CHAPTER

NEW 5G RADIO-ACCESS TECHNOLOGY

24

As described in the previous chapter, the overall 5G wireless access solution will consist of the evolution of LTE in combination with a new 5G radio-access technology (5G RAT).

In this chapter some key design principles and main technology components relevant for the new 5G RAT are discussed. As the detailed specification of the new 5G RAT has not even started in 3GPP at the time of this writing, there is obviously still a high degree of uncertainty in the details of this future technology. However, on a higher level, including general design principles and basic technology components, there seems to be a relatively high degree of commonality in the views of the major players in the industry.

24.1 5G: SOME GENERAL DESIGN PRINCIPLES

24.1.1 RADIO-ACCESS EVOLUTION AND FORWARD COMPATIBILITY

After their initial introductions, every cellular radio-access technology has gone through a sequence of evolutionary steps adding new features that provide enhanced performance and new capabilities to the technology. Examples of these evolutions include the evolution of GSM to EDGE, the evolution of WCDMA to HSPA, and the different steps of the evolution of LTE. In general, these evolutionary steps have been backward compatible implying that a legacy device can still access the network on a carrier frequency supporting the new features, although it may obviously not be able to fully benefit from the features as such.

For the new 5G RAT, the possibility to evolve the technology beyond its initial release will be even more important:

- As described in the previous chapter, 5G is envisioned to support a wide range of different use cases, many of which are yet unknown. Thus, the radio-access solution will have to be able to evolve and adapt to encompass new requirements and new service characteristics.
- As also described in the previous chapter, there will be a phased approach to the 3GPP specification of the new 5G RAT, with the initial phase having a relatively limited scope and later evolution ensuring full compliance with all identified 5G requirements.

4G, LTE-Advanced Pro and The Road to 5G. http://dx.doi.org/10.1016/B978-0-12-804575-6.00024-8

The new 5G RAT is not required to be backward compatibility to earlier generations. However, similar to earlier generations, the future evolution of the technology should be backward compatibility to its initial release. In order to minimize the constraints on this future evolution the concept of *forward compatibility* has been introduced as an additional requirement for the design of the new 5G RAT. In this context, forward compatibility simply means that the design of the radio-access technology should be such that the constraints on the future evolution of the technology, due to the requirement on retained backward compatibility to its initial release, will be as limited as possible.

Due to the obvious uncertainty of the characteristics and requirements of new, yet unknown applications and use cases, as well as the uncertainty of future technology directions, forward compatibility is inherently difficult to achieve. However, as described in the following sections, there are certain principles that, if followed, will at least enhance forward compatibility.

24.1.2 ULTRA-LEAN DESIGN: MINIMIZE "ALWAYS-ON" TRANSMISSIONS

For any cellular technology there are certain transmissions carried out regularly from every network node regardless of whether or not there is any ongoing user-data transmissions and even if there are no active devices at all within the coverage of the node. In the context of LTE such "*always-on transmissions*" include

- the primary and secondary synchronization signals;
- the cell-specific reference signals;
- the broadcast system information (MIB and SIBs).

The opposite of always-on transmissions are "*on-demand transmissions*"—that is, transmissions that can be initiated and deactivated on a per-need basis.

In high-traffic scenarios with a high traffic load per network node, which is the typically assumed scenario when evaluating cellular radio-access technologies, the always-on transmissions contribute only a relatively small fraction of the total node transmission and thus have relatively small impact on the overall system performance. However, in real-life cellular networks a large fraction of the network nodes is actually relatively lightly loaded on average:

- Especially in suburban and rural areas, network infrastructure (base stations) is typically deployed to provide coverage with a certain minimum end-user data rate and not because more network capacity is needed to handle the traffic volumes.
- Even when deployment of new infrastructure is driven by a need for more network capacity the deployment must be dimensioned to handle peak-traffic volumes. As traffic volumes typically vary significantly in time, the *average* load per network node will still be relatively low.

Furthermore, with data traffic typically being very bursty, even at relatively high load a fairly large number of subframes are actually not carrying any traffic.

As a consequence, in real-life networks always-on transmissions often have bigger impact on the overall system performance than what can be seen from "high-load" evaluations:

- The always-on transmissions will add to the overall system interference, thereby reducing the achievable data rates.
- The always-on transmissions will increase the overall network energy consumption, thereby limiting the network energy efficiency.

Expressed differently, minimizing the amount of always-on transmissions is an important component to enable very high achievable data rates and very high network energy efficiency.

The LTE small-cell on/off mechanism described in Chapter 15 is a step in the direction of minimizing the always-on transmissions. However, the small-cell on/off mechanism is still constrained by backward compatibility and the requirement that legacy LTE devices should still be able to access the carrier. The introduction of a new 5G RAT not constrained by backward compatibility to earlier technologies provides additional opportunities in this respect.

Minimizing the amount of always-on transmissions is also one important component for forward compatibility. As legacy devices expect the always-on transmissions to be present, such transmissions cannot be removed or even modified without impacting legacy devices and their ability to properly access the system.

The relation between always-on transmissions and forward compatibility is well illustrated by the use of MBSFN subframes to enable relaying functionality in LTE, see Chapter 18. Although introduced in order to provide support for MBSFN transmission for MBMS services, the MBSFN subframes have turned out to be very valuable for the evolution of LTE in general. The reason for this is simply that the MBSFN subframes include substantially less cell-specific reference symbols compared to normal subframes. By configuring subframes as MBSFN subframes from a legacy-device point of view, these subframes can be used for any kind of new transmissions without breaking backward compatibility. In the specific case of relaying, the almost-empty MBSFN subframes made it possible to create sufficiently large "holes" in the downlink transmission to enable in-frequency reception on the backhaul link. Without the MBSFN subframes this would not have been possible with retained backward compatibility as too many OFDM symbols would have included always-on cell-specific reference symbols. The key thing that allowed for this was that MBSFN subframes have been part of LTE specifications already since the first release implying that they are "understood" by all legacy devices. In essence, the baseline assumption from a device perspective on the new 5G RAT should be to treat each subframe as empty (or undefined) unless it has been explicitly instructed to receive or transmit.

Minimizing always-on transmissions can be seen as part of a more high-level *ultra-lean design principle* often expressed as *minimize all network transmissions not directly related to user-data transmission* as illustrated in Figure 24.1. The aim is, once again, to enable higher achievable data rates and enhance the network energy efficiency.

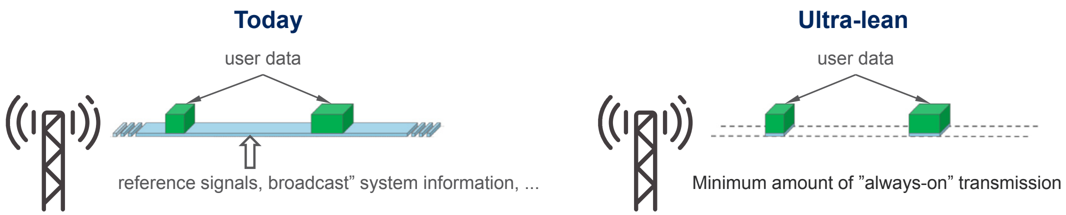

FIGURE 24.1

Ultra-lean transmission.

24.1.3 STAY IN THE BOX

The "stay-in-the-box" principle in essence says that a transmission should be kept together as illustrated in the right part of Figure 24.2 and not be spread out over the resource space (the time−frequency grid in case of OFDM) as shown in the left part of the figure. The aim is once again to enable a higher degree of forward compatibility. By keeping transmissions together, it is easier to later introduce new types of transmissions in parallel to legacy transmissions while retaining backward compatibility.

An example of LTE transmissions not fulfilling the "stay-in-the-box" principle is the set of physical channels (PCFICH, PHICH, and PDCCH) transmitted in the control region of each LTE subframe. As described in Chapter 6, each PDCCH/PHICH/PCFICH transmission is spread out over the entire carrier bandwidth in an apparently random way. The main reason for this structure was to enable a high degree of frequency diversity and to achieve randomization between transmissions. However, this also makes it very difficult to introduce new transmissions within the LTE control region unless they are fully aligned with the current control-channel structure. This is, for example, illustrated by the design of NB-IoT (Chapter 20) for which the solution was simply to have NB-IoT downlink transmissions avoiding the entire control region in case of inband deployment.

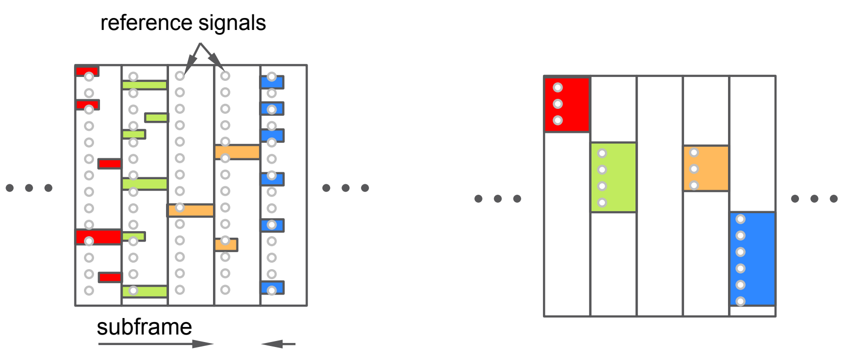

FIGURE 24.2

Illustration of signals spread out (left) vs. "stay-in-the-box" (right).

In contrast, the LTE EPDCCH is much more aligned with the “stay-in-the-box” principle. Each EPDCCH is contained within a single, or at most a few resource blocks making it much more straightforward to introduce new transmissions in parallel to the EPDCCH.

Related to the “stay-in-the-box” principle is the concept of *self-contained transmissions*. Self-contained transmission implies that, to the extent possible, the signals and information required for reception of data in a given beam and subframe are contained within the same beam and subframe. Although this approach provides a great deal of forward compatibility, some aspects, such as channel-estimation accuracy, may benefit from the possibility of exploiting “outside-the-box” signals and the final design must carefully take this into account.

24.1.4 AVOID STRICT TIMING RELATIONS

Another important design principle is to avoid static and strict timing relations across subframe borders as well as between different transmission directions.

An example of such static and strict timing relations is the LTE uplink hybrid-ARQ procedure where downlink acknowledgments as well as potential uplink retransmissions occur at fixed predefined time instants relative to the initial uplink transmission. These kinds of timing relations make it more difficult to introduce new transmission procedures that are not aligned in detail with the legacy timing relation. For example, the original LTE uplink hybrid-ARQ timing is not well matched to unlicensed spectrum and changes are required as discussed in Chapter 17. Avoiding strict and static timing relations is thus an important component for forward compatibility.

Strict and static timing relations may also prevent a radio-access technology to fully benefit from technology advances in processing capability. As an example, the LTE hybrid-ARQ timing was specified based on the processing capabilities, primarily in terms of channel decoding, estimated to be available at the time of the first commercial deployments of LTE. With time, the processing capabilities have advanced, allowing for faster decoding. However, the static timing relation of the uplink hybrid-ARQ protocol prevents this from being turned into a shorter hybrid-ARQ round-trip time and an associated lower latency.

24.2 5G: KEY TECHNOLOGY COMPONENTS

In the remainder of the chapter various technology components considered for 5G radio access are discussed, taking the design principles into account.

24.2.1 WAVEFORM

24.2.1.1 Scalable OFDM

At the core of the LTE RAT is an OFDM-based transmission scheme with a subcarrier spacing of 15 kHz and a cyclic prefix of about 4.7 μs.[1] This is true for both downlink and uplink,

[1]As described in Chapter 5 there is also an extended cyclic prefix of about 16.7 μs.

Table 24.1 Example of Scalable OFDM Numerology

	Baseline	Higher-Order-Derived Numerologies	
Scale factor	1	4	32
Subcarrier spacing	15 kHz	60 kHz	480 kHz
Symbol time (excl. CP)	66.7 μs	16.7 μs	2.1 μs
Cyclic prefix	4.7 μs	1.2 μs	0.15 μs
Subframe	500 μs	125 μs	15.6 μs

although additional DFT precoding is applied for uplink data (PUSCH) transmission to enable higher power-amplifier efficiency on the device side.

OFDM is the main candidate also for the new 5G RAT, both in uplink and downlink. Having the same waveform in both directions simplifies the overall design, especially with respect to wireless backhauling and device-to-device communication. However, in light of the very wide range of spectrum, deployment types, and use cases to be addressed by 5G, it is unrealistic to assume that a single OFDM numerology should be sufficient for the new 5G RAT.

As discussed in Chapter 23, 5G radio access should cover a very wide range of frequencies, from below 1 GHz up to at least several 10 GHz, and possibly as high as 70–80 GHz. For the lower part of the spectrum, perhaps as high as 5 GHz, a subcarrier spacing of the same order as LTE is sufficient. However, for higher frequencies, a larger subcarrier spacing is needed in order to ensure sufficient robustness to, especially, phase noise with reasonable cost and power consumption for mobile devices.

The new 5G RAT should also be able to operate in a wide range of deployment scenarios, ranging from extremely dense indoor and outdoor deployments to sparse rural deployments where each network node may cover a very large area. In the latter case a cyclic prefix similar to LTE or perhaps even larger is needed to handle large delay spread while, in the former case, a smaller cyclic prefix is sufficient.

The wide-area deployments for which a larger cyclic prefix is needed will typically operate at lower frequencies for which a lower subcarrier spacing is sufficient. At the same time, higher frequencies, for which a larger subcarrier spacing is needed, will typically be limited to denser deployments where a smaller cyclic prefix is sufficient. This speaks in favor of a *scalable* OFDM framework with different numerologies derived by frequency/time-domain scaling of a common baseline numerology.

Table 24.1 shows an example of such scaled OFDM numerology using the LTE numerology as baseline and with two derived higher-order numerologies based on the scale factors 4 and 32, respectively.[2] In this case, the baseline numerology with 15 kHz subcarrier spacing

[2]Note that in Table 24.1. we have redefined the term “subframe” to mean seven OFDM symbols, corresponding to what in LTE is referred to as a “slot.”

and a cyclic prefix of about 4.7 μs would be appropriate for wide-area deployments using lower-frequency spectrum.

In higher-frequency spectrum, the higher-order numerologies with larger subcarrier spacing should be used to ensure high robustness to, especially, phase noise. With a scaled numerology, this inherently leads to smaller cyclic prefix. However, as already mentioned, the use of higher-frequency spectrum will be limited to dense deployments with less delay spread. Furthermore, higher frequencies will typically be used in combination with extensive beam-forming. This will reduce the amount of large-delay reflections, thereby further reducing the delay spread of the received signal.

The benefit with a scaled numerology derived from a common baseline numerology, rather than a set of more independent numerologies, is that a main part of radio-interface specification can be made agnostic to the exact numerology. This also means that one could more easily introduce additional numerologies at a later stage by simply introducing additional scale factors.

Note that the higher-order numerologies with larger subcarrier spacing and smaller cyclic prefix could also be used at lower frequencies as long as the delay spread is limited. Thus, the higher-order numerologies could be used also for low-frequency dense deployments. One reason for doing this would be to enable further reduced latency also at lower frequencies by utilizing the shorter symbol time and corresponding shorter subframe of the higher-order numerologies.

It should be noted that one could use a larger subcarrier spacing also for wide-area deployments with larger delay spread by using an extended cyclic prefix similar to LTE. Assuming the LTE-derived numerology of Table 24.1, for the 4× numerology such an extended cyclic prefix would be roughly 4.2 μs (16.7/4). The drawback would be a significantly higher cyclic-prefix overhead. However, such large overhead could be justified under special circumstances.

It is important to point out that the set of numerologies provided in Table 24.1 is an example. One could obviously consider different scale factors than those of Table 24.1 One could also use a different baseline numerology not aligned with LTE. The use of an LTE-based baseline numerology provides benefits though. One example is that it would allow for deployment of an NB-IoT carrier (Chapter 20) within a carrier of the new 5G RAT in similar way as NB-IoT can be deployed within an LTE carrier. This can potentially be an important benefit for an operator migrating from LTE to the new 5G RAT on a carrier while maintaining support for existing NB-IoT-based massive-MTC devices as such devices typically have a very long lifespan.

24.2.1.2 Spectral Shaping

OFDM subcarriers have large sublobes due to the use of a rectangular pulse shape. Thus, orthogonality between subcarriers is not due to true spectral separation but rather due to the exact structure of each subcarrier. One implication of this is that orthogonality between different transmissions is only retained if the different transmissions are received time aligned within the cyclic prefix. In the LTE uplink this is achieved by devices updating their transmit timing based on timing-advance commands provided by the network initially as part of the random-access response (Chapter 11) and subsequently as MAC control elements on the DL-SCH (Chapter 7).

Although having uplink orthogonality relying on time alignment is, in general, a good approach, it has certain limitations:

- There must be regular uplink transmissions in order for the network to be able to estimate the uplink timing and provide timing-advance commands when needed.
- The need to establish time-alignment before user-data transmission can be initiated leads to additional delay in the initial access and prevents immediate data transmission from a not-yet synchronized device.

Another implication of the large sidelobes of OFDM subcarriers is that a relatively large guard band is needed in case of frequency multiplexing of an OFDM signal with a signal of a different structure. The latter could be a non-OFDM signal. However, it could also be an OFDM signal with a different numerology, for example, a different subcarrier spacing or just a different cyclic prefix.

To overcome these issues, one may consider modifications and/or extensions to OFDM that leads to a higher degree of spectrum confinement of the transmitted signal.

Filter-bank multi-carrier (FBMC) [69] is another kind of multi-carrier transmission, where spectrum shaping by means of filtering is applied to each subcarrier. In essence, the rectangular pulse shape of OFDM is, with FBMC, replaced by a nonrectangular pulse shape with a corresponding more confined frequency response. In order to retain orthogonality between modulated subcarriers, FBMC has to be used in combination with offset-QAM (OQAM) modulation on each subcarrier, rather than conventional QAM modulation as used, for example, in LTE. FBMC provides significantly more confined spectrum, with smaller per-subcarrier sidelobes, compared to conventional OFDM. However, there are also several drawbacks/issues with FBMC.

Although necessary to retain orthogonality between subcarriers, the use of OQAM modulation leads to difficulties with channel estimation, especially in combination with MIMO transmission. There are proposed solutions; however, these lead to degraded receiver performance and additional reference-signal overhead [75]. Another drawback of FBMC is that tight filtering in the frequency domain inherently leads to long pulses in the time domain, with FBMC pulse shaping typically having a length of several symbols. To avoid interference between transmission bursts, a corresponding guard time of a length corresponding to several symbols is needed between each burst. In case of transmission bursts consisting of many symbols, this guard-time overhead will be relatively small. However, for low-latency transmissions, short-burst transmission is needed leading to potentially large overhead.

A different way to improve the spectrum confinement, while maintaining the OFDM structure, is to apply filtering of the entire OFDM signal, see upper part of Figure 24.3. Such

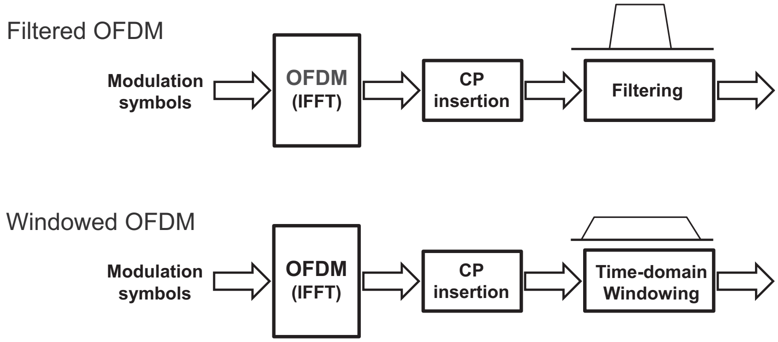

FIGURE 24.3

Filtered OFDM (upper) and windowed OFDM (lower).

filtering will be much less drastic than per-subcarrier filtering, with much less time-domain spreading of the signal. In practice, the filtering will simply use part of the cyclic prefix, making less cyclic prefix available to handle delay spread on the channel.

Alternatively, time-domain windowing can be used to control the spectral properties instead of filtering, see lower part of Figure 24.3. While filtering implies multiplication with a frequency response in the frequency domain or, equivalently, convolution with a time response in the time domain, windowing implies multiplication of the OFDM symbol with a window function in the time domain (convolution in the frequency domain).

In the end, these two approaches (filtering and windowing) lead to very similar result in terms of spectral confinement of the transmitted signal. However, windowing may be associated with somewhat less implementation complexity compared to filtering.

Filtering/windowing of the entire OFDM signal is very much an implementation issue and is actually, in practice, done for LTE already today in order to ensure that the transmitted OFDM signal fulfills the out-of-band-emissions requirements. However, filtering/windowing could also be used to spectrally confine certain parts of a carrier. This could, for example, be used to create “holes” in the spectrum to make room for other non-OFDM transmission. It could also allow for mixing different OFDM numerologies within one carrier, as illustrated Figure 24.4. The later could, for example, be beneficial when different services with different requirements are to be mixed on one carrier. The different numerologies could, for example, correspond to different subcarrier spacing (the case illustrated in Figure 24.4). In this case, the smaller subcarrier spacing could correspond to a conventional mobile-broadband service while the higher subcarrier spacing, which allows for lower latency, could correspond to a latency-critical service. However, the different numerologies could also correspond to the same subcarrier spacing with different cyclic prefix.

24.2.1.3 Low-PAPR Transmission

OFDM with DFT precoding is used for the LTE uplink in order to reduce the cubic metric (CM) [10] of the transmitted signal, thereby enabling higher power-amplifier efficiency on

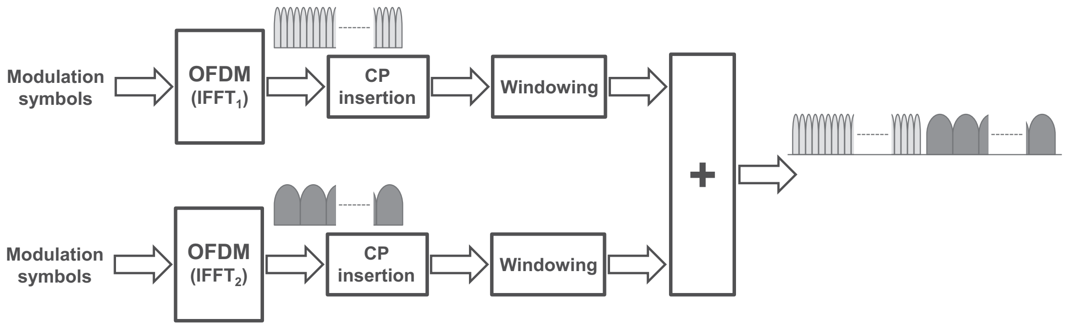

FIGURE 24.4

Mix of numerologies on one OFDM carrier using windowing.

the device side. The drawback of DFT precoding as it is done in LTE is that it limits the flexibility of the transmission. This is apparent, for example, in the design of the uplink reference signals and the design of the uplink control signaling (PUCCH) which are substantially less flexible and at least in some respects more complex than the corresponding downlink transmissions. From this point of view, it would be desirable to avoid DFT precoding, in the way that it is used in LTE, for the new 5G RAT. However, without the possibility for low-CM uplink transmission the new 5G RAT may have an uplink coverage disadvantage compared to LTE when operating in similar spectrum. This could negatively impact the migration of the new 5G RAT into spectrum currently used by LTE. Furthermore, for operation at very high frequencies, especially above 30–40 GHz, high power-amplifier efficiency is even more important compared to operation in lower-frequency spectrum:

- Operation in such high-frequency spectrum will typically be associated with a large number of antennas and, consequently, a large number of power amplifiers, especially at the base-station side.
- The small dimension and tight packing of electrical components at high frequencies make it more difficult to handle of excess heat generated due to power-amplifier inefficiency.

It should be noted that these arguments are relevant also for the base station. Thus, for such high frequencies the possibility for high power-amplifier efficiency on the base-station side may be as important as high power-amplifier efficiency on the device side.

Thus, at this stage one cannot discard the possible need for CM-reduction techniques on top of OFDM for the new 5G RAT. This could be in form of *complementary* DFT precoding that would be applied to data transmission when low CM is of essence, that is, in coverage-limited scenarios. However, it could also be in form of other CM-reduction techniques that can be added on top of OFDM such as, for example, tone reservation [8].

Also, as indicated earlier, any CM-reducing techniques should not only be considered for the uplink (device transmission) but also for the downlink (base-station transmission).

24.2.2 FLEXIBLE DUPLEX

LTE supports both FDD- and TDD-based duplex arrangements in order to match the existence of both paired and unpaired cellular spectra.

Both paired and unpaired spectra will exist also in the 5G era. Consequently, the new 5G RAT will have to support both FDD- and TDD-based duplex arrangement.

As discussed in the previous chapter, the new 5G radio access will cover a very wide range of frequencies, from below 1 GHz up to at least several 10 GHz, see also Figure 24.5. In the lower part of this spectrum, paired spectrum with FDD-based duplex arrangement will most likely continue to dominate. However, for higher frequencies which due to propagation constraints will be limited to dense deployments, unpaired spectrum with TDD-based duplex arrangement is expected to play a more important role.

24.2.2.1 Dynamic TDD

One benefit of unpaired spectrum with TDD-based duplex arrangement is the possibility to dynamically assign transmission resources (time slots) to different transmission directions depending on the instantaneous traffic conditions. This is especially beneficial in deployments and scenarios with more variable traffic conditions, which will, for example, be the case for dense deployments where each network node covers only a very small area.

One of the main concerns with unpaired spectrum and TDD operation has always been the possibility/risk for direct base-station-to-base-station and device-to-device interference, see Figure 24.6. In current commercial TDD-based cellular systems such interference is typically avoided by a combination of mutual time alignment between base stations and the use of the same downlink/uplink configuration in all cells. However, this requires a more or less static assignment of transmission resources preventing the resource assignment to adapt to dynamic traffic variations, thus removing one of the main benefits of TDD.

The eIMTA feature introduced in LTE release 12 (Chapter 15) is one step toward a more flexible assignment of TDD transmission resources to downlink and uplink. Nevertheless, a new 5G radio access should go even further, allowing for more or less fully dynamic assignment of transmission resources to the different transmission directions. However, this

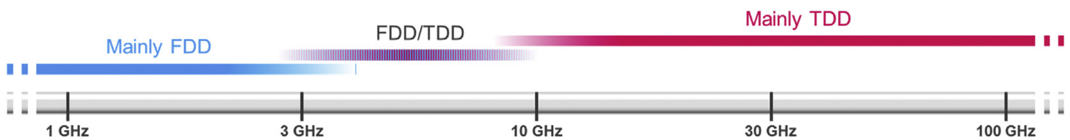

FIGURE 24.5

Typical duplex methods for different frequency bands.

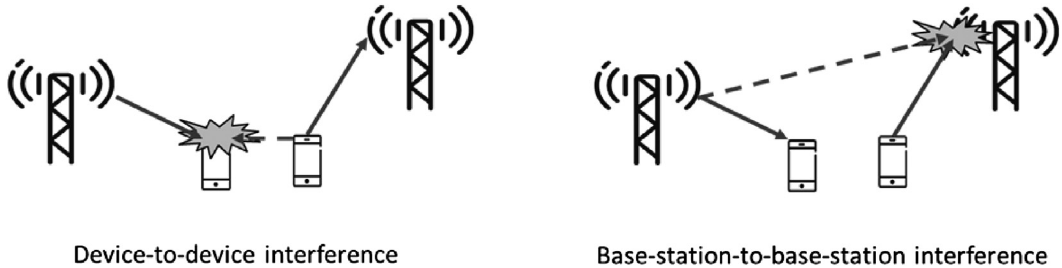

FIGURE 24.6

Direct base-station-to-base-station and device-to-device interference in case of TDD.

will then create a situation with potential base-station-to-base-station and device-to-device interference.

What makes base-station-to-base-station and device-to-device interference special and potentially much more severe, compared to the base-station-to-device and device-to-base-station interference that occur in any cellular system regardless of the duplex arrangement, is the difference in transmission characteristics between base stations and devices. This is especially the case for wide-area-covering ("macro") deployments where

- base stations have high transmit power, are located at elevated positions ("above roof top"), and are often transmitting with a high duty cycle serving many active devices;
- devices have much lower transmit power, are typically located indoor or outdoor on street level and are typically transmitting with, on average, a relatively low duty cycle.

However, in the future there will also be many very dense deployments, especially at higher frequencies. In such cases, the transmission characteristics of base stations and devices will be more similar:

- Compared to wide-area deployments, base stations in dense deployments will have a transmit power more similar to the transmit power of devices.
- Base stations in dense deployments are deployed indoor and outdoor on street level, that is, similar to device locations.
- Base stations in dense deployments will typically operate with, on average, lower duty cycle due to more dynamic traffic variations.

As a consequence, in dense deployments the base-station-to-base-station and device-to-device interference will be more similar to the base-station-to-device and device-to-base-station interference occurring regardless of the duplex arrangement, making dynamic assignment of TDD transmission resources a more viable option. The instantaneous traffic conditions, including the downlink vs. uplink traffic demands, will also vary more extensively in such deployments, making dynamic assignment of TDD transmission resources more beneficial.

It is important to understand that supporting fully dynamic TDD does not mean that the transmission resources should always be dynamically assigned. Especially when using the new 5G RAT for more wide-area deployments in unpaired spectrum on lower frequencies, the typical situation would be a synchronized deployment with the downlink/uplink configuration aligned between cells. The key thing is that the new 5G RAT *should allow for* fully dynamic assignment of transmission resources when operating in unpaired spectrum, leaving it to the network operator to decide what to use in a given deployment.

24.2.2.2 What About Full Duplex?

There have recently been different proposals for "true" full-duplex operation [70]. In this context, full-duplex operation means that transmission and reception is carried out *at the same frequency at the same time*.[3]

Full-duplex operation obviously leads to very strong "self" interference from the transmitter to the receiver, an interference that needs to be suppressed/cancelled before the actual target signal can be detected.

In principle, such interference suppression/cancellation is straightforward as the interfering signal is in principle completely known to the receiver. In practice, the suppression/cancellation is far from straightforward due to the enormous difference between the target signal and the interference in terms of received power. To handle this, current demonstrations of full-duplex operation rely on a combination of spatial separation (separate antennas for transmission and reception), analog suppression, and digital cancellation.

Even if full duplex would be feasible in real implementation, its benefits should not be overestimated. Full duplex has the potential to double the link throughput by allowing for continuous transmission in both directions on the same frequency. However, there will then be two simultaneous transmissions, implying increased interference to other transmissions, something which will negatively impact the overall system gain. The largest gain from full duplex can therefore be expected to occur in scenarios with relatively isolated radio links.

One scenario where full duplex is more likely to be beneficial is the wireless backhaul scenario—that is, wireless connectivity between base stations:

- Backhaul links are in many cases more isolated, compared to conventional base-station/device links.
- The extensive receiver complexity associated with full duplex may be more feasible to include in backhaul nodes, compared to conventional mobile devices.
- For backhaul nodes it is easier to envision more spatial separation between transmitting and receiving antennas, relaxing the requirements on active interference suppression.

When talking about full duplex one is typically assuming full duplex *on link level*—that is, simultaneous transmission in both directions on a bidirectional base-station/device link (left part of Figure 24.7). However, one can also envision full duplex on *cell level*, where a base

[3]Not to be up mixed with *full-duplex FDD* as used in LTE.

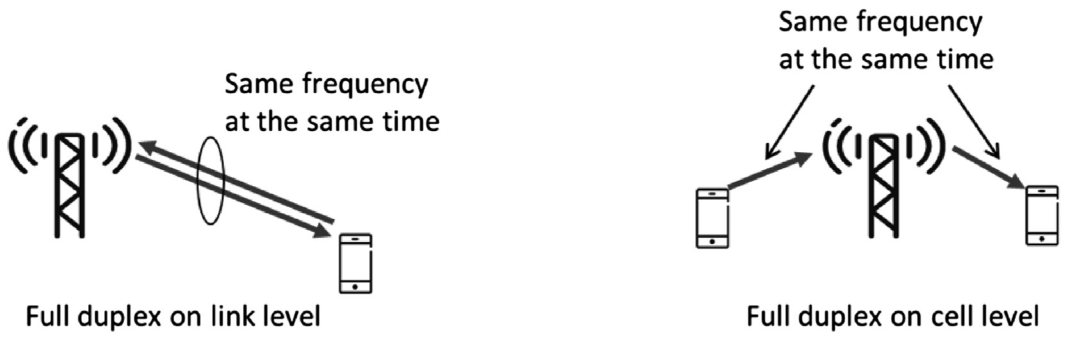

FIGURE 24.7

Full duplex on link level vs. cell level.

station transmits to one device and is simultaneously receiving/detecting the transmission *of another* device on the same frequency (right part of Figure 24.7). The benefit of full duplex on cell level, compared to full duplex on link level, is that simultaneous same-frequency transmission and reception does not need to be supported on the device side. Furthermore, one can typically achieve a higher degree of spatial separation between transmitting and receiving antennas on the base-station side, relaxing the requirements on active interference suppression.

24.2.3 FRAME STRUCTURE

The new 5G RAT should have a frame structure that supports operation in both paired and unpaired spectrum, and in licensed as well as unlicensed spectrum. The frame structure should also be applicable to device-to-device (sidelink) connectivity, see Section 24.2.9.2.

The frame structure is a key factor to enable low latency over the radio interface. To enable a low latency, there is, for example, need for a short TTI and, consequently a short subframe. As an example, to enable the required 1 ms end-to-end latency, a subframe in the order 200 µs or less is needed. Note that this is aligned with the higher numerologies of Table 24.1.

Low link-level latency also requires the possibility for fast demodulation and decoding of data. To enable this, control information needed by the receiver for demodulation and decoding of the data within a subframe should be located at the beginning of the subframe, similar to the PDCCH of LTE.[4] Reference signals for channel estimation should also be located early in the subframe. In this way demodulation and decoding can start as early as possible without having to wait until the entire subframe has been received. Figure 24.8 illustrates a relatively generic frame structure fulfilling the above requirements. In case of downlink data transmission (upper part of Figure 24.8), control information (scheduling

[4]Note that the control signaling should still follow the "stay-in-the-box" principle of Section 24.1.3 and be transmitted jointly with the data, in contrast to the PDCCH which is spread out over the entire frequency domain.

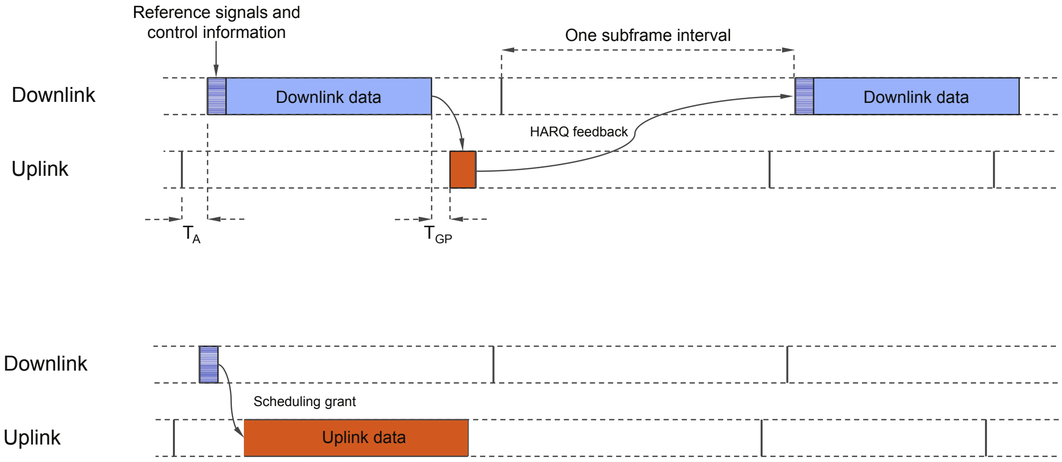

FIGURE 24.8

Generic 5G frame structure (TDD assumed in the figure).

assignments) and reference signals are located at the beginning of the downlink subframe, enabling early start of demodulation and decoding. In case of TDD operation, the downlink transmission ends prior to the end of the downlink subframe interval. Assuming decoding can be finalized during the guard period for the downlink-to-uplink switch, a hybrid-ARQ acknowledgment may then be transmitted on the uplink already in the last part of the downlink subframe interval, enabling very fast retransmissions. In the example of Figure 24.8, retransmissions occur with only one subframe delay.[5] This can be compared to LTE where there is a roughly three-subframe latency between the downlink transmission and the hybrid-ARQ acknowledgment and typically eight subframes between retransmissions. In combination with the shorter subframe, the outlined frame structure would thus enable substantially reduced hybrid-ARQ round-trip time compared to LTE.

Uplink data transmissions can be handled in a similar way; the scheduling grant is transmitted at the beginning of the subframe and the corresponding uplink data fills the remaining part of the uplink subframe as shown at the bottom of Figure 24.8. Thus, the scheduling decisions controls the "direction" of the data transmission, uplink or downlink, at a given point in time, resulting in a dynamic TDD scheme as discussed in Section 24.2.2.

Although 5G should enable very low latency in the order of 1 ms, many applications are not that latency critical. At the same time, the short transmissions needed to enable that kind of very low latency leads to inefficiencies in terms of, for example, control overhead. It may also lead to degraded channel estimates assuming that only reference signals within the set of subframes corresponding to the data transmission can be used for channel estimation

[5]Although one could in principle envision a base station decoding data as fast as the device, deployment aspects such as the use of remote radio units implies that retransmission earlier than shown in the figure cannot be generally assumed.

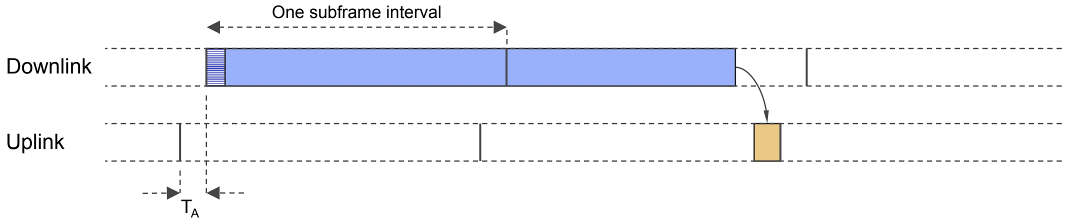

FIGURE 24.9

Subframe aggregation.

(self-contained transmissions as discussed in Section 24.1.3). To handle this, there should be a possibility to dynamically aggregate multiple subframes for a single transmission as illustrated in Figure 24.9

24.2.4 CHANNEL CODING

LTE uses Turbo coding for data transmission—see, for example, Chapter 6. Turbo coding is one candidate for channel coding also for the new 5G RAT. However, there are also other channel-coding approaches being considered, including low-density parity check (LDPC) codes [72] and Polar codes [73].

LDPC codes have been around for a relatively long time. They are block codes based on sparse ("low-density") parity matrixes with decoding based on iterative message-passing algorithms. In contrast, Polar codes have emerged much more recently. Their main fame comes from the fact that they are the first known structured codes that reach the Shannon bound.

In terms of basic performance, that is, the required E_b/N_0 needed for a certain error rate, the difference in performance between the different coding schemes is relatively small for medium to large sizes of the block of bits to be decoded (in the order of 1000 bits and beyond). On the other hand, for smaller block lengths, Polar codes currently seem to have a slight advantage. Polar codes also benefit from not having an error floor, something that makes it easier to achieve a very low error rate after decoding. In contrast, Turbo codes and, in many cases, also LDPC codes have such error floor. This may speak in favor of Polar codes for high-reliability applications with small payloads but requiring extremely good error performance.

The main issue with Polar codes is their relative immaturity which also implies that the implementation experience of Polar codes is much more limited compared to both Turbo codes and LDPC codes. The latter have, as already mentioned, been around for many years and are already part of widespread wireless technologies. The immaturity and lack of implementation experience of Polar codes implies that there may be practical issues not yet discovered. It also implies that it is more difficult today to fully understand the implementation complexity of Polar codes.

The main benefit of Turbo codes is flexibility in terms of block length and code rate. As an example, LTE uses a single basic Turbo code of code rate 1/3 regardless of the block length.

Lower and higher codes rates are then straightforwardly achieved by means of rate matching (puncturing or repetition) with performance degrading only gracefully with the amount of puncturing. The flexibility in terms of puncturing of Turbo codes also makes is straightforward to implement hybrid ARQ based on incremental redundancy.

On the other hand, an LDPC code is designed for a certain block length and a certain code rate. To use a different block length for a given LDPC code, zero padding of the information block is needed, inherently leading to a lower code rate. To increase the code rate, puncturing can be applied. However, any substantial amount of puncturing may lead to significant degradation of the decoder performance. Thus, in order to support multiple block lengths and multiple code rates one may have to define multiple LDPC codes with different parity matrices, something which will increase the complexity and storage requirements.

The main drawback of Turbo codes versus LDPC codes and, as it looks today, also versus Polar codes, is in terms of decoder complexity. Simply speaking, at least LDPC codes allow for substantially higher throughout per chip area and lower energy consumption per bit compared to Turbo codes. LDPC codes also allow for a higher degree of parallelism of the decoder, enabling lower latency in the decoding process. Note that the possibility for fast decoding is one component to enable low latency over the radio interface.

24.2.5 MULTI-ANTENNA TRANSMISSION AND BEAM-FORMING

Multi-antenna transmission has been a key part of LTE already since its first release. As has been extensively described in previous chapters, there has also been a continuous evolution of LTE introducing support for new transmission schemes exploiting multi-antenna configurations at the network side. Spatial multiplexing (SU-MIMO) and SDMA (MU-MIMO) have been the main focus for LTE, targeting increased end-user data rates and system throughput within a limited transmission bandwidth and amount of spectrum.

SU/MU-MIMO will remain important technology components also for the new 5G RAT. However, for higher frequencies, the limiting factor will typically not be bandwidth and spectrum but coverage. As a consequence, especially for higher frequencies, beam-forming as a tool to provide enhanced coverage will be extremely important and even a necessity in many cases (Figure 24.10).

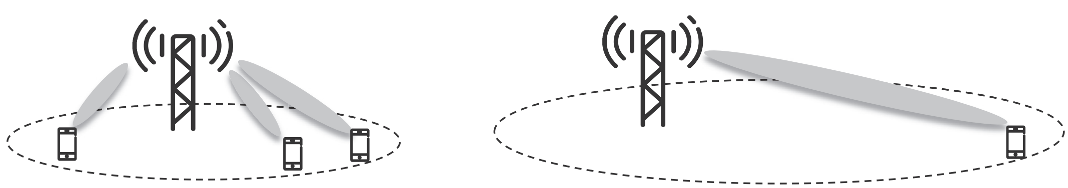

FIGURE 24.10

Beam-forming for capacity (left) and for coverage (right).

The extensive use of beam-forming to ensure sufficient coverage will have fundamental important on the new 5G RAT. One example is broadcast channels—for example, to deliver system information. If the link budget is challenging, requiring extensive use of beam-forming to provide coverage, broadcast channels may not work and other solutions have to be considered as discussed in Section 24.2.7 below. Beam finding and beam tracking are other challenges that need to be addressed.

Recent developments in implementation with tight integration of antenna elements and RF components such as power amplifiers and transceivers allow for a significantly larger number of controllable antenna elements than previously used, as discussed in Chapter 10. The use of a massive number of antenna elements enables extensive use of the spatial domain. *Massive MIMO* is a term commonly used in this context. Strictly speaking, this term only means the usage of a large number of antenna elements although it is often used in a more narrow meaning, namely exploiting channel reciprocity together with a large number of transmission antennas for simultaneous transmission to multiple receiving devices, in essence reciprocity-based multi-user MIMO. A key assumption in this more narrow interpretation of massive MIMO is the availability of the instantaneous channel impulse response as the base station, knowledge that can be obtained by exploiting channel reciprocity. Theoretical results indicate that the effective channel from the base station to each device is nonfrequency selective and exhibits no fast channel variations when the number of antenna elements goes to infinity. This would in theory allow for very simple scheduling strategies and very high capacity with simple receivers. However, in practice the channel knowledge is not perfect and the number of transmission antennas finite, thus frequency-domain scheduling is still relevant. With realistic traffic behavior instead of a full-buffer scenario, there is sometimes only one or a few devices to transmit to, meaning that single-user MIMO and spatial multiplexing are important complements to massive multi-user MIMO. To follow the traffic variations dynamic switching between SU-MIMO and MU-MIMO is therefore essential.

In the discussion above so-called *digital beam-forming* was assumed. In essence, each antenna element has its own DA converter and power amplifier, and all beam processing is done in baseband. Clearly, this allows for the largest flexibility and is in principle the preferred scheme with no limitations on the number of simultaneously formed beams. However, despite technology advances in integration and DA converters, such implementations may not be feasible in the near to mid-term perspective for several reasons. One aspect is the large power consumption and the challenges of cooling a large number of tightly integrated DA converters and RF chains. Analog or hybrid beam-forming schemes are therefore of interest from a practical perspective but implies that only one, or in the best case a few, simultaneous beams can be formed. Not only does this impact the possibilities for massive MIMO as discussed in the preceding paragraphs, it may also impact areas such as control signaling. For example, if only a single beam can be formed at a time, receiving a single-bit hybrid-ARQ acknowledgment will be very costly. Furthermore, frequency multiplexing multiple low-rate signals from different non-colocated devices is not possible as only a single beam can be formed at a time.

To summarize, a whole range of multi-antenna schemes are required to allow for efficient support not only of different traffic and deployment scenarios, but also to support a wide range of implementation alternatives.

24.2.6 MULTI-SITE CONNECTIVITY AND TIGHT INTERWORKING

Multi-site connectivity implies that a device has simultaneous connectivity to multiple sites. Fundamentally, this is nothing new. Soft handover in WCDMA/HSPA is one example of multi-site connectivity. LTE joint transmission CoMP (Chapter 13) and dual connectivity (Chapter 16) are other examples. However, in the 5G era, multi-site connectivity is expected to play a larger role, in particular when operating at very high carrier frequencies or when very high reliability is required,

At high carrier frequencies, the propagation conditions are different than at lower frequencies as already discussed. The diffraction losses are higher and the possibilities for radio waves to penetrate an object are lower. Without multi-site connectivity, this could result in the connection being (temporarily) lost when, for example, a large bus or a truck is passing between the device and the base station. To mitigate this, diversity through multi-site connectivity is beneficial. The likelihood of links to multiple antenna sites being in poor shape is much lower than the likelihood of a single link being in inferior.

Ultra-reliable low-latency communication, URLLC, is another example where multi-site connectivity is beneficial. To obtain the very low error probabilities discussed in Chapter 23, all types of diversity need to be considered, including site diversity through multi-connectivity.

Multi-connectivity can also be used to increase the user data rates at low loads. By transmitting from two or more antenna sites simultaneously, the effective rank of the channel toward the device can be increased and a larger number of layers being transmitted compared to the single-site scenario, see Figure 24.11. This is sometimes denoted *distributed MIMO*.

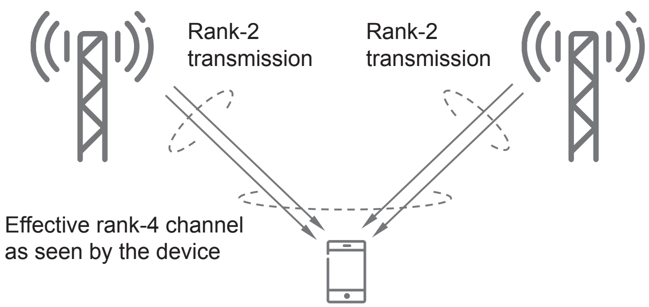

FIGURE 24.11

Multi-site transmission as a way to increase the effective channel rank.

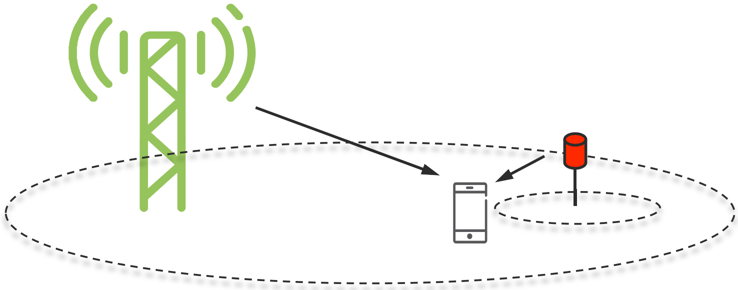

FIGURE 24.12

Multi-site connectivity between LTE and the new 5G RAT.

In essence, the transmission resources at the neighboring site otherwise momentarily being left unused due to lack of traffic can be used to increase the data rates experienced by the receiving device.

Multi-site connectivity could include sites within the same layer (intra-layer connectivity). However, a device could also have simultaneous connectivity to sites of different cell layers (inter-layer connectivity), see Figure 24.12. Especially in the latter case, the multi-site connectivity could include connectivity via different radio-access technologies (multi-RAT connectivity). This is closely related to the tight interworking between LTE and the new 5G radio access discussed in Chapter 23. One scenario where this makes sense is LTE on a low-frequency band providing ubiquitous and reliable access, complemented by the new 5G RAT in a higher-frequency band for providing very high data rates and high capacity in hot spots. Since the propagation conditions on high frequencies can be less predictable than at low frequencies, a tight connection between the two is required to provide a consistent user experience.

Multi-site connectivity can be realized at different levels in the protocol stack. Carrier-aggregation-like structures can be envisioned, although they typically have tight latency requirements on the backhaul between the sites. Another approach is to aggregate the data streams higher up in the protocol stack, similar to dual connectivity for LTE as discussed in Chapter 16 where aggregation is done at the PDCP layer. If a common PDCP layer is used for LTE and the new 5G radio access, tight interworking is straightforward. Related to this are also discussions on the interface between the radio-access network and the core network. This interface does not have to be identical to the existing S1 interface, but could be either a new interface or an evolution of the S1 interface.

24.2.7 SYSTEM-ACCESS FUNCTIONALITY

A very important part of a radio-access technology is the system-access functionality—that is, functionality not directly related to user-data delivery but necessary for devices to be able to even access the system. The system-access functionality includes the means by which devices acquire information about the configuration of the system, the means by which the network notify/page devices, and the functionality by which devices access the system normally referred to as *random access*.

As described in Chapter 11, LTE system information is broadcast on a per cell layer. This is a reasonable approach for large cells with wide antenna beams and a relatively high number of users per cell. However, there are some key 5G aspects and characteristics that will impact the system access functionality of the new 5G RAT and partly call for new types of solutions compared to LTE:

- the reliance of beam-forming for coverage, especially at higher frequencies;
- the need to support very high network energy efficiency in some deployments;
- the aim to support a high degree of forward compatibility.

Relying on beam-forming for coverage implies that the possibilities for broadcasting a lot of information are limited from a link-budget perspective. Minimizing the amount of broadcast information is therefore crucial to enable the use of massive beam-forming for coverage. It is also well in line with the ultra-lean design principle and the aim to minimize the amount of "always-on" transmission as outlined in Section 24.1.2. Instead it should be possible to provide a major part of the system information on a per-need basis, including dedicated signaling to a specific device, see Figure 24.13. This will allow for using the full battery of beam-forming capabilities also for the system-information delivery. Note that this does not mean that the system information should *always* be provided in this way but it should be *possible* when motivated by the scenario. In some scenarios, especially in case of a large number of devices in a cell, it is undoubtedly more efficient to broadcast the system information. The key point is that the new 5G RAT should have the flexibility to deliver system information by different means, including broadcast over the entire coverage area, as joint transmission to a set of devices and by means of dedicated signaling on a device-by-device bases, depending on the scenario.

There should also be flexibility in terms of from what nodes system information is broadcasted. As an example, in a multi-layer network with a large number of low-power

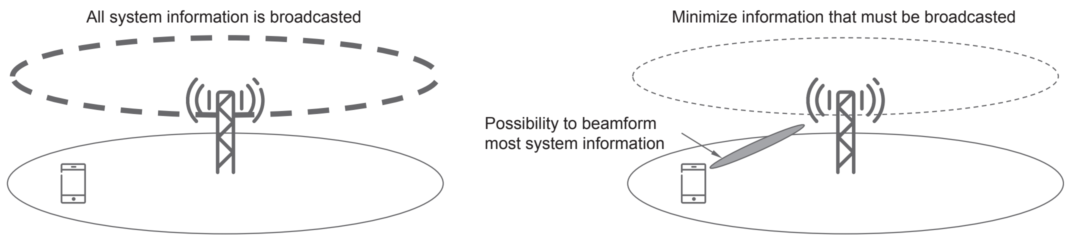

FIGURE 24.13

Broadcasted vs. dedicated transmission of system information.

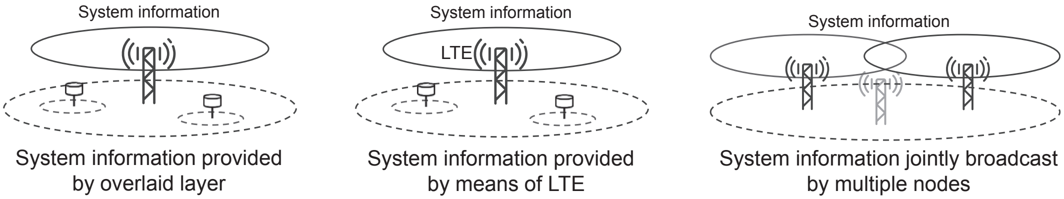

FIGURE 24.14

Broadcasting (parts of) system information.

nodes deployed under the coverage of an overlaid macro layer, system information may only be broadcast from the overlaid layer (left part of Figure 24.14). This would mean that nodes in the underlaid layer may, from a transmission point of view, be completely inactive when there is no device to be served. Once again, this is very well aligned with the ultra-lean design principle. At least in an early-phase deployment of the new 5G RAT, a common situation would be a layer of lower-power nodes based on the new 5G radio access, with an overlaid LTE-based macro layer. In other words, the multi-layer deployment would also be a multi-technology deployment. In that case, the system information relevant for the underlaid low-power layer could be provided via the *LTE-bas*ed overlaid layer as illustrated in the middle part of Figure 24.14.

Taking into account that a large part of the system information is actually identical between neighbor cells, system information can also be broadcasted jointly from a set of nodes using MBSFN transmission (right part of Figure 24.14). This would improve the coverage of the system information broadcast making it possible to deliver it using less resources, both time—frequency resources and power. Note that only a subset of the nodes may be involved in the MBSFN transmission.

Random-access procedures may also be impacted by excessive use of beam-forming as a way to provide coverage. Depending on the beam-forming solution used on the network side it may not be possible to listen for random-access transmissions in all directions simultaneously, which, for example, can be addressed through different types of beam-sweeping solutions. Furthermore, if the (small amount of) broadcasted system information is identical across multiple sites and delivered using MBSFN transmission, the device cannot use the LTE approach of targeting a certain cell with the random-access attempts. However, the device is actually not interested in contacting a specific cell, it is interested in establishing a connection with the network, a network which may consist of nodes currently not transmitting, and therefore being silent. Furthermore, a node having the best downlink toward a device may not necessarily be the best node to receive the random-access transmission. This could be reasons to consider a random-access scheme different from LTE—for example, where the device transmits a random-access request without targeting a particular node and being capable of handling a situation where multiple nodes may respond to the request.

24.2.8 SCHEDULED AND CONTENTION-BASED TRANSMISSIONS

In LTE, all uplink data transmissions are scheduled. Scheduled uplink transmission will most likely remain the normal case also for the new 5G RAT. Scheduling provides dynamic and tight control of transmission activities, resulting in efficient resource utilization. However, scheduling requires the devices to request resources from the base station, which, after taking a scheduling decision, may provide the device with a scheduling grant indicating the resource to use for the uplink transmission, see Figure 24.15. This request-grant procedure, described for LTE in Chapter 9, will add to the overall latency. Therefore, for latency-critical services with small and infrequent uplink transmissions, a possibility for immediate transmission, without a preceding request-grant phase, is of interest.

One way of avoiding the latency associated with the request-grant phase is to provide a device with a scheduling grant in advance valid for a certain time. During the time during which the scheduling grant is valid, a device can transmit on the uplink without having to go through the request-grant phase, see Figure 24.16. In order to retain efficiency, the grant of Figure 24.16 would typically not provide exclusive access to the uplink. Rather, multiple devices would typically get grants covering the same set of resource with the network handling any collisions that might occur.

Another way to avoid the latency associated with the request-grant phase is to allow for unscheduled transmissions not requiring any grant, see Figure 24.17. It should be noted that

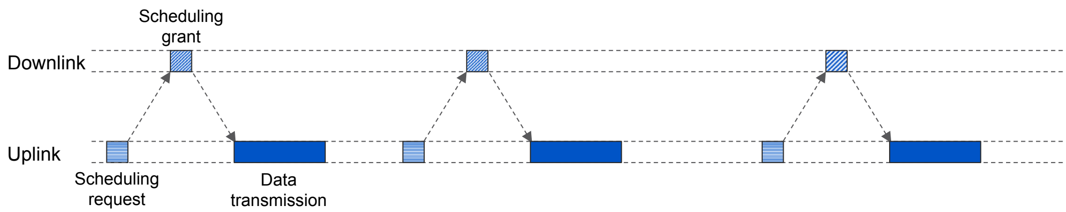

FIGURE 24.15

Scheduled uplink transmission.

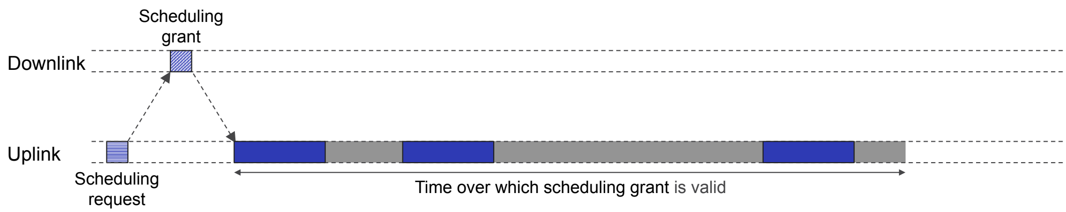

FIGURE 24.16

“Pre-scheduled” uplink transmissions.

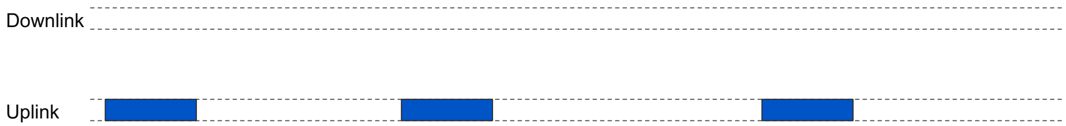

FIGURE 24.17

Unscheduled uplink transmission.

under the assumption that the scheduling grant of Figure 24.16 does not provide exclusive access, there are many similarities between unscheduled uplink transmission and the scheduled transmission of Figure 24.16. In principle, unscheduled transmission can be seen as a special case of the scheduled transmission of Figure 24.16 where the scheduling grant is implicitly provided when the connection is established and is then valid for the duration of the connection.

Without a grant providing exclusive access, collisions between transmissions from different devices cannot be avoided. This can be handled in different ways. One way is to accept the collision and assume that the colliding devices will retransmit at a later stage, hopefully not colliding again. One could also have a situation where the processing gain is sufficient to actually allow for detection and decoding of the colliding transmissions. Note that this is essentially what is done for the WCDMA uplink where "colliding" transmissions is the normal case.

There are also proposals for more specific transmission structures allowing for more efficient detection of colliding signals. On example is *low-density spreading* (LDS) [71]. LDS spreads the transmitted signal with special spreading sequences for which only a small fraction of the sequence elements are nonzero and where the set of nonzero elements differ between different users. With such spreading sequences, two colliding transmissions will only partly collide, enabling more efficient and low-complex multi-user detection based on message-passing algorithms in a way similar to the decoding of LDPC codes. *Sparse-code multiple access* (SCMA) [76] is a modification/extension of LDS where the direct sequence spreading of LDS is replaced by sparse code words providing extended Euclidean distance and enhanced link performance.

24.2.9 NEW TYPES OF WIRELESS LINKS

Traditionally, mobile communication has solely been about wireless links between base stations and mobile devices. This has partly been changed already with LTE, with the introduction of relaying in release 10 (Chapter 18) and device-to-device connectivity in release 12 (Chapter 21).

In the 5G era this will be even more pronounced and support for both base-station-to-base-station communication ("wireless backhaul") and device-to-device connectivity are expected to be integrated parts of the new 5G RAT.

24.2.9.1 Access/Backhaul Convergence

The use of wireless technology for backhaul has been used extensively for many years. Actually, in some regions of the world, wireless backhaul constitutes more than 50% of total backhaul. Current wireless-backhaul solutions are typically based on proprietary (non-standardized) technology operating as point-to-point line-of-sight links using special frequency bands above 10 GHz. The wireless backhaul is thus using different technology and operating in different spectrum, compared to the access (base-station/device) link. Relaying, introduced in release 10 of LTE, is basically a wireless-backhaul link although with some restrictions. However, it has so far not been used in practice to any significant extent. One reason is that small-cell deployments, for which relaying was designed, has not yet taken off in practice. Another reason is that operators prefer to use the precious low-frequency spectrum for the access link. Wireless backhauling, if used, relies on non-LTE technologies capable of exploiting significantly higher-frequency bands than LTE, thereby avoiding wasting valuable access spectrum for backhaul purposes.

However, in the 5G era, a convergence of backhaul and access can be expected for several reasons:

- In the 5G era, the access link will expand into higher-frequency bands above 10 GHz—that is, the same frequency range that is currently used for wireless backhaul.
- In the 5G era, the expected densification of the mobile networks, with a large number of base stations located indoor and outdoor on street level, will require wireless backhaul capable of operating under non-line-of-sight conditions and, more generally, very similar propagation conditions as the access link.

The requirements and characteristics of the wireless-backhaul link and the access link are thus converging. In essence, with reference to Figure 24.18, there is, radio-wise, no major difference between the wireless backhaul link and the normal wireless link. As a consequence, there are strong reasons to consider a convergence also in terms of technology and spectrum. Rather, there should be a single new 5G RAT that can be used for both access and wireless backhaul. There should preferably also be common spectrum pool for both the access link and the wireless backhaul.

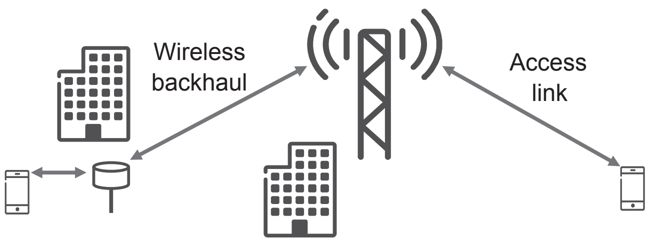

FIGURE 24.18

Wireless backhaul vs. the access link.

A consequence of this is that the wireless-backhaul use case needs to be taken into account in the design of the new 5G RAT. Despite the claim mentioned earlier that the requirements and characteristics are converging, there are still certain important attributes of the wireless-backhaul link that needs to be taken into account in order for the wireless-backhaul use case to be efficiently supported:

- In the wireless-backhaul scenario, the "device" is typically not mobile.
- Compared to normal devices, the wireless-backhaul "device" can have more complexity, including more antennas and possibility for separation of transmit and receive antennas.

It should be noted that these characteristics are not necessarily unique for the wireless backhaul use case. As an example, one may envision MTC applications where the device is not mobile. At the same time, there may be wireless-backhaul scenarios where the backhauled node is mobile, for example, in case of wireless backhaul to base stations on trains and buses.

It should also be noted that a common spectrum pool for access and wireless backhaul does not necessarily mean that the access link and the wireless-backhaul link should operate on the same frequency ("inband relaying"). In some cases, this will be possible. However, in many other cases, having a frequency separation between the backhaul link and the access link is preferred. The key point is that the separation of spectrum between backhaul and access should, as much as possible, not be a regulatory issue. Rather, an operator should have access to a single spectrum pool. It is then an operator decision how to use this spectrum in the best possible way, taking into account both need for both access and backhaul.

24.2.9.2 Integrated Device-to-Device Connectivity

As described in Chapter 21, support for direct device-to-device connectivity, or sidelink connectivity, using LTE was introduced in 3GPP release 12, with further extensions introduced in release 13. As also described, LTE device-to-device connectivity consists of two parts:

- device-to-device communication, focusing on the public-safety use case;
- device-to-device discovery, targeting not only public safety but also commercial use cases.

Device-to-device connectivity should be an even more integrated part of the new 5G RAT. This should be possible taken into account that, in this case, the support for device-to-device connectivity can and should be taken into account already in the initial design of the RAT. In contrast, LTE device-to-device connectivity was introduced into an already existing RAT that was initially designed without any considerations of a later introduction of device-to-device connectivity. By taking device-to-device connectivity into account already in the initial design, it would be possible to have much more of a common framework for downlink, uplink, and sidelink (Figure 24.19).

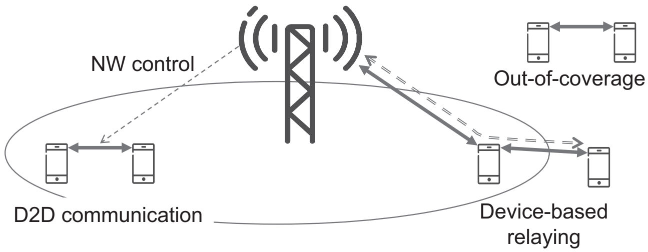

FIGURE 24.19

Device-to-device connectivity.

Device-to-device connectivity for the new 5G RAT should not be seen as a tool only targeting specific use cases such as public safety. Rather, the device-to-device connectivity should be seen as a general tool to enhance connectivity within the 5G network. In essence, direct data transfer between devices should be configured if the network concludes that this is more efficient (requires less resources) or provides better quality (higher data rates and/or lower latency) compared to indirect connectivity via the infrastructure. The network should also be able to configure device-based relay links to enhance the connectivity quality.

In order to maximize efficiency device-to-device connectivity should, as much as possible, take place under network control. However, similar to LTE, device-to-device connectivity should also be possible when no network is available (i.e., out of coverage).