frequency bands are used in a complementary fashion to form both sharp and broad beams. Also, most importantly, these access points self-backhaul by forming a hierarchical wireless mesh network (Figure 1.4), avoiding the need to use cables or fiber to form the backhaul network. In such a way, the network can grow organically to serve the demands for wireless traffic. The access point can wake up, identify other nodes in the network, and begin routing traffic on demand, with links going up or down in a dynamic fashion, much like the original vision for the Internet and the need for packet switching. In parallel, people started dreaming of 5G and what it should encompass. Many people came to the realization that to serve these visions, we need to utilize higher-

frequency bands to realize higher spectral efficiencies and to circumvent interference, and the idea that 5G would also operate in the mm-wave bands was born.[3]

5G Wishlist

Given this xG vision, which is more or less the same as what people were thinking for 5G, let's enumerate our wishlist a bit more carefully. Are today's networks fast enough both in terms of speed of transmission and latency per user? While a lot of progress has been made on speed, even exceeding 1 Gbps, these improvements are mostly for marketing and don't bear out in practice. But nevertheless, being able to get a mobile wireless connection over 10 Mbps is quite impressive and certainly sufficient for many applications such as video. The problem is that many times we cannot get sufficient coverage and we are all too familiar with video streams coming to a screeching halt at just the right moment. The other issue with today's networks is the latency is typically tens to hundreds of milliseconds long, and sometimes even longer. The latency is also unpredictable, making it difficult to design a closed-loop control system. For this reason, many applications that could benefit from wireless technology have not embraced wireless. Examples include industrial control, semiautonomous driving, multiuser gaming, and virtual reality and augmented reality devices driven from the cloud.

While speed is definitely a great marketing specification, another revolution is under way, the proliferation of low-cost devices with wireless connectivity. This is the well-known and much anticipated IoT revolution, which requires very small footprint and low-power wireless connectivity, and in most cases the speed is not an issue. More important than speed is the power consumption. Today people are adding Bluetooth, Bluetooth Low Energy (BTLE), Zigbee, WiFi, or other radios for wireless connectivity. These radios are short range and cannot actually connect to the Internet without a nearby access point (such as a WiFi router connected to the Internet). Why not just put LTE radios in such devices? The problem with LTE is cost and power consumption, and a lack of a clear business model. For example, many smart watches today have an LTE radio inside but suffer from poor connectivity and require frequent recharging, and each device requires registration with the carrier (and a not-so-insignificant fee per month). Clearly this does not lend itself well to IoT, where we imagine thousands of devices operating on small coin cell batteries.

This brings us to another point. WiFi technology has advanced tremendously in the past 20 years, and for a long time there was a clear boundary between mobile carrier connectivity and wireless connectivity with WiFi and Bluetooth. But today the boundary is blurring, and in many cases these technologies compete. In a crowded café, dozens of users are streaming video from the Internet and one may find that LTE outperforms WiFi. LTE technology operates in licensed spectrum and interference is managed much better than in WiFi unlicensed spectrum, where the access point may only have control over a subset of devices operating in the same band. In many ways, both WiFi and

[3] Samsung, "Pioneer in 5G Standards, Part 1: Finding the 'Land of Opportunity' in 5G Millimeter-Wave." http://bit.ly/2GBDoiA.

mobile standards have converged, for example the use of OFDM to manage equalization in a wideband channel, power control for interference mitigation, similar modulation and coding schemes, and MIMO. LTE even now operates in unlicensed bands, and carriers are encouraging users to use the WiFi infrastructure to relieve traffic demands on the operator. So why should we have two standards if they are so similar? While it's unlikely that WiFi and LTE will ever merge into one standard (politics alone will prevent this from happening), we could wish for more interplay and compatibility between the radios. Too often we are frustrated by our wireless devices not connecting to the Internet only to find that the WiFi has taken over without truly connecting to the Internet. Many users have to actually manually shut off their WiFi on a daily basis to prevent their phones from connecting to a weak network. The situation has worsened because traditional broadband carriers are trying to compete with the wireless carriers by deploying citywide outdoor WiFi networks.

All of these problems arise because today's mobile networks are simply not up to the task of serving the exponentially growing needs of our modern devices. The spectral capacity of today's wireless networks are in fact operating close to Shannon capacity limits, and MIMO techniques are not as effective in outdoor channels (see Chapter 3). In dense urban environments, this is especially problematic because of high population densities (about 7,282 people in a square kilometer in San Francisco). If 10% of the population is actively watching videos at a given time in a given square kilometer (25 Mbps per HD stream), then the base station has to have a capacity of over 18 Gbps. To serve that much data with a 100 MHz swath of spectrum translates into a spectral efficiency of 180 bits/Hz, which is impossible without enormous signal-to-noise ratio (SNR) in a single channel scenario (not using MIMO). Base stations could be deployed over increasingly smaller areas to solve this problem, but then we are plagued with interference and cost barriers. On the other hand, massive MIMO demonstrations have already showed nearly 100 bits/Hz of spectral capacity in a multiuser MIMO scenario, which is a technique that can improve the aggregate capacity of a system rather than the per-user capacity, and this is an exciting technology on our wishlist for 5G. The other approach is to just go to higher carrier frequencies where wider bandwidths make it easier to serve high data rates. Higher frequencies have propagation issues but offer the ability to use beam-forming to reduce interference as well.

Finally, let's consider the enormous cost to deploy a new network, especially a network with an order of magnitude more base stations to serve dense urban environments. Such an investment should pay off in less than a decade to allow the operators to be profitable. This means that the cost of base stations has to go down, especially in terms of rents on property, backhaul access, and electricity costs. Since mm-wave radios are shorter range, one can anticipate a 10 $\times$ densification of base stations, which must be accommodated by a 10 $\times$ reduction in building a base station. For this to happen, wireless backhaul is a must, as many locations cannot be served by fiber without tearing up the streets and installing new access. Also, wireless backhaul using a phased array, rather than a dish, is clearly advantageous to reduce the setup cost for a new base station. A phased array can dynamically find other nodes and point the beam appropriately, whereas a fixed point-to-point link requires precision antenna alignment. Even a massive

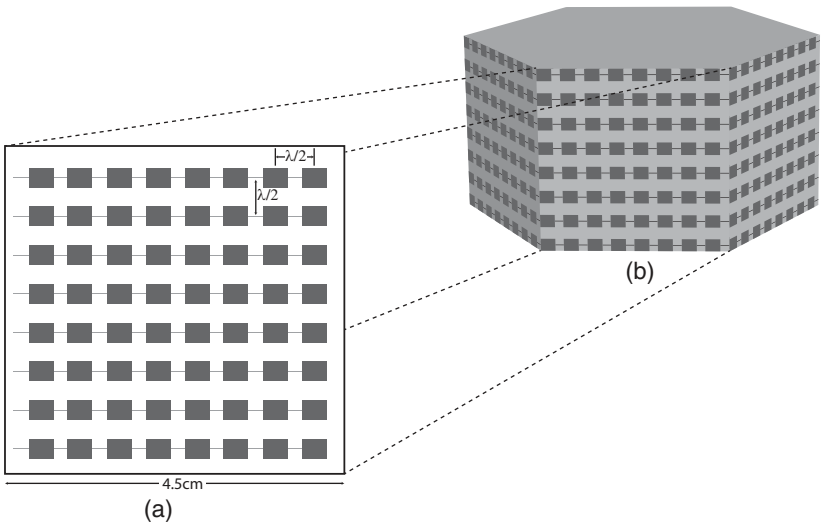

Figure 1.5 An array of 64 patch antennas only occupies an area of 4.5 cm by 4.5 cm as shown in (a). In (b), an array of such panels forms a four-sector base station that can serve thousands of users simultaneously using multiple spatial beams.

array of antennas in mm-wave bands does not occupy much area. Take a linear array consisting of 64 elements, or antenna subarrays, each with 8 elements, as shown in Figure 1.5. Even at 28 GHz, one of the lowest frequency mm-wave bands, the size of the array is $8 \cdot \lambda/2 \times 8 \cdot \lambda/2$ or about 4.5 cm $\times$ 4.5 cm. A base station may consist of a half a dozen of such panels, which means the entire base station could fit in a cube with an edge.

The Cloud

Today we have an enormous amount of data moving from edge devices (say your mobile phone) all the way up the cloud, a room full of servers running the applications. These data have to move back and forth, and it means a lot of data transport over hundreds of kilometers and also a lot of latency. For example, using the web service cloudping.info, the measured ping speed from a mobile phone to the Amazon Web Services is around 50 ms, whereas the ping speed to the carrier is only 25 ms. Clearly any applications such as gaming or augmented reality require millisecond delays, both for health concerns (to avoid making people dizzy) and to make the experience more real. If we could run applications much closer to edge devices, we could greatly improve the latency. This is exactly what people are proposing in industry, putting the servers in base stations, or rather moving base stations into server racks. To keep costs low and allow base station densification, the base station is split into a remote radio head and then backend processing is moved offsite into a server rack. This architecture has other benefits, such as making the network more software defined and flexible. Traffic to/from remote radio heads can be managed on the fly, serving demand (a stadium during the game) when and where it's needed.

Recently there's a lot of buzz around the concept of full duplex communication. Full duplex means a radio can transmit and receive at the same time, in the same bandwidth. Traditionally this was achieved with a circulator or isolator, or a nonreciprocal element. A circulator is a three-port device that allows both the PA and low-noise amplifier (LNA) to be connected to the antenna without any interference (or only a small amount of the transmitter signal leaks into the receiver). A four-port hybrid can do the same thing, at the cost of insertion loss. A practical circulator has loss too, but there's no fundamental limit to how low this loss can be. On the other hand, more importantly, circulators are bulky and narrowband, and cannot be integrated into a chip due to the need for non-linear magnetics. Recently the coauthors of this book have demonstrated new CMOS-compatible architectures for circulators, and these are described in detail in Chapter 4. Another active cancellation approach, which is applicable to novel full-duplex systems and also traditional frequency division duplex (FDD) systems is presented in Chapter 5. FDD is common today and allows simultaneous transmit and receive in two nearby bands by the application of a sharp filter, or duplexer, to provide isolation. These filters are also band-specific and difficult to integrated into CMOS. This chapter will take a different route and use active impedance synthesis to cancel the transmit signal in the receive band.

While much of the buzz around 5G is in the new mm-wave bands, such as 28 GHz and 39 GHz, the 60 GHz band will likely play an equally important role as an unlicensed spectrum, much as WiFi today plays a complementary (and sometimes competing) role to LTE. The amount of bandwidth available in the 60 GHz band is enormous, and we are witnessing multiband radios that can pump tens to hundreds of gigabytes per second through this spectrum. This capability will enhance local area networks and provide backhaul mesh networking to 5G systems. In Chapter 8, we describe the latest chipset, which can push the limits of CMOS in the 60 GHz band to demonstrate record data transfer speeds.

1.2 Radar

Advanced Driver Assistance Systems (ADAS) are all systems to support the driver for safety and enhance driving convenience. There is a strong and important focus on safety, as many if not most accidents are a result of human behavior or error. Consequently, the ultimate goal of ADAS is to avoid any kind of accidents or collisions, by facilitating automated systems ranging from obstacle detection (e.g., vehicle, parking, pedestrian, etc.) to traffic sign detection and driver monitoring (e.g., drowsiness) or communication (car-to-car, car-to-infrastructure; see Figure 1.6). A key technology of ADAS is the detection of any kind of obstacles by specialized radar systems.

The use cases for automotive radar are diverse and include the following scenarios that demand specific requirements on the detection device:

Adaptive cruise control (ACC) is applicable in normal driving conditions to adapt the drive speed to the cars ahead as well to detect obstacles in the far distance to avoid