CHAPTER

NR PHYSICAL LAYER: OVERVIEW 2

Like for any wireless technology, the physical layer forms the backbone of 5G NR. The NR physical layer has to support a wide range of frequencies (from sub-1 GHz to 100 GHz) and various deployment options (pico cells, micro cells, macro cells). There are human-centric and machine-centric use cases with extreme and sometimes contradictory requirements. There may also be unforeseen applications with new requirements in the future. To successfully address these challenges, 3GPP is developing a flexible physical layer for NR. The flexible components can be properly optimized with an accurate understanding of radio wave propagation and hardware imperfections in networks and devices. This is a challenge, because these characteristics are less understood. NR is the first ever mobile radio access technology going into millimeter-wave frequency range (with frequencies as high as 100 GHz), targeting channel bandwidths in the GHz range, and enabling massive multiantenna systems.

The first release of NR (3GPP NR Release 15) was completed in June 2018. Any future releases of NR will be backwards compatible to its first release. This is usually referred to as forward compatibility of NR, that is, NR will be developed in such a manner that any future releases of NR will be backwards compatible to its initial release(s). For the reader interested in understanding what an NR physical layer is, in this chapter we provide an overview of the NR physical layer (based on the first NR release) and discuss the radio wave propagation and hardware impairment related challenges associated with enabling the NR physical layer. Our focus in the following chapters will be on fundamental principles, models, and technology components for the NR physical layer. If the reader is interested in the details of the NR specification, we strongly recommend [6].

The chapter is organized as follows. In Section 2.1, we briefly describe the protocol stack of the NR radio interface and the role of the physical layer therein. Due to brevity, this section may be hard to fully grasp if the reader is not familiar with 3GPP-based cellular technologies (e.g., 4G LTE [5]). The rest of the chapter focuses only on physical layer aspects. Section 2.2 gives an overview of the key physical layer technology components of NR—modulation, waveform, multiantenna, and channel coding schemes. These technology components are explained in detail in Chapters 5–8. Section 2.2 introduces the physical time-frequency resource structure of NR. Sections 2.4 and 2.5 describe how the time-frequency resources are allocated to different types of signals. Sections 2.6 and 2.7 explain flexible duplexing schemes and a flexible transmission structure (frame structure) available in NR. Finally, Section 2.9 briefly summarizes some challenges associated with radio wave propagation and hardware impairments. This section motivates Chapters 3 and 4, which, respectively, cover radio wave propagation and hardware impairments in depth.

2.1 RADIO PROTOCOL ARCHITECTURE

In 3GPP terminology, a base station is an implementation of a logical radio access network node. For example, in 3G UMTS and 4G LTE, the network node is termed Node B (NB) and evolved Node B

5G Physical Layer. https://doi.org/10.1016/B978-0-12-814578-4.00007-2

(eNB), respectively. A 5G NR radio access network node has been named the next generation Node B (gNB) by 3GPP. It is important to stress that gNB is a logical entity and not a physical implementation of a base station. A base station can be realized in different ways based on a standardized gNB protocol. Similarly, any device is referred to as a UE in 3GPP specifications.

The radio protocol architecture for NR can be separated into control-plane architecture and user-plane architecture. The user-plane delivers user data, whereas the control-plane is mainly responsible for connection setup, mobility, and security. Fig. 2.1 illustrates the user-plane protocol stack of NR. The protocol is split into the following layers: physical (PHY) layer, medium access control (MAC) layer, radio link control (RLC) layer, packet data convergence protocol (PDCP) layer, and service data adaptation protocol (SDAP) layer. The main functionalities of these layers are briefly described now.

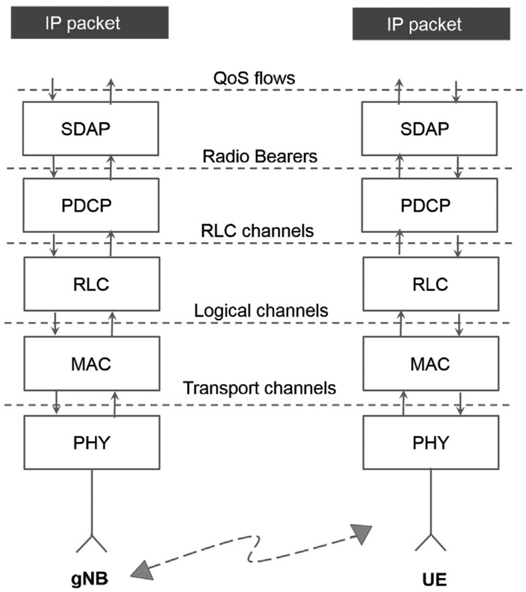

FIGURE 2.1

NR user-plane protocol stack.

- The SDAP layer handles the mapping between quality of service (QoS) flow and radio bearers.[1] IP packets are mapped to radio bearers according to their QoS requirements.
- The PDCP layer is primarily responsible for IP header compression/decompression, reordering and duplicate detection, ciphering/deciphering and integrity protection. The header compression mechanism reduces the number of bits to transmit over radio interface. The ciphering protects from

[1] A radio bearer can be seen as a pipe that carries internet protocol (IP) packets through a network and gets prioritized according to the specified QoS requirement.

eavesdropping and ensures message integrity. The reordering and duplication detection mechanisms allow in-sequence delivery of data units and removes duplicate data units.

- The RLC layer mainly performs error correction through an automatic repeat request[2] (ARQ) mechanism, segmentation/resegmentation of (header compressed) IP packets, and in-sequence delivery of data units to higher layers.
- The MAC layer is mainly responsible for error correction through hybrid ARQ[3] (HARQ) mechanism and uplink and downlink scheduling. The scheduler controls the assignment of uplink and downlink physical time-frequency resources for transmission. The MAC layer also takes care of multiplexing data across multiple component carriers when carrier aggregation is employed.
- The PHY layer handles coding/decoding, modulation/demodulation, multiantenna processing, and mapping of signals to physical time-frequency resources.

The control-plane is mainly responsible for control signaling for connection setup, mobility, and security. Control signaling originates either from a core network or from a radio resource control (RRC) layer in gNB. The main services from a RRC layer include broadcasting of system information, transmission of paging messages, security management including key management, handovers, cell-selection/reselection, QoS management, and detection of and recovery from radio link failures. The RRC messages are transmitted using the same PDCP, RLC, MAC, and PHY layers as for the user-plane. Therefore, from a physical layer perspective, there is not at all a fundamental technological difference in providing services to higher layers in control-plane and user-plane protocol stacks.

2.2 NR PHY: KEY TECHNOLOGY COMPONENTS

The key technology components of the NR physical layer are modulation, waveform, multiantenna transmission, and channel coding. In the following, we provide a brief overview of these physical layer components.

2.2.1 MODULATION

NR supports quadrature phase shift keying (QPSK), 16 quadrature amplitude modulation (QAM), 64 QAM and 256 QAM modulation formats for both uplink and downlink, as in LTE. Moreover, $\pi/2$-BPSK is supported in uplink to enable a further reduced peak-to-average power ratio and enhanced power amplifier efficiency at lower data rates, which is important for mMTC services. Since NR will cover a wide range of use cases, it is likely that the set of supported modulation schemes may be expanded. For example, 1024 QAM may become part of the NR specification, since fixed point-to-point backhaul already uses modulation orders higher than 256 QAM. Different modulation schemes for different UE categories may also be included in the NR specification.

[2] Automatic repeat request (ARQ) is an error-control method for data transmission that uses acknowledgments and timeouts to achieve reliable data transmission.

[3] Hybrid ARQ is a combination of forward error-correction coding and ARQ error-control.

Table 2.1 Scalable OFDM numerology for 5G NR (3GPP Release 15)

OFDM numerology	15 kHz	30 kHz	60 kHz	120 kHz
Frequency band	0.45–6 GHz	0.45–6 GHz	0.45–6 GHz 24–52.6 GHz	24–52.6 GHz
OFDM symbol duration	66.67 µs	33.33 µs	16.67 µs	8.33 µs
Cyclic prefix duration	4.69 µs	2.34 µs	1.17 µs	0.59 µs
OFDM symbol with CP	71.35 µs	35.68 µs	17.84 µs	8.91 µs
Maximum bandwidth	50 MHz	100 MHz	200 MHz	400 MHz

2.2.2 WAVEFORM

NR employs cyclic prefix OFDM (CP-OFDM) in both uplink and downlink up to at least 52.6 GHz. This is in contrast to LTE, where CP-OFDM is only used for downlink transmissions and DFT-Spread OFDM (DFTS-OFDM) is used for uplink transmissions. Having the same waveform in both directions simplifies the overall design, especially with respect to wireless backhauling and device-to-device (D2D) communications. Additionally, there is an option for DFT-spread OFDM (DFTS-OFDM) in uplink for coverage-limited scenarios, with single stream transmissions (that is, without spatial multiplexing). In practice, a gNB can select the uplink waveform (either CP-OFDM or DFTS-OFDM) and a UE should be able to support both OFDM and DFTS-OFDM. Any operation that is transparent to a receiver can be applied on top of the NR waveform, such as windowing/filtering to improve spectrum confinement.

NR has a scalable OFDM numerology to enable diverse services on a wide range of frequencies and deployments. The subcarrier spacing is scalable and specified as 15×2^n kHz, where n is an integer and 15 kHz is the subcarrier spacing used in LTE. In Chapter 6, we provide details on numerology design for NR. In 3GPP Release 15, four subcarrier spacings are specified: 15 kHz, 30 kHz, 60 kHz, and 120 kHz (i.e., $n = 1/2/3/4$), all with 7% CP overhead as in LTE (see Table 2.1). For the 60 kHz numerology, an extended CP is also defined. Different numerologies are specified for different frequency bands. Currently, there is no spectrum identified for NR between 6 and 24 GHz; therefore, the corresponding numerology has not been specified. When new frequencies will become available, the corresponding numerologies can be specified according to 15×2^n kHz. For all numerologies, the number of active subcarriers is 3300. Considering 3300 active subcarriers, the maximum bandwidths enabled by different numerologies are given in Table 2.1. To support even larger channel bandwidths, carrier aggregation can be employed. In Release 15, up to 16 component carriers are supported, where each component carrier can have up to 3300 active subcarriers.

The spectrum of an OFDM signal decays rather slowly outside the transmission bandwidth. In order to limit out-of-band emission, the spectrum utilization for LTE is 90 percent. For NR, it has been agreed that the spectrum utilization will be 94 to 99 percent. Windowing and filtering operations are viable ways to confine the OFDM signal in the frequency domain. In Chapters 5 and 6, we provide details on the spectrum confinement techniques for multicarrier waveforms.

2.2.3 MULTIPLE ANTENNAS

Multiantenna techniques were important already in LTE, but in NR they have a more fundamental role in the system design. The extension of the spectrum for mobile communication to include also the millimeter-wave bands has led to a beam-centric design of NR in order to support analog beam-forming for achieving sufficient coverage. Furthermore, multiantenna techniques are crucial for fulfilling the performance requirements for 5G also for the traditional cellular frequency bands.

For low frequencies, multiantenna techniques are mainly enhancements of features developed in the later releases of LTE. The purpose with these enhancements is improved spectral efficiency driven by the ever-increasing quest for higher data rates and capacity in a congested spectrum. Advances in active array antenna technology have made it possible to have digital control over a large number of antenna elements, sometimes referred to as massive multiple-input multiple-output (MIMO). This enables higher spatial resolution in the multiantenna processing which can give higher spectral efficiencies. To this end, NR provides better support for multiuser MIMO (MU-MIMO) and reciprocity-based operation. A new framework for acquiring channel state information (CSI) has been developed to allow for more flexibility in the transmission of reference signals and to enable CSI with higher spatial resolution. This framework also provides a leaner system design and makes it easier to adapt to diverse use cases and to introduce new features in future releases of NR.

For high frequencies, obtaining coverage is the main challenge rather than obtaining high spectral efficiency. The reason for this is that transmission losses, when using legacy transmission techniques, are considerably higher, while there is a large amount of bandwidth available in the millimeter-wave spectrum. To overcome higher transmission losses and provide sufficient coverage, beam-forming is useful, particularly under LoS conditions and possibly both at the gNB and UE. With current hardware technology, analog beam-forming is expected to be prevalent at millimeter-wave frequencies. Therefore, procedures for supporting analog beam-forming in both the gNB and the UE have been developed in NR. Unlike previous generations of mobile communication systems, NR supports beam-forming not only for the data transmission but also for initial access and broadcast signals.

NR multiantenna techniques for both the gNB and the UE are discussed in detail in Chapter 7.

2.2.4 CHANNEL CODING

NR employs low density parity check (LDPC) codes for the data transmission for mobile broadband (MBB) services and polar codes for the control signaling. LDPC codes are attractive from an implementation perspective, especially at multigigabits-per-second data rates. Unlike the LDPC codes implemented in other wireless technologies, the LDPC codes considered for NR use a rate-compatible structure. This allows for transmission at different code rates and for HARQ operation using an incremental redundancy.

For the physical layer control signaling where the information blocks are small compared to data transmission and HARQ is not used, NR employs polar codes. By concatenating the polar code with an outer code and by performing successive cancellation list decoding, good performance is achieved at shorter block lengths. For the smallest control payloads, Reed–Muller codes are used.

NR LDPC and polar codes are discussed in detail in Chapter 8. For URLLC services, channel codes have not been agreed in 3GPP yet. In Chapter 8, a selection of coding schemes that exhibit a favorable performance/complexity tradeoff in the short block length regime is discussed. These schemes may be candidates for the future NR releases in addition to LDPC and polar codes.

2.3 PHYSICAL TIME-FREQUENCY RESOURCES

Physical time-frequency resources correspond to OFDM symbols and subcarriers within the OFDM symbols. The smallest physical time-frequency resource consists of one subcarrier in one OFDM symbol, known as a resource element. The transmissions are scheduled in group(s) of 12 subcarriers, known as physical resource blocks (PRBs). An example of NR physical time-resource structure is shown in Fig. 2.2.

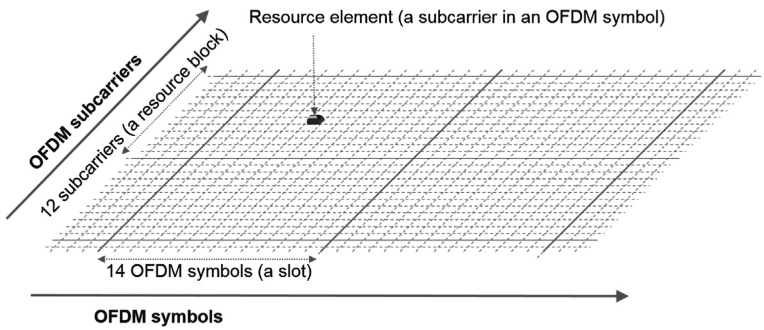

FIGURE 2.2

NR physical time-frequency structure.

In the time domain, the radio transmissions are organized into *radio frames*, *subframes*, *slots*, and *mini-slots*. As illustrated in Fig. 2.3, each radio frame has a duration of 10 ms and consists of 10 subframes with a subframe duration of 1 ms. A subframe is formed by one or multiple adjacent slots, each slot having 14 adjacent OFDM symbols. A mini-slot, in principle, can be as short as one OFDM symbol, but in Release 15 mini-slots are restricted to 2, 4, and 7 OFDM symbols. The time duration of a slot/mini-slot scales with the chosen numerology (subcarrier spacing) since the duration of an OFDM symbol is inversely proportional to its subcarrier spacing.

Physical layer uses time-frequency resources for transmission. In NR, the time-frequency resources (resource elements) represent either *physical channels* or *physical signals*, as in LTE. In the 3GPP terminology, a physical channel corresponds to a set of resource elements carrying information originating from higher layers, whereas a physical channel corresponds to a set of resource elements (used by physical layer) that do not contain information from higher layers. In the following sections we discuss both.

2.4 PHYSICAL CHANNELS

The time-frequency resources carrying information from higher layers (layers above PHY) are termed physical channels [1]. There are a number of physical channels specified for uplink and downlink:

- Physical downlink shared channel (PDSCH), used for downlink data transmission.

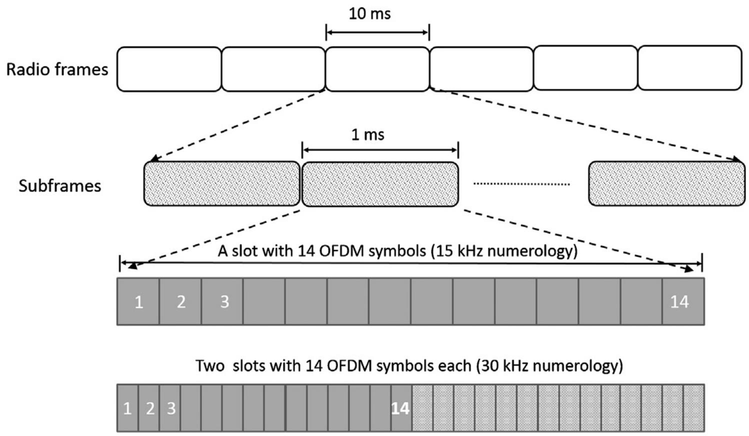

FIGURE 2.3

NR frame structure.

- Physical downlink control channel (PDCCH), used for downlink control information, which includes scheduling decisions required for downlink data (PDSCH) reception and for scheduling grants giving permission for uplink data (PUSCH) transmission by a UE.
- Physical broadcast channel (PBCH), used for broadcasting system information required by a UE to access the network.
- Physical uplink shared channel (PUSCH), used for uplink data transmission (by a UE).
- Physical uplink control channel (PUCCH), used for uplink control information, which includes: HARQ feedback acknowledgments (indicating whether a downlink transmission was successful or not), scheduling request (requesting time-frequency resources from network for uplink transmissions), and downlink channel-state information for link adaptation.
- Physical random access channel (PRACH), used by a UE to request connection setup referred to as random access.

On a high level, the downlink and uplink transmissions between gNB and UE work as follows. In downlink, a UE monitors the PDCCH, typically once per slot (PDCCH can also be configured more than once to enable ultralow-latency transmissions). Upon detection of a valid PDCCH, the UE receives one unit of data (named a transport block) on the PDSCH following the scheduling decision of gNB. Afterwards, the UE responds with a hybrid ARQ acknowledgment indicating whether the data was successfully decoded or not. In the case of unsuccessful decoding, retransmission(s) is scheduled. For uplink data transmission, a UE first requests gNB for physical time-frequency resources for the data awaiting transmission. This is termed a scheduling request and is sent over the PUCCH. In response, gNB sends a scheduling grant (over the PDCCH) which gives permission to a UE to use certain time-frequency resources for transmission. Following the scheduling grant, the UE schedules its data transmission over the PUSCH. gNB receives uplink data and sends hybrid ARQ acknowledgment indicating whether the uplink data transmission was successfully decoded or not. In the case of a decoding failure, retransmission(s) is scheduled. To enable ultralow-latency communication, the network can also preconfigure data transmission resources for a UE to avoid the signaling involved in sending scheduling request and scheduling grants. This is known as grant-free transmission. A drawback of this scheme is that time-frequency resources can be unnecessarily reserved for a UE even if there is no data awaiting transmission at the UE.

2.5 PHYSICAL SIGNALS

The time-frequency resources that are used by the PHY layer but do contain information from higher layers (i.e., layers above the PHY layer) are termed physical signals [1]. The physical signals are reference signals used for different purposes, for example, demodulation, channel estimation, synchronization, and channel-state information. There are different physical signals in the uplink and the downlink. The downlink physical signals include:

- demodulation reference signal (DM-RS)
- phase tracking reference signal (PT-RS)
- channel state information reference signal (CSI-RS)
- primary synchronization signal (PSS)
- secondary synchronization signal (SSS)

Furthermore, the following uplink physical signals are defined:

- demodulation reference signal (DM-RS)
- phase tracking reference signal (PT-RS)
- sounding reference signal (SRS)

NR has an ultralean design, which minimizes always-on transmissions to enhance network energy efficiency, reduce interference, and ensure forward compatibility. In contrast to the setup in LTE, the reference signals in NR are transmitted only when necessary. Next, we briefly discuss four main reference signals: DM-RS, PT-RS, CSI-RS, and SRS.

DM-RS is used to estimate the radio channel for demodulation. DM-RS is UE-specific, can be beam-formed, confined in a scheduled resource, and transmitted only when necessary, both in the downlink and the uplink. The DM-RS design takes into account the early decoding requirement to support low-latency applications. For this reason, the DM-RS is placed in the beginning of a slot (known as a front-loaded DM-RS). For low-speed scenarios, DM-RS uses low density in the time domain (i.e., fewer OFDM symbols in a slot contain DM-RS). For high-speed scenarios, the time density of DM-RS is increased to track fast changes in the radio channel.

PT-RS is introduced in NR to enable compensation of the oscillator phase noise. Typically, phase noise increases as a function of the oscillator carrier frequency. PT-RS can therefore be utilized at high carrier frequencies (such as by millimeter wave) to mitigate phase noise. One of the main degradations caused by phase noise in an OFDM signal is an identical phase rotation of all the subcarriers, known as common phase error (CPE). (This is discussed in detail in Chapter 6 and Chapter 7.) PT-RS is designed so that it is sparse in the frequency domain and has a high density in the time domain and the reason is as follows. The phase rotation produced by CPE is identical for all subcarriers within an OFDM symbol, but there is low correlation of phase noise across OFDM symbols. The density of PT-RS in frequency domain is one subcarrier in every PRB, every second PRB, or every fourth PRB. The density in time domain is every OFDM symbol, every second OFDM symbol, or every fourth OFDM symbol. Fig. 2.4 shows an example of DM-RS and PT-RS time-frequency structure. Like DM-RS, PT-RS is also UE-specific, it is confined in a scheduled resource, and it can be beam-formed. PT-RS is configurable depending on the quality of the oscillators, the carrier frequency, the OFDM subcarrier spacing, and on the modulation and coding schemes used for transmission.

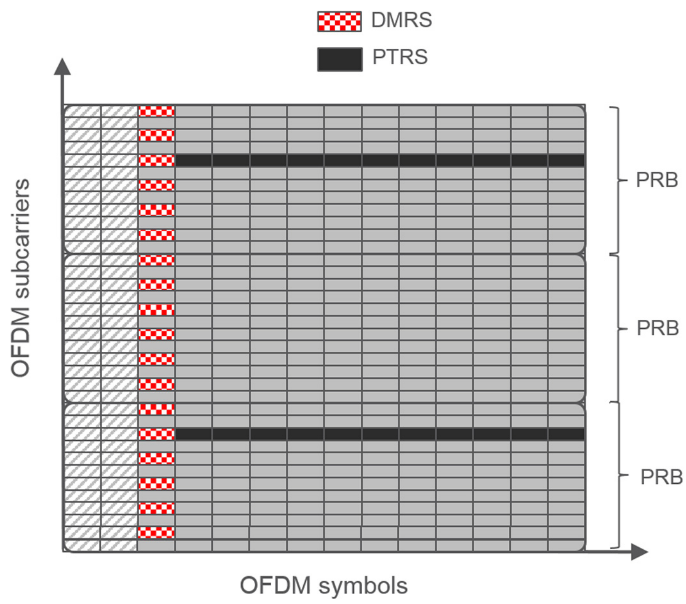

FIGURE 2.4

An example of time-frequency structure DM-RS and PT-RS.

CSI-RS is a downlink reference signal which is used mainly for CSI acquisition, beam management, time/frequency tracking and uplink power control. It has a very flexible design in order to support diverse use cases. CSI-RS for CSI acquisition is used for determining CSI parameters such as channel quality indicator (CQI), rank indicator (RI), and precoding matrix indicator (PMI), which are needed for link adaptation and for determining precoders. Furthermore, a so-called CSI interference measurement (CSI-IM) resource, which is a zero-power CSI-RS (ZP CSI-RS) resource, can be configured for interference measurements in the UE. For beam management, CSI-RS is used for evaluating candidate transmission beams by measuring reference signal received power (RSRP) for each beam. It can also be used for beam recovery purposes. A CSI-RS configured for time/frequency tracking is called a tracking reference signal (TRS). TRS can be used for fine time and frequency synchronization, and Doppler and delay spread estimation. This is needed for channel estimation and demodulation.

The SRS is transmitted in uplink to perform CSI measurements mainly for scheduling and link adaptation. For NR, it is expected that the SRS will also be utilized for reciprocity-based precoder design for massive MIMO and uplink beam management. SRS has a modular and flexible design to support different procedures and UE capabilities.

Table 2.2 Semi-static TDD configurations

OFDM numerology	Uplink/downlink switching periodicities [ms]
15 kHz	0.5, 1, 2, 5, 10
30 kHz	0.5, 1, 2, 2.5, 5, 10
60 kHz	0.5, 1, 1.25, 2, 2.5, 5, 10
120 kHz	0.5, 0.625, 1, 1.25, 2, 2.5, 5, 10

2.6 DUPLEXING SCHEME

NR supports TDD and FDD transmissions, as in LTE. The duplex scheme typically depends on spectrum allocation. At lower frequencies, the spectrum allocations are mostly paired, implying FDD transmission. At higher frequencies the spectrum allocations are often unpaired, implying TDD. In addition, NR supports dynamic TDD, where uplink and downlink allocations dynamically change over time. This is one of the key enhancements over LTE which is useful in scenarios with rapid traffic variations. The transmission scheduling decisions are made by the gNB scheduler and the UEs follow these scheduling decisions. The network can coordinate scheduling decisions between neighboring network sites to avoid interference, if necessary. TDD can also be semi-statically configured with certain uplink/downlink switching periodicities. There are uplink/downlink switching periodicities specified in Release 15 for different OFDM numerologies given in Table 2.2.

2.7 FRAME STRUCTURE

The NR frame structure follows three key design principles to enhance forward compatibility and reduce interactions between different functionalities. The first principle is that transmissions are self-contained. Data in a slot and in a beam is decodable on its own without dependency on other slots and beams. This implies that reference signals required for the demodulation of data are included in a given slot and a given beam. The second principle is that transmissions are well confined in time and frequency. Keeping transmissions confined makes it easier to introduce new types of transmissions in parallel with legacy transmissions in the future. The NR frame structure avoids the mapping of control channels across the full system bandwidth. The third principle is to avoid static and/or strict timing relations across slots and across different transmission directions. For example, asynchronous HARQ is used instead of a predefined retransmission time.

The NR frame structure supports TDD and FDD transmissions and operation in both the licensed and the unlicensed spectrum. It enables very low latency, fast HARQ acknowledgments, dynamic TDD, coexistence with LTE and transmissions of variable length (for example, short duration for URLLC and long duration for enhanced MBB (eMBB)). Considering the TDD operation, Fig. 2.5 provides examples of NR frame structure for different scenarios.

NR can also employ mini-slots to support transmissions with a flexible start position and a duration shorter than a regular slot duration. In principle, a mini-slot can be as short as one OFDM symbol and can start at any time. In Release 15, mini-slots are limited to 2, 4, and 7 OFDM symbols. Mini-slots can be useful in various scenarios, including low-latency transmissions, transmissions in unlicensed spectrum and transmissions in the millimeter-wave spectrum. In low-latency scenarios, transmission needs

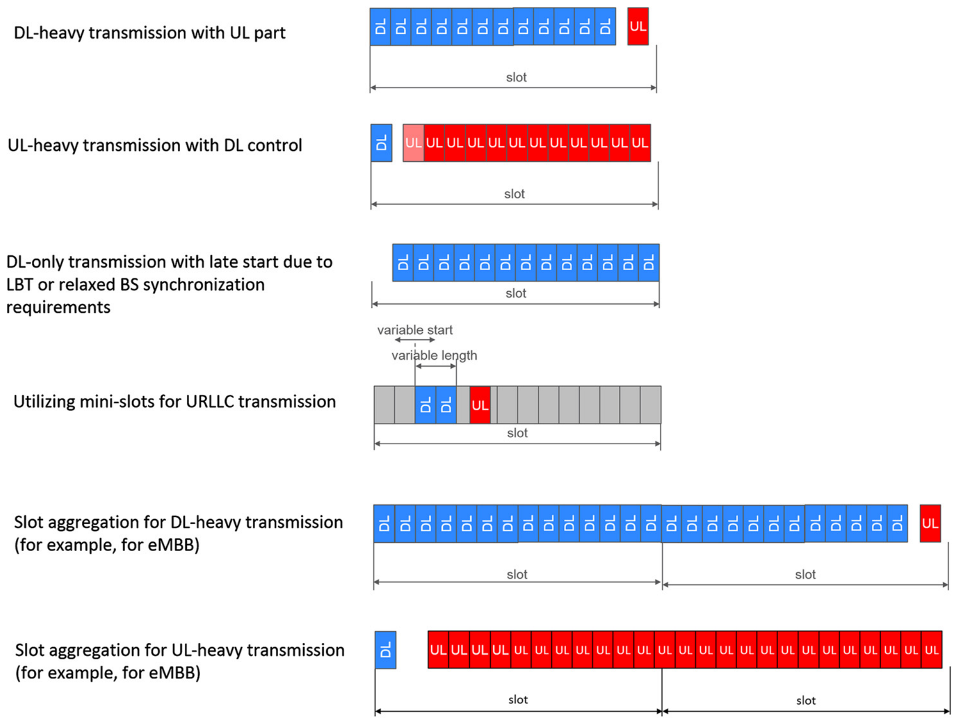

FIGURE 2.5

Examples of NR physical layer frame structure.

to begin immediately without waiting for the start of a slot boundary (ultra-reliable low-latency communications (URLLC), for example). When transmitting in the unlicensed spectrum, it is beneficial to start transmission immediately after the listen-before-talk (LBT) mechanism. When transmitting in the millimeter-wave band, the large amount of bandwidth available implies that the payload supported by a few OFDM symbols is large enough for many of the packets. Fig. 2.5 provides examples of URLLC- and LBT-based transmission in unlicensed spectrum via mini-slots and illustrates that multiple slots can be aggregated for services that do not require extremely low latency (eMBB, for example). Having a longer transmission duration helps to increase coverage or reduce the overhead due to switching (in TDD), transmission of reference signals, and control information.

The same frame structure can be used for FDD, by enabling simultaneous reception and transmission (that is, downlink and uplink can overlap in time). This frame structure is also applicable to device-to-device (D2D) communications. In that case, the downlink slot structure can be used by the device that is initiating (or scheduling) the transmission, and the uplink slot structure can be used by the device responding to the transmission.

To obtain low latency, a slot (or a set of slots in the case of slot aggregation) is front loaded with control signals and reference signals at the beginning of the slot (or set of slots), as illustrated in Fig. 2.6. An NR frame structure also allows for rapid HARQ acknowledgment, in which decoding is performed during the reception of downlink data and the HARQ acknowledgment is prepared by the UE during the guard period, when switching from downlink reception to uplink transmission. NR supports very short duration uplink control signals (along with longer formats) to provide fast HARQ feedbacks from a device to the base station. Fig. 2.6 shows an example of a self-contained slot where the delay from the end of the data transmission to the reception of the acknowledgment from the device is on the order of just one OFDM symbol.

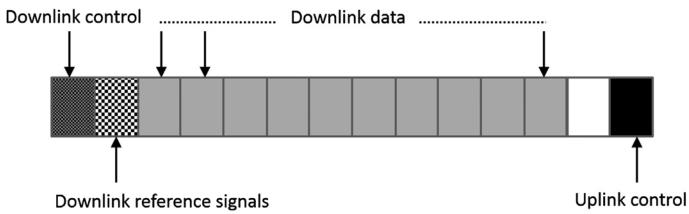

FIGURE 2.6

A self-contained slot.

2.8 PHY PROCEDURES AND MEASUREMENTS

There are a number physical layer procedures defined in the NR specifications, for example for cell search, power control, uplink synchronization and timing control, random access, and beam management, and there are channel-state information related procedures [2,3]. The NR specification also defines physical layer measurements for intra- and inter-frequency handover, inter RAT handover, timing measurements, and measurements for radio resource management [4]. These details (apart from the beam management procedure) are out of scope of this book. We strongly recommend [6] for understanding these details.

2.9 PHYSICAL LAYER CHALLENGES

5G NR is the first cellular technology to operate at millimeter-wave frequencies, support GHz of bandwidths, and utilize a massive number of antennas. These aspects impose a number of challenges for operation of NR physical layer mainly due to less understood radio wave propagation characteristics and hardware impairments (at the base stations and the devices). To enable a high performing NR, it is important to accurately understand the characteristics of radio propagation and hardware impairments. Chapters 3 and 4 address radio wave propagation, channel modeling, and hardware impairment modeling in great detail. In the following, we provide a brief discussion of some of these challenges.

2.9.1 PROPAGATION RELATED CHALLENGES

When going up in frequency the antenna related transmission loss increases as the square of the frequency, for any fixed receive antenna pattern, because the aperture is proportional to the square of the wavelength. Some additional increase in loss due to propagation effects is likely, particularly under NLoS conditions. The challenge here is to utilize advanced multiantenna techniques to steer the signal into favorable directions and to increase the receive antenna aperture. At this point it is not fully clear to what extent the channel conditions may be improved at higher frequencies by means of these antenna techniques. In some scenarios, like LoS, it may even be favorable to go up in frequency. It is clear that the outdoor-to-indoor scenario will be challenging as the penetration loss goes up substantially with frequency. Another effect of using beam-forming and narrow beams is that the channel's dynamic changes will be larger and faster due to sudden blockage of the beam. Furthermore, the directional spread characteristics of the propagation channel is not well known. For channels which are highly dispersive in direction (rich scattering) high gain antennas are not very useful as they will only catch/direct a small fraction of the signal from the transmitter/to the receiver. In this case fully coherent antenna combining techniques are more suitable. However, such techniques are very complex and resource costly when the array antenna sizes are large in terms of the number of elements.

The main challenge for NR is to what extent novel multiantenna techniques may compensate for loss of, or even gain in, performance at higher frequencies comparing with legacy techniques used at lower frequencies. Under LoS conditions it is favorable to use higher frequencies and beam-forming as shown in Chapter 3. Moreover, the propagation losses in NLoS do not increase dramatically with frequency (in the range $0 \log f$ to $6 \log f$) except for outdoor-to-indoor transmissions. However, as early deployments of NR are expected to largely rely on analog beam-forming, the propagation channel is required to be highly directive. There are unfortunately only few highly directionally resolved channel measurements, implying that the possible beam-forming performance at millimeter-wave frequencies is largely unknown.

2.9.2 HARDWARE RELATED CHALLENGES

As regards the demands for increasingly high data rates moving into NR, the need for spectral efficiency and more bandwidth is rapidly increasing. The technology components used to address this increase involve advanced multiantenna techniques such as massive multi user (MU)-MIMO or, moving into the millimeter-wave regime, analog beam-forming. This raises new challenges in terms of efficient radio implementation as both the number of deployed transceivers and their operating frequencies and bandwidths increase.

As one of the important building-blocks in radio transceiver front-ends, the RF power amplifier (PA) continues to play a critical role as a major consumer of power. Operating in dense and highly integrated antenna arrays, the PA may also suffer the effects of mutual coupling unless sufficient isolation is implemented. This introduces another source of distortion caused by load-modulation as the active impedance presented to the amplifier changes over time. Another critical issue in antenna systems using directive transmission is assessing the distortion behavior in terms of how it is distributed over space. With conventional methods, the distortion follows the same pattern as the desired signal, but using MU-MIMO precoding or beam-forming techniques, this is generally not the case.

RF oscillators are another source of imperfections in the analog front-end, which, when moving toward the millimeter-wave bands, becomes a limiting factor. Maintaining a stable oscillation becomes more difficult at very high frequencies as the increasing losses decrease the quality factor of the resonator-tanks and there is a lack of power generation due to fundamental limitations in transistor technology. Thus, as the phase noise increases, coherent transmission may become increasingly difficult. Thus, phase noise needs to be accurately modeled in order to assess its impact and to provide a basis for countermeasures such as tracking. Phase-locked loop (PLL) techniques may be used in order to decrease the narrow-band $1/f$ part of the phase noise, which may have more significant negative effects on OFDM-based transmission, but this comes at the cost of increased white noise levels.

As advanced signal processing hardware and algorithms make it possible to compensate for some of the more prominent impairments, data-converters such analog-to-digital converter (ADC) and digital-to-analog converter (DAC) are crucial components connecting the analog and digital domains. Unfortunately, data-converters do not enjoy the same beneficial scaling as described by Moore's law, as these are mixed signal components and generally require linear transistor technologies. Scaling down the geometries to increase the available processing speed is not necessarily a good option in this case, as it makes the transistors operate more as switches. As shown in the literature, massive MU-MIMO opens up possibilities to decrease the effective resolution. This will, however, inflict an increasing amount of quantization noise which, if left unchecked, may corrupt the signal to a large extent. In order to assess the impact of a coarser quantization, different modeling techniques based on either deterministic behavioral models or stochastic processes may aid us.

Overall, the need for improved mathematical tools capable of assessing the behavior of non-ideal radio components over time/frequency and space is larger than ever. Specifically this is so since these models may serve as a basis for advanced compensation techniques such as digital pre-distortion (DPD) or phase-noise tracking. Current developments in this field have further opened up possibilities of new analysis and understanding of how radio imperfections behave in the context of a large antenna array. Another relevant field of modeling which currently is in its cradle is that of stochastic modeling of radio impairments aimed to aid performance analysis on link- or system-level, as the kind of oversampled passband or baseband data most behavioral models are built upon may not always be available. This modeling framework will be further discussed in Chapter 4.

REFERENCES

[1] 3GPP, TS 38.211 NR Physical channels and modulation, June 2018.
[2] 3GPP, TS 38.213 NR Physical layer procedures for control, June 2018.
[3] 3GPP, TS 38.214 NR Physical layer procedures for data, June 2018.
[4] 3GPP, TS 38.215 NR Physical layer measurements, June 2018.
[5] E. Dahlman, S. Parkvall, J. Sköld, 4G LTE-Advanced Pro and The Road to 5G, Academic Press, 2016.
[6] E. Dahlman, S. Parkvall, J. Sköld, 5G NR: The Next Generation Wireless Access Technology, Academic Press, 2018.