Chapter 8

5G Networks

5G is the fifth and latest generation of cellular networks that has just started to roll out in 2019-2020. While the previous four generations mainly sought to improve the data rate and capacity of the cellular systems, 5G is designed to improve several other aspects of communications and connectivity beyond the data rates. This chapter discusses the new applications promised by 5G and the networking technologies behind them.

8.1 Key 5G targets

5G promises to massively surpass 4G in the following three main categories:

1. **Data Rate:** While 4G offered the maximum data rate of 1Gbps per user under ideal conditions, 5G promises 20Gbps under the same conditions.
2. **Latency:** Radio contribution to latency between send and receive is an important metric for any wireless networks. Latencies of cellular networks have been very high in the past; typically, about 100ms with 3G and then improved to 30ms in 4G. 5G promises latencies as low as 1ms.
3. **Connection Density:** Number of devices per km^2 that can connect to a cellular base station becomes important as more and more devices need wireless connectivity. While 4G was able to connect only 100 thousand devices per km^2, 5G promises to increase that number to 1 million.

8.2 New applications enabled by 5G

Massive improvements in data rates, latency, and connection density are expected to enable new applications in the following key areas:

1. **Enhanced Broadband.** The huge data rates of 5G have made cellular networks a viable option for residential broadband, which is also referred to as fixed wireless. With fixed wireless, no cabling is required to provision broadband service to the home. A home wireless router can simply be connected to the nearest 5G tower using a SIM card. With high data rates, 5G mobile devices can enjoy new video standards, such as 4K streaming, augmented reality, virtual reality, and blazing fast photo uploads.

2. **Ultra-reliable low latency communications.** Latencies below 1ms will support real-time control of any devices, with will enable new applications such as industrial robotics, autonomous driving, remote medical procedures, and so on.
3. **Massive machine-to-machine (M2M) and Internet of Things (IoT) Communications:** One million connections per square kilometers will help support many M2M and IoT applications that involve connecting billions of devices at a scale not seen before. This has the potential to revolutionize almost all vertical markets including agriculture, manufacturing, health, and defense.

8.3 5G technologies

To meet the massive capacity and data rate increase targets, enhancements will be made in the three fundamental areas:

Increase bps/Hz or spectral efficiency: develop new coding and modulation techniques as well as new spectrum sharing methods to squeeze more bits out of the given spectrum. Enhancements in this sector of R&D will linearly increase the capacity. For example, increasing bps/Hz by a factor of 2 will directly double the capacity of a given cell.

Reduce cell radius or increase spectral reuse: By reducing the cell size, the same spectrum can be reused many times in a given service area. This is the most effective method to increase capacity. Cell sizes have been consistently reduced over the 4 generations. 5G will continue to follow this trend.

Use new spectrum: It has been known all along over the four generations that despite advancements in improving spectral efficient and spectral reuse, eventually we will need new spectrum to cope with the increasing demand for mobile traffic. 5G will be the first generation where new spectrum from the high frequency bands, notable millimeter wave bands, will be used.

Enhancements are also made in spectrum access techniques to address the aggressive new targets for low latency and massive connectivity. In the rest of this chapter, we discuss some of the key new developments in 5G to address these challenges.

8.4 Non-Orthogonal Multiple Access (NOMA)

NOMA [ALDA2018] proposes to use the power as the fourth dimension for multiplexing. The previous generations used only orthogonal multiple access in the sense that the same communication resource could not be allocated to multiple users at the same time. Remember that initially, in 1G, only frequency was used to separate users using the so-called frequency division multiple access (FDMA). Then in 2G, time was used as the 2nd dimension using the concept of time division multiple access (TDMA). In 3G, code was introduced in the form of code division multiple access (CDMA) as a 3rd dimension to separate users who are using the same frequency at the same time. In 4G, OFDMA simply introduced highly flexible multiplexing techniques for the time and frequency dimensions. Use of power as a tool to separate users who are using the same frequency at the same time has not been implemented so far. With increasing handset computational powers, it has now become feasible to consider this.

In NOMA, multiple users' signals are superimposed at the transmitter side using different power coefficients for different users based on their individual channel conditions. For example, a receiver closer to the transmitter is likely to be allocated higher power (due to higher channel coefficient) compared to the one located further away from the transmitter (lower channel coefficient). The superimposition of signals for N users can be mathematically described as follows:

$$X(f,t) = \sum_{i=1}^{N} \sqrt{a_i P}\, x_i(f,t) \quad (8.1)$$

Where f is the frequency, $x_i(f,t)$ is the information signal for user i, P is the transmitter power, and a_i is the power coefficient for user i subjected to $\sum_{i=1}^{N} a_i = 1$ and $a_1 \geq a_2 \geq \cdots \geq a_N$ when channel gains are assumed to be ordered as $h_1 \leq h_2 \leq \cdots \leq h_N$.

At the receiver side, successive interference cancellation (SIC) is applied for decoding the signals one by one until the desired user's signal is obtained. SIC relies on the signal processing that allows a receiver to immediately decode the signal with the highest power by considering all other signals as noise. Once the signal with the highest power is decoded, it can be subtracted and removed from the combined signal. The signal with the second highest power now becomes the most powered signal in the residual superimposed signal, hence can be decoded using the same technique. Thus, any receiver can use SIC to actually decode all the signals, but they stop decoding as soon as they receive their own signals.

The concept of NOMA and the associated SIC process is illustrated in Figure 8.1, where a base station serves N users located at different distances from the base station. Using the same frequency, the base station transmits a combined superimposed signal with the highest power allocated to the farthest user (the worst channel), U_1, and the lowest power to the closest one (the best channel), U_N. U1 decodes the highest-powered signal easily and stops the decoding process because the decoded signal is addressed to it, i.e., it does not really carry out SIC. U_2 employs SIC once to remove the signal for U_1 before detecting its own signal, which is the second highest powered signal. U_3 has to repeat SIC twice and so on.

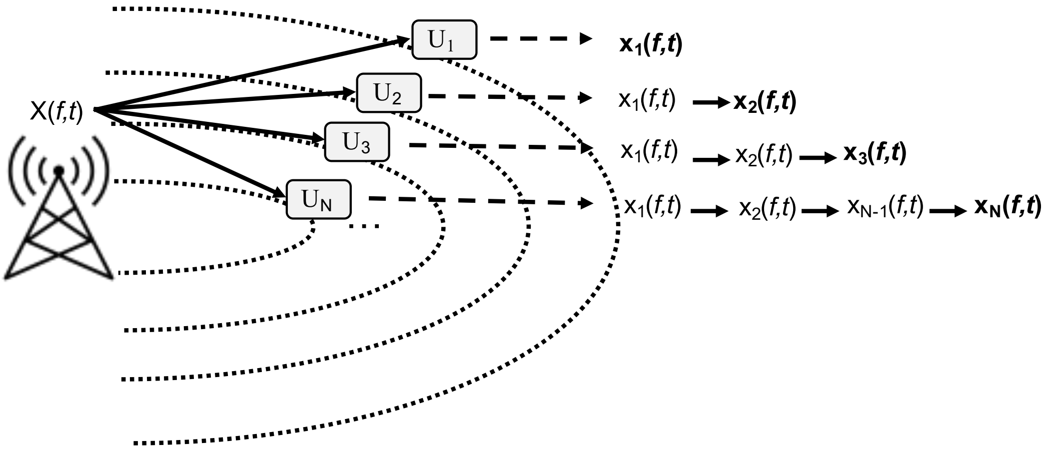


Figure 8.1 Non-Orthogonal Multiple Access (NOMA)

8.5 Full-duplex wireless

Recall that for FDD, separate frequencies have to be allocated in uplink and downlink to achieve full-duplex communication. For a single frequency, full-duplex had not been possible so far due to the transmitter overwhelming the receiver causing too much interference as illustrated in Figure 8.2. Therefore, if single frequency is to be used for both DL and UL, then it would have to be half-duplex, like it is in TDD. In that case, when the frequency is used for DL, there is no traffic allowed in UL, and vice versa. Clearly, half-duplex reduces capacity and increases latency.

With advancements in DSP and processing powers, it is now contemplated to implement self-interference cancellation to realize full-duplex over the same frequency, so that simultaneous transmission and reception may be possible [FDUPLEX2011]. Figure 8.3 illustrates how the self-interference can be conceptually cancelled through additional signal processing and circuits implemented within the wireless radio. Basically, an attenuated and delayed transmit signal should be combined with the received signal to cancel the interference within the received signal that was caused by the over-the-air interference from the transmitting antenna. Such full-duplex communication would double the throughput, reduce end-to-end latency, and allow transmitters to monitor (estimate) the channel.

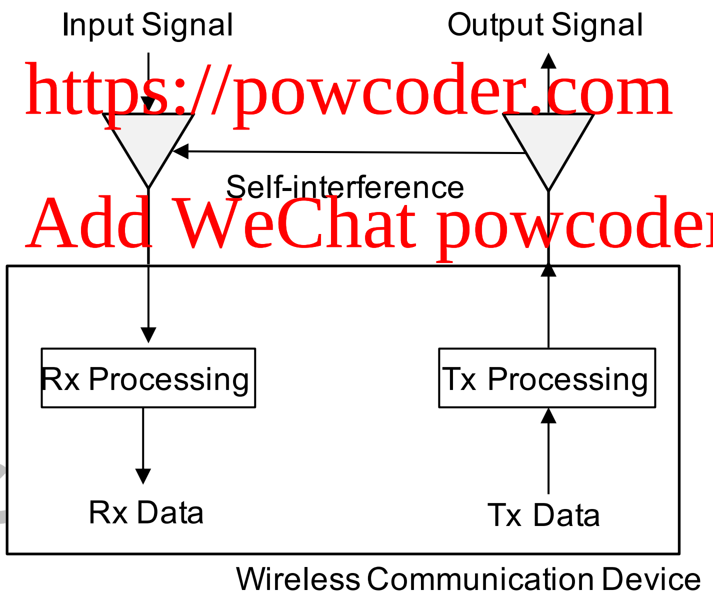


Figure 8.2 The self-interference problem in wireless communications

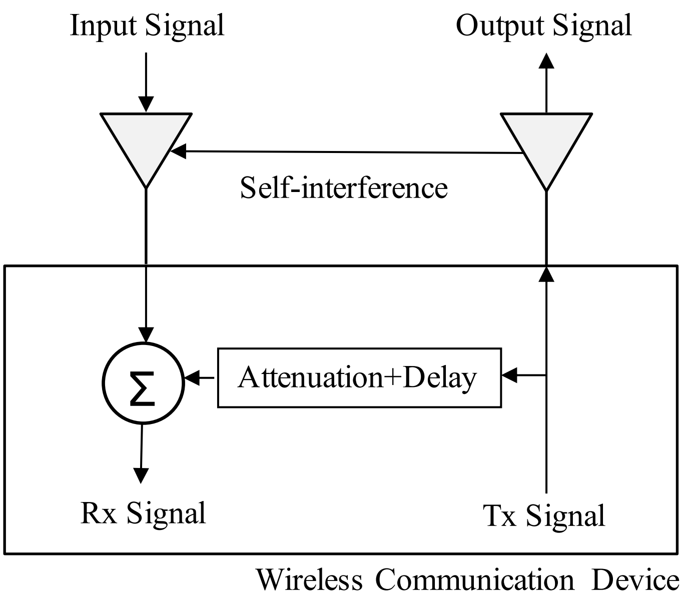
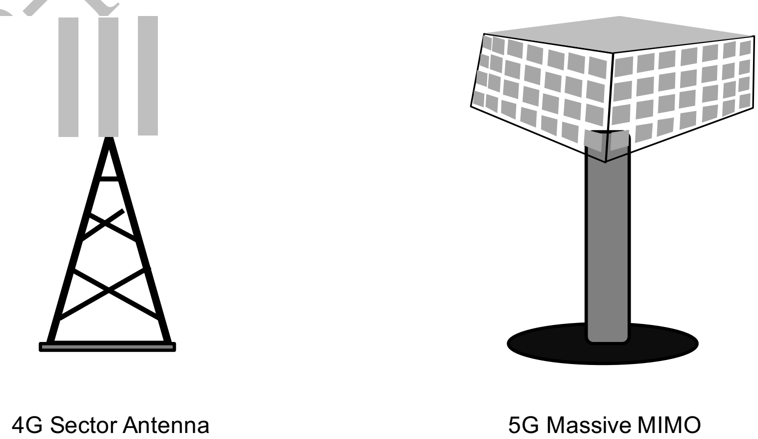


Figure 8.3 Self-interference cancellation for full-duplex wireless communications

8.6 Massive MIMO and 3D Beamforming

Most of the current radio towers use vertical sector antennas to focus energy in the 2D horizontal plane to serve people on the ground. Increasingly, people are now living in high rise apartments. With popularity of drones, cellular networks are also facing the issue of connecting devices that may fly above the ground. Thus, new mechanisms are required to reach devices that are spread in 3D.

5G is expected to serve users in 3D coverage spaces by using a new type of base stations that use massive MIMO and 3D beamforming [[5GMIMO]] as shown in Figure 8.4. Instead of using a few vertical antennas to cover geographical sectors on the ground, 5G base stations are expected to deploy planar arrays with many (>100) antenna elements. By configuring the phase and amplitude coefficients of each elements, the base station can form many beams of different shapes in both vertical (elevation) and horizontal (azimuth) planes.



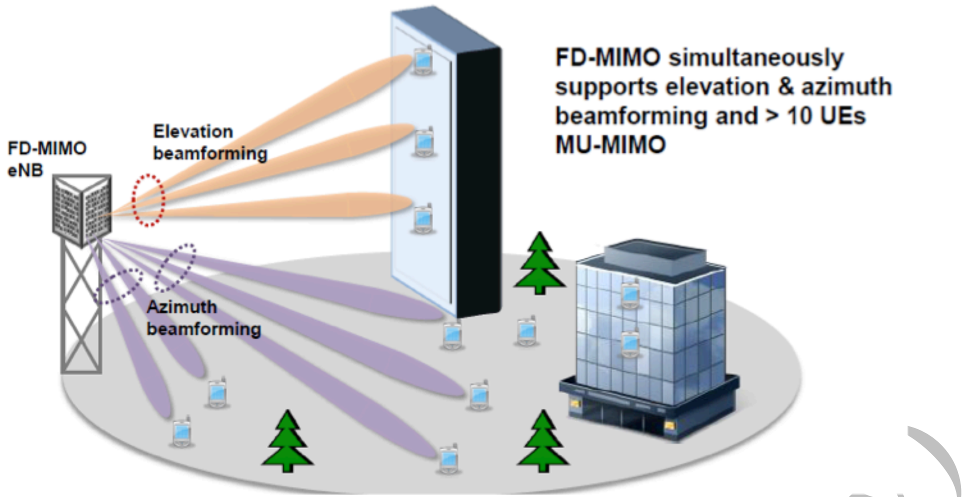


Figure 8.4 Antenna shapes for 4G vs. 5G (top) and 3D beamforming with massive MIMO (bottom)

8.7 Mobile Edge Computing (MEC)

In future, mobile phones will need access to many computations that are not feasible to do in the handset. For example, speech recognition, augmented reality, and so on. However, sending these computation tasks to the cloud, which may be far away from the handset, would be costly and increase latency. Also to service IoT, where many machines may need some computing help from the cloud, the computation needs to come closer to the device. The idea behind MEC [MEC2018] is to store such computing resources, a *mini cloud*, in every radio tower to make this very efficient. This concept is shown in Figure 8.5.

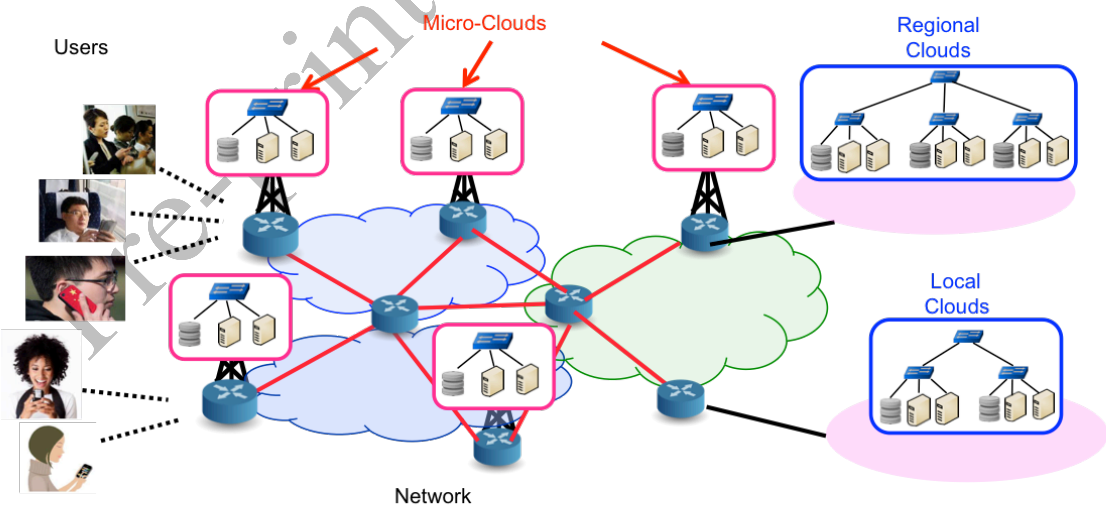


Figure 8.5 Mobile Edge Computing (MEC)

8.8 New Spectrum

Finally we come to the point when we must discuss the opportunities for new spectrum. All those spectrum efficiency and spectrum reuse factor enhancements techniques we have discussed so far will help improving the capacity, but eventually we will need access to new spectrum to keep increasing the capacity.

While previous generations used frequencies in the highly congested bands below 6GHz, there are plenty of spectrum available at higher frequencies, between 6-100 GHz, which is also referred to as high band. In this high band, 26GHz and 28GHz have emerged as two of the most important 5G spectrum bands [5Gmmwave] as they can be utilised with the minimal user equipment complexity. These bands are also called millimeter wave (mmWave) bands as their wave lengths are close to 1 mm.

Use of mmWave bands will give 5G the much-needed spectrum boost to address the massive capacity increase targets. As the antenna size is proportional to the wavelength, the mmWave band will facilitate building massive MIMO base stations with hundreds of small antenna elements for efficient beam forming. However, signals at such high frequencies need line-of-sight for good performance, which will force 5G to exploit them for high data rate short distance communications.

8.9 Chapter Summary

1. 5G is being launched in 2020 promising to offer ultra-high data rates, ultralow latency, and massive connectivity for Internet of Things
2. 5G will use NOMA as a new access technology that enables serving multiple users over the same frequency at the same time; NOMA uses power as a new dimension to differentiate users.
3. 5G promises full-duplex wireless communications where both the Tx and Rx antennas can function at the same time.
4. 5G base stations will use planar array antennas for massive MIMO and 3D beamforming.
5. 5G base stations will host computing and storage resources to reduce latency for applications requiring cloud support.
6. 5G will use new spectrum in the mmWave band.

References

[ALDA2018] Mahmoud Aldababsa, Mesut Toka, Selahattin Gökçeli, Güneş Karabulut Kurt, Oğuz Kucur, "A Tutorial on Nonorthogonal Multiple Access for 5G and Beyond", *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 9713450, 24 pages, 2018. https://doi.org/10.1155/2018/9713450

[FDUPLEX2011] Jain et al., “Practical, Real-time, Full Duplex Wireless,” ACM Mobicom 2011.

[5GMIMO] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou and W. Hong, "Digital Beamforming-Based Massive MIMO Transceiver for 5G Millimeter-Wave Communications," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 7, pp. 3403-3418, July 2018, doi: 10.1109/TMTT.2018.2829702.

[MEC2018] N. Abbas, Y. Zhang, A. Taherkordi and T. Skeie, "Mobile Edge Computing: A Survey," in *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450-465, Feb. 2018, doi: 10.1109/JIOT.2017.2750180.

[5Gmmwave] S. Sun, T. S. Rappaport, M. Shafi, P. Tang, J. Zhang and P. J. Smith, "Propagation Models and Performance Evaluation for 5G Millimeter-Wave Bands," in *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8422-8439, Sept. 2018, doi: 10.1109/TVT.2018.2848208.

End of Chapter 8