Chapter 10
5G NR Overview and Physical Layer

10.1 Introduction

This text deals primarily with the physical layer. However, to better understand its operation, it is helpful to stand back a bit and study, at a high level, its surrounding architecture and protocols. To this end, we review in this chapter the main architecture options for connection to the core network followed by the *Radio Access Network* (RAN) protocol architecture. The RAN is responsible for all radio-related functions such as coding, modulation, HARQ operation, physical transmission, scheduling, etc. Following this, we narrow our study within the RAN to the physical layer, with emphasis on how physical channels are structured, be they user data conveying or control data conveying, as well as how physical signals are created, such signals being those originating solely in the physical layer. Procedures such as initial access, scheduling, and uplink power and timing control are introduced. How maximum user data rates and low latency are achieved is demonstrated, 5G operating frequency spectrum reviewed, some typical base station and UE parameters presented, and relevant 3GPP specifications listed.

10.2 Connection to the Core Network

The *core network* is responsible for the overall control of the UE. Functionally, it sits above the RAN and handles functions not related to radio access but required for the providing of a complete network such as authentication and the establishment, maintenance, and release of communication links. To ease the transition to a fully independent 5G network, 3GPP specified two primary core connection architectures: *non-standalone*, for which there are a number of variations based on the routing of user data, and *standalone*. Shown in Fig. 10.1 is the stand-alone architecture called Solution 2 and one of the non-stand-alone architectures called Solution 3×.

D. H. Morais, *Key 5G Physical Layer Technologies*,
https://doi.org/10.1007/978-3-030-51441-9_10

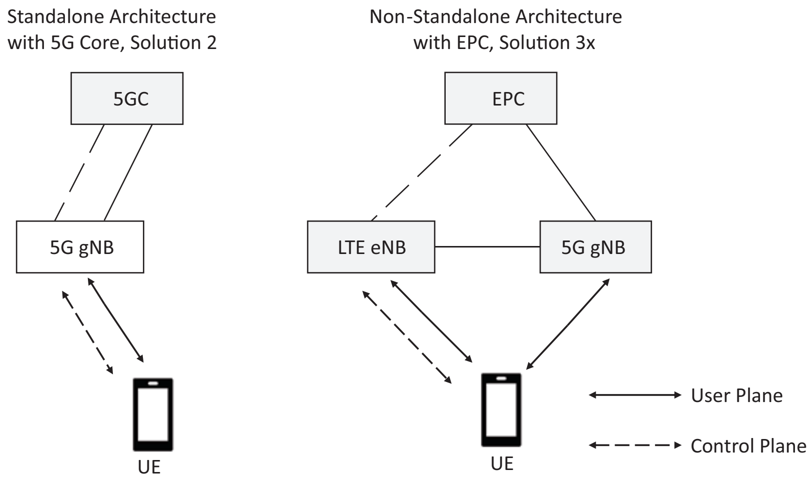

Fig. 10.1 Non-stand-alone and stand-alone core connections

In the non-stand-alone architecture, both the LTE core, called the Evolved Packet Core (EPC), and the LTE base station, called the eNodeB, are utilized. Control plane data is all done via connection to the eNodeB, and user plane data to/from the UE is via either the eNodeB or the NR base station called the gNB or simultaneously with both. When the latter is the case, this is referred to as *dual connectivity*, which can increase the user data rate and increase reliability. From the point of view of available 5G service, only enhanced mobile broadband (eMBB) is available with the non-stand-alone architecture. Note, however, that dual connectivity can also be 5G–5G, where the UE communicates with two gNBs. 3GPP Release 15 assumes synchronization between the different gNBs; however, in Release 16, this limitation is removed.

In the stand-alone architecture, no elements of LTE are utilized. The UE is connected to the 5G core (5GC) via the gNB. This makes available all new functionality provided by the 5GC. Thus, in addition to eMBB, ultra-reliable low-latency communication (URLLC) is available via Rel. 15 and massive machine-type communications (mMTC) via Rel. 16.

10.3 RAN Protocol Architecture

The 5G RAN protocol architecture of the gNB and the UE [1–3] is shown in Fig. 10.2. The protocols consist of a *user plane* (UP) and a *control plane* (CP). The UP transports user data which enters/leaves from above in the form of IP packets. The control plane transports control signaling information and is mainly responsible for connection establishment and maintenance, mobility, and security. The protocol

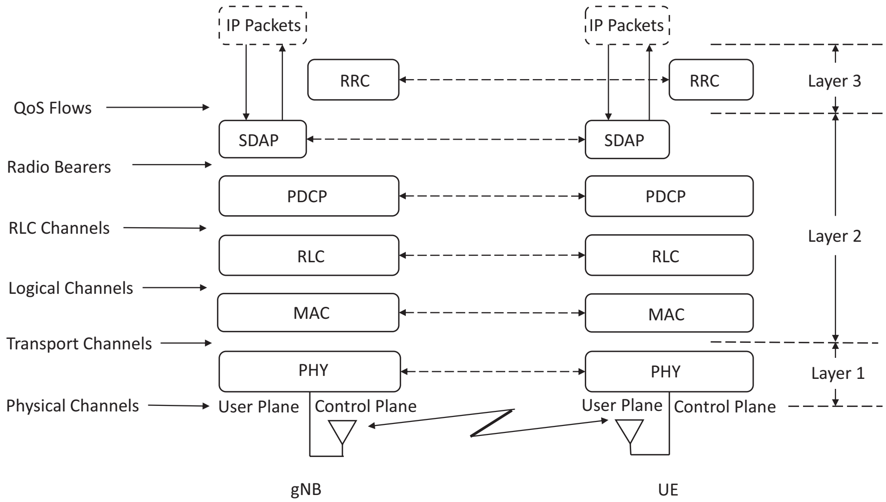

Fig. 10.2 User plane and control plane protocols

stack is divided into layers. Layer 1 is the physical layer. Layer 2 encompassed the MAC, RLC, PDCP, and SDAP sublayers. Layer 3 is the RRC layer. It will be noted that most of the protocols are common to the user plane and control plane. Shown in Fig. 10.2 is the nomenclature of connections between layer/sublayer and sublayer/sublayer. Logical channels, which are offered to the RLC sublayer by the MAC sublayer, are characterized by the type of information they carry. Transport channels, which are offered to the MAC sublayer by the physical layer, are characterized by how and with what characteristic information is transferred over the radio interface. Data on a transport channel is arranged into *transport blocks* (TBs). Physical channels carry data from higher layers between the gNB and UE.

Before proceeding further, the definition of some commonly used terms in describing the various protocols is in order:

– A d*ata unit* is the basic unit exchanged between different layers of a protocol stack.
– A *service data unit* (SDU) is a data unit passed by a layer above to the current layer for transmission using services of the current layer.
– A *protocol data unit* (PDU) is a data unit created by the current layer via the adding of a header to the received SDU prior to transportation to the layer below. The header added describes the processing carried out by the current layer.
– A *data radio bearer* (DRB) is a radio bearer conveying user data received from the SDAP.
– A *signaling radio bearer* (SRB) is a radio bearer conveying control information received from the RRC.

10.4 Layer 3 (RRC) Description

As seen in Fig. 10.2, Level 3 has one protocol, namely, the *Radio Resource Control* (RRC) protocol, which resides in the control plane. The overall task of the RRC is to configure the UE with the parameters required by the other protocol layers to establish and maintain connectivity between the UE and the gNB. It does this by handling control layer procedures, including:

– Broadcast of *system information* (SI). System information is all the non-device-specific information that a UE needs to be able to communicate with the network via the gNB. System information consists of a *master information block* (MIB) and a number of *system information blocks* (SIBs).
– Transmission of paging messages emanating from the 5GC or the gNB to the UE via the gNB to notify the UE about connection requests.
– The establishment, maintenance, and release of an RRC connection between the UE and the gNB.
– Security functions.
– Establishment, configuration, maintenance, and release of SRBs and DRBs.
– Mobility functions, including UE cell selection and reselection and handover.
– QoS management functions.
– UE measurement reporting.
– Detection and recovery from a failure of the radio link.

RRC messages are sent via SRBs as these always have a higher priority than DRBs, thus ensuring that they are sent as quickly as possible.

The RRC can be in one of three different states depending on the traffic activity. These states are RRC_IDLE, RRC_INACTIVE, and RRC_CONNECTED.

In RRC_IDLE state, the parameters necessary for communication between a UE and a gNB have not been assigned to a specific gNB. Thus, no data transfer can take place to/from the UE, and the UE “sleeps” most of the time, waking up only periodically to see if there is a paging message from the network which may lead to a change of RRC state.

In the RRC_INACTIVE state, the parameters necessary for communication between the UE and a specific gNB have been assigned to a specific gNB, but there is no data transfer. However, when there is data to be transferred, the RRC quickly changes state to RRC_CONNECTED.

In the RRC_CONNECTED state, the parameters necessary for communication between the UE and a specific gNB have been assigned to a specific gNB, and data transfer takes place.

10.5 Layer 2 User Plane and Control Plane Protocol Description

A detailed layer 2 user plane protocol and control plane structure is shown in Fig. 10.3.

Following is a description of the various protocol sublayers, starting from the top.

10.5.1 The Service Data Adaptation Protocol (SDAP) Sublayer

In NR, IP packets (IPv4 or IPv6) as described in Sect. 2.2 are mapped from the 5G core network to data radio bearers (DRBs) according to their quality of service (QoS) requirements. The *Service Data Adaptation Protocol* (SDAP) is a user plane sublayer and is responsible for this mapping. It also marks both the DL and UL packets with an identifier as to the specified QoS. The SDAP is a new protocol and not present in 4G. If the gNB is connected to the EPC, as in the non-stand-alone case shown in Fig. 10.1, the SDAP is not used.

10.5.2 The Packet Data Convergence Protocol (PDCP) Sublayer

The *Packet Data Convergence Protocol* (PDCP) is responsible for several services and functions [1]. The following is a description of the main ones:

– Sequence numbering, in the downward flow in the user plane, of data radio bearers so that in-sequence reordering is possible in the upward flow.
– Header compression, in the user plane only, to reduce the number of bits transmitted over the radio interface. Header compression is accomplished via robust header compression (ROHC) as discussed in Sect. 2.3. In the upward flow, header decompression is applied.
– Integrity protection, in the downward flow in the control plane, to ensure that control messages originate from the correct source. In the upward flow, integrity verification is applied.
– Ciphering, also known as encryption, in the downward flow in both the user plane and control plane, to ensure that intruders cannot access the data and signaling messages that the UE and the gNB exchange. In the upward flow, deciphering is applied.
– Duplicate detection and removal in the upward flow in the user plane and, if in order delivery to layers above is required, reordering to provide in-sequence delivery.

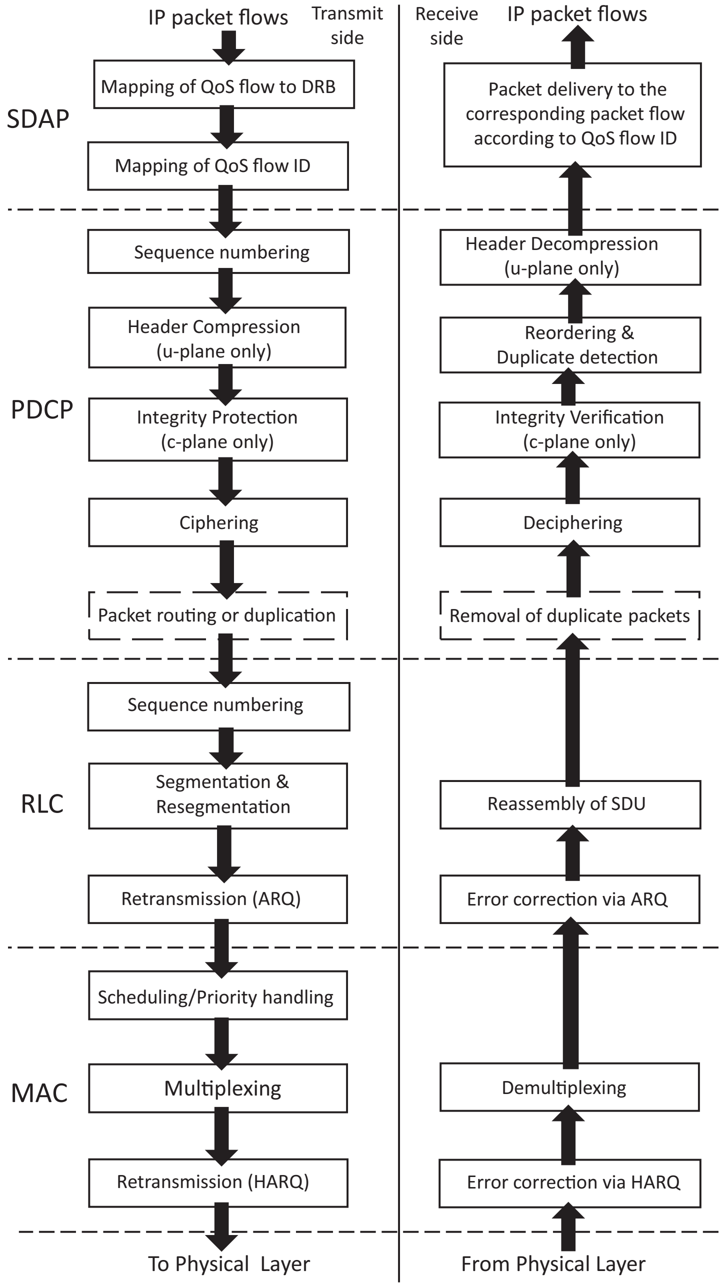

Fig. 10.3 NR user plane and control plane protocol layer structure

10.5.3 The Radio Link Control (RLC) Sublayer

The RLC sits below the PDCP sublayer and interfaces with the MAC sublayer below via logical channels. Depending of the class of service to be provided, the *Radio Link Control* (RLC) sublayer can be structured in one of three transmission modes, namely, the *transparent mode* (TM), the *unacknowledged mode* (UM), and the *acknowledged mode* (AM). In the transparent mode, data unit flow is transparent, and no headers are added. In the unacknowledged mode, segmentation (discussed below) is supported, and, in the acknowledged mode, segmentation and retransmission of erroneous packets are supported. The RLC is responsible for several functions and services including:

– Sequence numbering of user data in the downward flow, independent of the one in the PDCP above. This numbering is in support of HARQ retransmissions. In the upward flow, RLC does not facilitate in-sequence delivery, minimizing the overall latency. This task is left to the PDCP sublayer above.
– Segmentation in the downward flow, of RLC SDUs, received from the PDCP above, into suitably sized RLC PDUs. In the upward flow, reassembly of the original RLC SDUs.
– Retransmission, in the acknowledgment mode, in the downward flow, via HARQ, of erroneously received RLC PDUs at the receiving end.
– Error correction in the upward flow via HARQ.

10.5.4 The Medium Access Control (MAC) Sublayer

The *Medium Access Control* (MAC) sublayer is the lowest sublayer of Level 2 and interfaces with the physical layer below via transport channels. As indicated above, data on a transport channel is arranged into transport blocks (TBs), and the transmission time of each transport block is called the *Transmission Time Interval* (TTI). One transport block of variable size can be transmitted in each TTI over the radio interface to/from an EU except when there is spatial multiplexing of more than four layers, in which case two transport blocks are transmitted per TTI.

The MAC sublayer is responsible for several functions and services including:

– Mapping between logical channels and transport channels.
– In the downward flow, multiplexing of MAC SDUs (RLC PDUs) belonging to one or different logical channels into transport blocks delivered to the physical layer on transport channels. Associated with each transport block is a *transport format* (TF) which defines how the transport block is to be transmitted over the radio interface. The TF conveys information about the size of the transport block, the coding and modulation scheme to be employed, and the antenna mapping.

- In the upward flow, demultiplexing of MAC SDUs belonging to one or different logical channels from transport blocks delivered from the physical layer on transport channels.
- Error correction via HARQ.
- Scheduling and scheduling-associated functions.

The MAC sublayer provides services to the RLC via logical channels that can be classified into two groups: control channels used for transporting control plane signaling and traffic channels used for carrying user plane data. The logical channels defined in 3GPP NR specifications are:

- The *Broadcast Control Channel* (BCCH): A DL channel used for broadcasting system information (SI) received from the RRC above to all UEs in a cell. In order to access the network, a UE needs to obtain the SI to learn how the system is configured and how to operate within the cell. As indicated above, SI consists of an MIB and several SIBs. The MIB is passed down to the BCH transport channel and the SIBs to the DL-SCH transport channel (transport channels described below). Among the SIBs is SIB1, referred to as *remaining minimum system information* (RMSI), and consists of the system information that the UE needs over and above the MIB in order to access the network.
- The *Paging Control Channel* (PCCH): A DL channel transmitted in multiple cells to page UEs whose location in a given cell is not known to the network and to notify UEs of system information changes and Public Warning System broadcasts.
- The *Common Control Channel* (CCCH): A DL and UL channel used for transmitting control information between UEs and the network for UEs having no RRC connection with the network.
- The *Dedicated Control Channel* (DCCH): A point-to-point DL and UL channel used to transmit dedicated control information between a UE and the network. It is used by UEs having an RRC connection for the configuration of the UE.
- The *Dedicated Traffic Channel* (DTCH): A point-to-point DL and UL channel, dedicated to one UE, for the transfer of user information to/from the UE.

Logical channels are mapped to transport channels. The transport channels defined in NR specifications are:

- The *Broadcast Channel* (BCH): A DL channel broadcast in the entire coverage area of the cell. It conveys the master information block (MIB) received from the BCCH logical channel above. The MIB is a block of data that contains a limited amount of the total information that the UE requires in order to acquire the remaining system information (SIBs) broadcast by the network via the DL-SCH. The transport block that it delivers to the physical layer has a fixed format defined by the specifications.
- The *Paging Channel* (PCH): A DL channel broadcast in the entire coverage area of the cell, used for the transmission of paging information from the PCCH logical channel above. It supports UE *discontinuous reception* (DRX) to enable UE

power saving by sleeping and waking up only at predefined times to receive the PCH.

– The *Downlink Shared Channel* (DL-SCH): The DL channel used to for the transmission of user data received from the DTCH above. It is also used for transmission of SIBs received from the BCCH above, transmission of UE-specific control information received from the DCCH above, and transmission of common control information received from the CCCH. It supports HARQ with soft combining and dynamic link adaptation via the varying of coding, modulation, and transmit power. It further supports both dynamic and semi-static resource allocation, discontinuous reception (DRX), and the possibility to use beamforming.
– The *Uplink Shared Channel* (UL-SCH): Similar to the DL-SCH, the UL-SCH is the UL channel used for the transmission of user data. It supports HARQ with soft combining and dynamic link adaptation via the varying of coding, modulation, and transmit power. It further supports both dynamic and semi-static resource allocation and the possibility to use beamforming.
– The *Random-Access Channel* (RACH): A UL channel used by the UE to request access to the network when the UE does not have accurate UL timing synchronization or does not have any allocated UL transmission resource. It carries the risk of collision with other transmissions in which event it must back off and try again. Note that although it is defined as a transport channel, it does not carry transport blocks.

The mapping between logical channels and transport channels as well as that between transport channels and physical channels is shown Fig. 10.4 for both the DL and UL. Also shown is physical signals which are described below.

An example of Layer 2 user plane downward data flow is shown in Fig. 10.5. Here, at the top, are three IP packets, two on radio bearer RB_x and one on radio

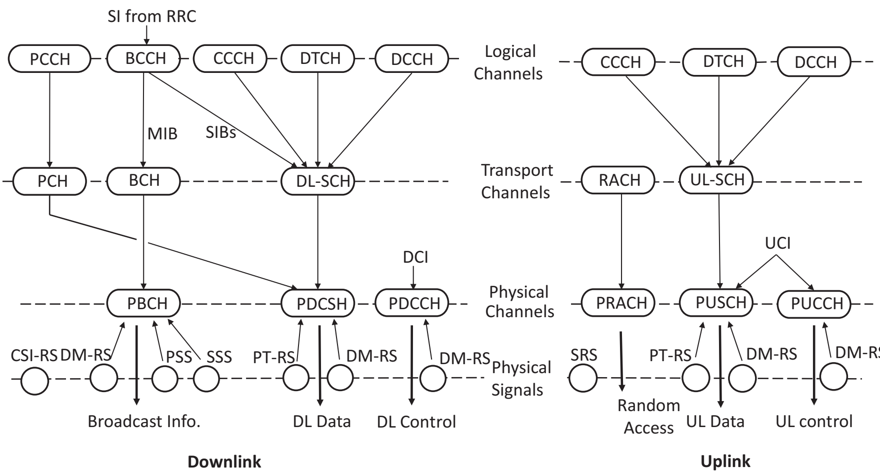

Fig. 10.4 Physical signals and mapping between logical, transport, and physical channels

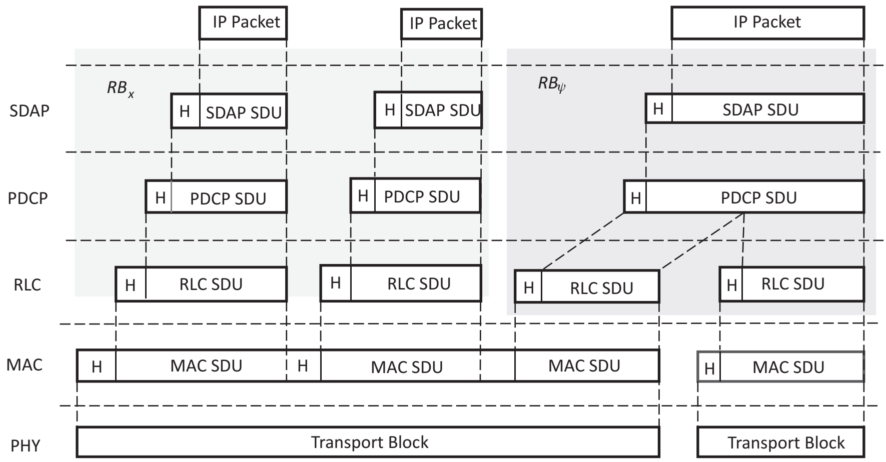

Fig. 10.5 An example of Layer 2 user plane data flow

bearer RB_y. First the IP packets enter the SDAP layer as SDAP SDUs where they each have a header added to create SDAP PDUs. These PDUs enter the PDCP layer as PDCP SDUs where, after processing including header compression, each has a header added to create PDCP PDUs. These PDUs enter the RLC layer where the ones created from radio bearer RB_x become RLC SDUs and the one created from radio bearer RB_y is segmented into two RLC SDU segments. After processing, the segmented and unsegmented SDUs each have a header added to create RLC PDUs which are then forwarded to the MAC layer where they are received as MAC SDUs. These SDUs, after processing, each have a header added creating MAC PDUs. The first three MAC PUDs counting from left to right (two from RB_x and one from RB_y) are multiplexed to form one transport block which is forwarded to the physical layer, and the fourth MAC PDU (from RB_y), on its own, is forwarded to the physical layer as another transport block.

10.6 Layer 1 (Physical Layer) Description

The physical layer, the main focus of this book, is the lowest layer, Layer 1, in the RAN protocol architecture. Its input on the downward flow and output on the upward flow is transport blocks. As mentioned above, one transport block of variable size can be transmitted in each TTI over the radio interface to/from an EU except when there is spatial multiplexing of more than four layers, in which case two transport blocks are transmitted per TTI.

The physical layer is responsible for many functions including CRC attachment, coding, rate matching and HARQ, scrambling, linear modulation, layer mapping, and mapping of the signal to the assigned physical time-frequency resource. It also handles the mapping of transport channels to physical channels. The physical channels defined in NR specifications are:

– The *Physical Downlink Control Channel* (PDCCH): A DL channel used to convey control information to the UE, including scheduling decisions required by the UE to know when and where to receive data, parameters used by the PDSCH for its transmission, and scheduling grants for uplink transmission.
– The *Physical Downlink Shared Channel* (PDSCH): A DL channel that carries user data, system information, paging information, and control information from layers above. Specifically, it provides the physical layer to transport information from the DL-SCH and PCH transport channels. Its allocation and other parameters used for its transmission are signaled to the UE by the PDCCH.
– The *Physical Broadcast Channel* (PBCH): A DL channel fed by the BCH above, which carries some of the required system information to allow a UE to access the network.
– The *Physical Uplink Control Channel* (PUCCH): A UL channel used by the UE to send HARQ acknowledgments to the gNB, indicating whether or not the DL transport block was successfully received, to send reports of the state of the DL channel to aid in the DL channel-dependent scheduling and to request UL scheduling grants.
– The *Physical Uplink Shared* Channel (PUSCH): The UL version of the PDSCH, providing the physical layer to transport user information from the UL-SCH.
– The *Physical Random-Access Channel* (PRACH): A UL channel used to enable the random-access procedure by physically transmitting information from the RACH.

Besides the physical channels, *physical signals* are also employed in NR transmission and their relationship to physical channels shown in Fig. 10.4. Physical signals are time-frequency resources used by the physical layer, but which do not contain information acquired from layers above. They are reference signals (pilot subcarriers) used for different purposes, for example, demodulation and channel estimation, and synchronization signals used for UE synchronization with the gNB. The physical signals defined in NR specifications are:

– *Demodulation Reference Signals* (DM-RSs): Used in both the DL and UL to estimate the radio channel (the channel coefficients) for demodulation purposes. It is UE-specific, confined to a scheduled resource, can be beam-formed, and transmitted only as necessary. In the DL, there is a DM-RS for the PDSCH, one for the PDCCH, and one for the PCBH. In the UL, there is one for the PUSCH and one for the PUCCH.
– The *Phase-Tracking Reference Signal* (PT-RS): Used in both the DL and UL to facilitate compensation of oscillator phase noise (Sect. 7.2.4) and thus employed when transmission is in a millimeter wave band where oscillator phase noise is likely to be high. It allows the reduction of common phase error at the receive end. It is UE-specific, confined in a scheduled resource, and can be beam-formed.

– The *Channel State Information Reference Signal* (CSI-RS): Used in the DL only to allow the UE to acquire channel state information (CSI) and report this to the gNB to facilitate link adaptation. It also supports reference signal received power (RSRP) measurements for mobility and beam management, time/frequency tracking for demodulation, and UL reciprocity-based precoding. It is UE-specific but multiple users can share the same CSI-RS resource. It is defined as either zero power (ZP-CSI-RS) or nonzero power (NZP-CSI-RS). When configured as ZP-CSI-RS, the resource elements assigned to it are made unavailable for PDSCH transmission as they are configured for CSI-RS transmission on another device.
– The *Tracking Reference Signal* (TRS): Used to aid the UE in time and frequency tracking made necessary because of time and frequency variations in its local oscillator. It is not a defined physical signal per se, but rather a resource set consisting of multiple periodic NZP-CSI-RSs.
– The *Sounding Reference Signal* (SRS): Used in the UL only to enable the gNB to perform CSI measurements to be used mainly for frequency domain scheduling and link adaptation. It is also useful in the case of FDD where there is no channel reciprocity and thus no UL CSI information available from UE DL measurements.
– The *Primary Synchronization Signal* (PSS) and the *Secondary Synchronization Signal* (SSS): This pair of signals, used in the DL only, is employed by the UE during initial access (Sect. 10.13). Together, the PSS and the SSS create the capability to transmit one of 1008 possible physical cell identities (cell IDs). In the first stage of initial access, the UE searches for the PSS and SSS in order to detect the presence of a gNB, acquire accurate timing and frequency of the system, and determine the physical cell identity. The PSS provides initial timing and frequency information, and the SSS provides refinement to the initial findings.

10.7 The Frame (Time-Domain) Structure

The NR frame structure supports both frequency division duplexing (FDD) and time division duplexing (TDD) operations in both licensed and unlicensed frequency bands. It facilitates very low latency, fast acknowledgment of HARQ and dynamic TDD. Transmissions in the time domain, in both the DL and UL, are organized into *frames*. There is one set of frames in the UL and one in the DL on a carrier. Each UL frame transmitted from the UE starts at a specified time [4] before the start of its corresponding DL frame at the gNB to ensure that the UL frame, when it arrives at the gNB, is synchronized with the DL frame at the gNB.

Frames are of 10 ms duration, each frame being divided into ten subframes, with each subframe thus being of 1 ms duration. In turn, each subframe is divided into *slots*, each slot consisting of 14 OFDM symbols. Note, however, that there is an exception to this when the subcarrier spacing is 60 kHz and the extended cyclic prefix (Sect. 9.6) is employed. Here there are only 12 OFDM symbols per slot. A

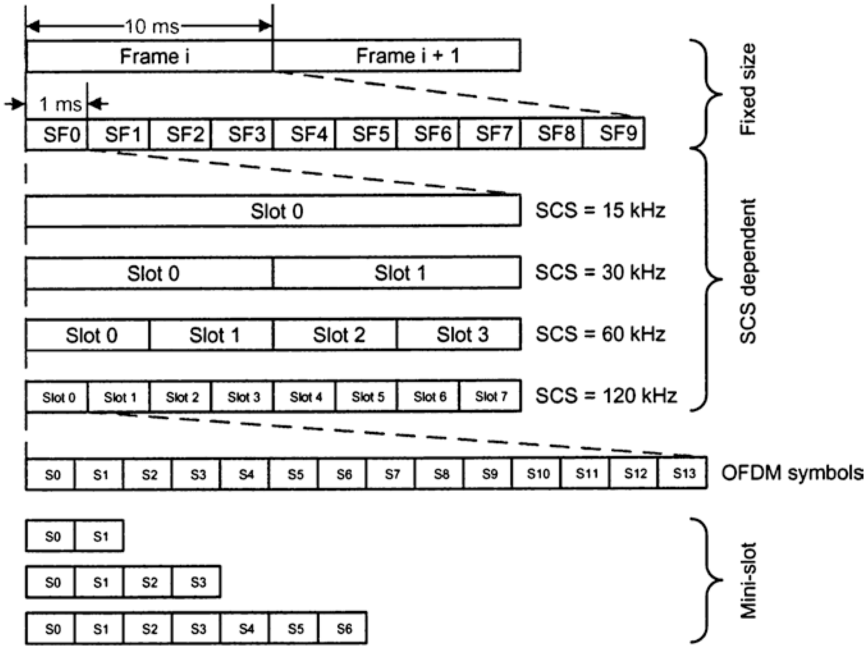

Fig. 10.6 Frame structure and slots. (From [4], with the permission of Elsevier)

slot represents the nominal minimum scheduling interval, referred to as the Transmission Time Interval (TTI), but data can be scheduled to span more than one slot, the latter being referred to as slot aggregation. Recall from Table 9.1 that the OFDM symbol varies with numerology. Thus, the duration of a slot and hence the number of slots per subframe depend on the numerology. This is demonstrated in Fig. 10.6 [4]. The larger the subcarrier spacing, the shorter the slot duration. For subcarrier spacings of 15, 30, 60, and 120 kHz, the associated slot durations are 1, 0.5. 0.25, and 0.125 ms, respectively. In terms of numerology, slot duration is equal to $1/2^{\mu}$ ms. The shorter the slot duration, the lower the latency. This lowered latency is achieved, however, at the expense of a shorter cyclic period which, in a large cell deployment, may not be acceptable due to large rms delay spread. To further aid in latency reduction, reference signals and control signals are located at the beginning of the slot or the set of slots in the case of slot aggregation, as this speeds up the receiver processing.

To allow greater flexibility regarding latency, NR allows using less than 14 symbols for transmission, the resulting smaller effective slots being referred to as *mini-slots*. With the mini-slot structure, 2, 4, or 7 symbols can be allocated with a flexible start position allowing transmission to commence as soon as possible without waiting for the start of a slot boundary. Mini-slots permit shorter latencies even with the 15 kHz subcarrier spacing. There are three good reasons for allowing the use of mini-slots. The first is lowered latency in lower-frequency transmissions where typically 15 or 30 kHz subcarriers are employed. The second is support of analog beamforming in very high-frequency transmission. Here, even though high

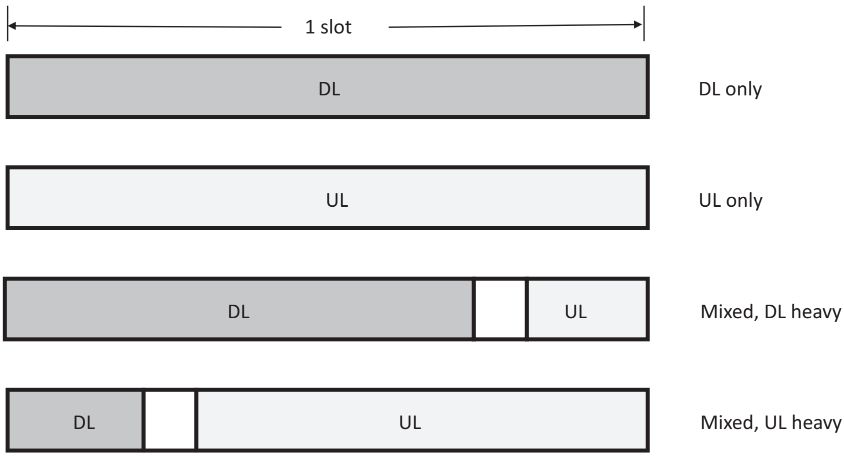

Fig. 10.7 TDD-based frame structure

subcarrier frequencies are used and hence shorter slots, only one beam at a time can be used for transmission, as discussed in Sect. 8.3.7.2, and thus transmission to multiple users must be on a time division basis. Without mini-slots, all 14 symbols would have to be used per UE, before moving on to the next UE. The net effect would be increased multi-transmission latency. With the use of mini-slots, this can be reduced even when transmitting large payloads as with the very large bandwidths available, a few OFDM symbols can be of large data-carrying capacity. The third good reason for employing mini-slots is to facilitate efficient transmission in unlicensed bands. In such bands, the transmitter must constantly monitor the band usage and only transmit when the band is clear. Once it's determined that the band is clear, then the quicker transmission can begin, the better the chance of seizing the spectrum. Mini-slots facilitate this quick action.

In the TDD mode, as shown in Fig. 10.7, a slot can be scheduled for all DL transmission, all UL transmission, or a mixture of both DL and UL, where guard periods are inserted for UL/DL switching.

10.8 The Frequency Domain Structure

In NR, a *resource element*, consisting of one subcarrier during one OFDM symbol, is the smallest physical resource specified. In the frequency domain, 12 consecutive subcarriers of the same spacing are called a *resource block* (RB), the width of a resource block thus being a function of its numerology. The basic scheduling unit is the *physical resource block* (PRB), which is defined as 12 consecutive subcarriers of the same spacing in the frequency domain, i.e., one RB, over one OFDM symbol in the time domain, with all subcarriers within the PRB having the same CP length. Multiple numerologies are supported on the same carrier, and resource block

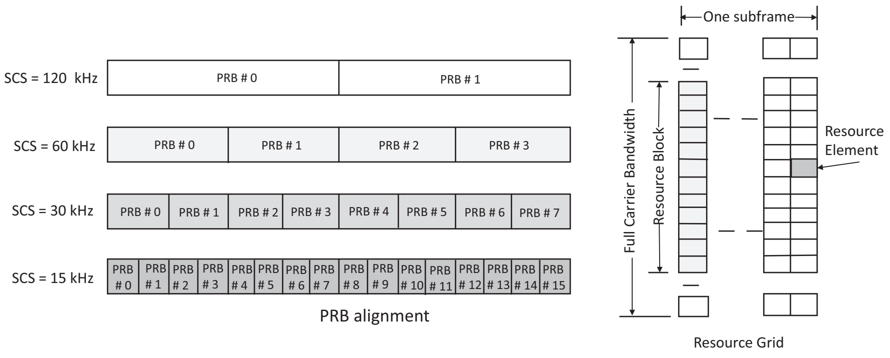

Fig. 10.8 PRB alignment and resource grid

Table 10.1 Minimum and maximum number of resource blocks and corresponding transmission bandwidths

Freq. range	Subcarrier spacing (kHz)	Min (N_{PRB})	Max (N_{PRB})	Minimum bandwidth (MHz)	Maximum bandwidth (MHz)
FR1	15	25	270	4.50	48.60
FR1	30	11	273	3.96	98.20
FR1	60	11	135	7.92	97.20
FR2	60	66	264	47.52	190.08
FR2	120	32	264	46.08	380.16

locations are specified so as to have their boundaries aligned. Thus, as shown in the PRB alignment portion of Fig. 10.8, two PRBs of subcarrier spacing 30 kHz, for example, occupy the identical frequency range as one PRB of 60 kHz subcarriers. Shown in resource grid portion of Fig. 10.8 is a resource element (dark shading) and a PRB (light shading) within a physical resource grid, the latter being described below.

NR specifies minimum and maximum values of RBs (same for DL and UL) per carrier as a function of numerology. Table 10.1 shows these values and their corresponding transmission bandwidths for frequency ranges FR1 and FR2 (Sect. 10.18). For each numerology and carrier, a *resource grid* is defined that covers, in the frequency domain, the full carrier bandwidth being utilized and, in the time domain, one subframe. The resource grids for all subcarrier spacings overlap, and there is one set of resource grids per transmission direction. The resource grid defines the transmitted signal space as seen by the UE for a given subcarrier spacing. However, the UE needs to know where exactly in the available transmitted bandwidth the resource blocks are located. To address this need, NR specifies a common reference point for resource grids referred to as *point A* as well as two classes of resource blocks, namely, *common resource blocks* and *physical resource blocks*. Common resource blocks are numbered from 0 and upward in the frequency domain for a given subcarrier spacing. The center of subcarrier 0 of common resource block 0 for a given subcarrier spacing coincides with "point A." Point A serves as a reference from which the frequency structure can be described and need not be the actual carrier frequency. As a part of the initial access procedure, the location of point A is transmitted to the UE as part the information broadcasted by the PBCH. The physical resource blocks, which indicate the actual transmitted signal spectrum, are then located relative to point A.

10.9 Antenna Ports

In NR, a logical *antenna port* is defined such that the channel over which a symbol on the antenna port is conveyed can be inferred from the channel over which another symbol on the same antenna port is conveyed. In other words, symbols that are transmitted over an antenna port are subject to the same propagation conditions. Thus, the receiving device can assume that two transmitted signals have travelled over the same radio channel and hence experienced the same propagation conditions if and only if they are transmitted from the same antenna port. Demodulation Reference Signals (DM-RSs) are used to help receiving devices estimate the channel that corresponds to the signal transmitted from a specific antenna port. Thus, for the DM-RS associated with the PDSCH, for example, the channel over which a PDSCH symbol on one antenna port is transmitted can be inferred from the channel over which a DM-RS symbol on the same antenna port is transmitted if the two symbols are within the same scheduled PDSCH resource, in the same slot, and undergo the same precoding.

Key to understanding an antenna port is the fact that it doesn't necessarily correspond to a physical antenna. For example, two different signals can be transmitted over several physical antennas via the same pre-antenna paths, as is done with analog beamforming. Nonetheless, the receiver will still see these signals as travelling over the same composite channel and thus emanating from the same antenna port. On the other hand, two different signals transmitted over these same set of antennas but with different precoding for each signal will be seen by the receiver as having been transmitted from two different antenna ports. NR channels/signals and their associated antenna ports are shown in Table 10.2.

10.10 Physical Layer Processing of Transport Channels

The DL transport channels are the DL-SCH, the BCH, and the PCH. The UL ones are the UL-SCH and the RACH. In this section, we examine how transport blocks created by these channels are processed in the physical layer, with the exception of the RACH, which, though defined as a transport channel, does not deliver transport blocks but rather preamble sequences. Time/frequency resource mapping, which can be quite detailed, is only addressed here in the most general terms. For those interested in a more detailed description, it is well covered in [4].

Table 10.2 NR antenna ports

Downlink		Uplink	
Channel/signal	Antenna port starting with	Channel/signal	Antenna port starting with
PDSCH	1000	PUSCH	0
DM-RS for PDSCH	1000	DM-RS for PUSCH	0
PT-RS for PDSCH	1000*	PT-RS for PUSCH	0*
PDCCH	2000	SRS	1000
DM-RS for PDCCH	2000	PUCCH	2000
CSI-RS	3000	DM-RS for PUCCH	2000
TRS	3000	PRACH	4000
PBCH, PSS, SSS	4000		
DM-RS for PBCH	4000		

Note: * signifies lowest number in group

10.10.1 Physical DL Shared Channel (PDSCH) and Physical UL Shared Channel (PUSCH) Processing

Shown in Fig. 10.9 is a block diagram summary of the various processing steps taken on the transmit side by the PDSCH and PUSCH from transport block(s) into the gNB or UE to antenna out (up-conversion is omitted). In the case of the PDSCH, transport blocks are received from the DL-SCH and the PCH. For the PUSCH, transport blocks are received from the UL-SCH. The process shown is the same as that discussed in Chap. 9 and shown in Fig. 9.1 and so requires little additional comment here. Pertinent specifications will, however, be stated. Starting at transmitter input we have:

CRC processor: Receives transport blocks from the MAC layer. For transport blocks larger than 3824 bits, a 24-bit CRC is added. For transport blocks less than or equal to 3824 bits, a 16-bit CRC is used to reduce overhead.

Coding: LDPC coder. See Sect. 6.3.4.

Modulation symbol mapper:

- For the DL with CP-OFDM: QPSK, 16-QAM, 64-QAM, 256-QAM.
- For the UL with CP-OFDM: QPSK, 16-QAM, 64-QAM, 256-QAM.
- For the UL with DFTS-OFDM: π/2 BPSK, QPSK, 16-QAM, 64-QAM, 256-QAM. DFTS-OFDM is optional and used in link budget-limited cases.

Layer mapper:

- For the DL: One coded transport block mapped on up to four layers. If there is a second coded transport block, it is mapped on up to four more layers.
- For the UL: One coded transport block mapped on up to four layers.

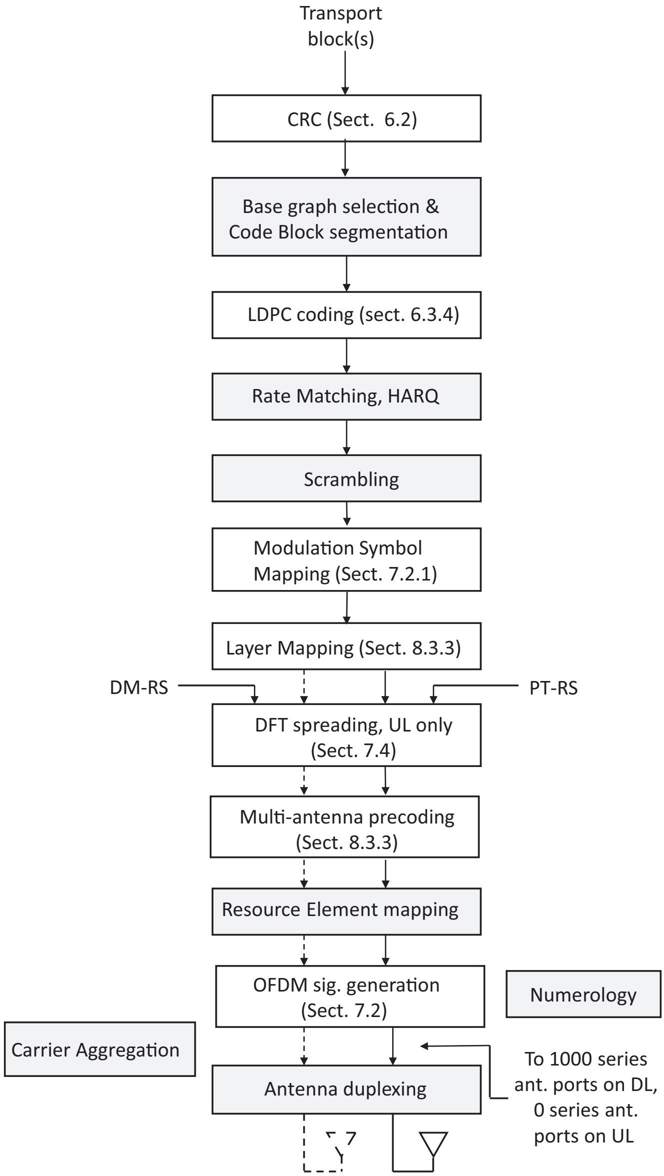

Fig. 10.9 PDSCH and PUSCH processing

OFDM signal generator:

– FFT size: 512 minimum, 4094 maximum.
– Subcarrier spacing: 15 kHz minimum, 120 kHz maximum.

Resource and physical antenna mapper: Broadly speaking, modulation symbols to be transmitted on each antenna port can be mapped to any resource element not specifically designated to other physical channels or signals. Resource elements in turn are mapped to the 1000 series antenna port(s) on the DL and 0 series antenna ports on the UL.

In the receiver, the processing is the inverse of that in the transmitter.

10.10.2 Physical Broadcast Channel (PBCH) Processing

Shown in Fig. 10.10 is a block diagram summary of the various steps taken by the PBCH on the transmit side to process transport blocks received from the BCH containing MIB data. The transport blocks are of size 32 bits and arrive at the PBCH processor at the rate of one every 80 ms. Note that there is no layer mapper in PBCH processing as PBCH uses a single antenna port transmission scheme. It uses the same antenna port as the PSS and the SSS. Starting at the physical layer input, we have:

CRC processor: A 24-bit CRC is added to each received transport block.

Coding: Polar coder. See Sect. 6.3.5.

Rate matcher: The output sequence length is 864 bits.

Modulation symbol mapping: QPSK modulation. The output stream is 432 complex modulation symbols.

Resource and physical antenna mapper: The PBCH together with the PSS and SSS are transmitted together in a block referred to as a *synchronization signal block* or SS block. The SS block is transmitted on a set of defined resource elements as shown in Fig. 10.11. The total number of resource elements used per SS block for

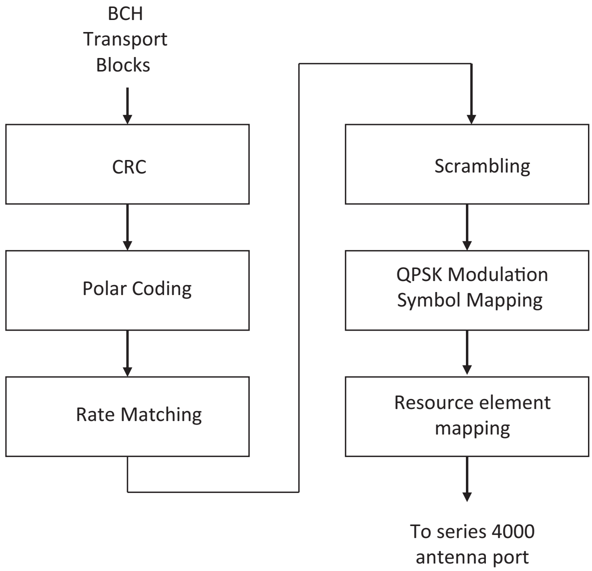

Fig. 10.10 PBCH transmitter processing

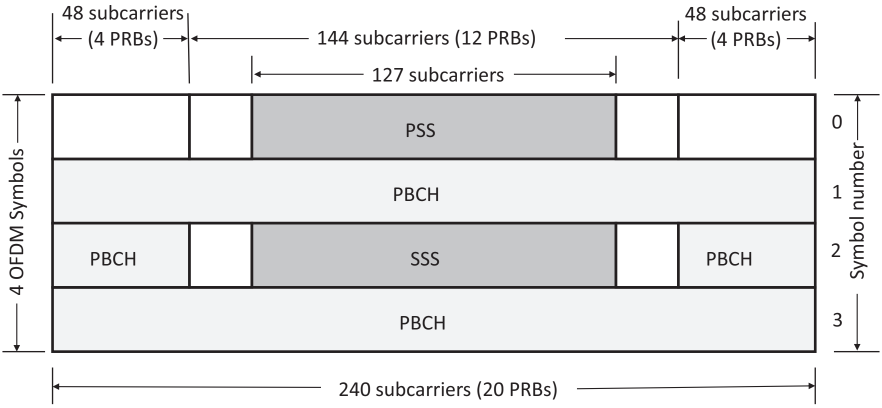

Fig. 10.11 Frequency/time structure of an SS block

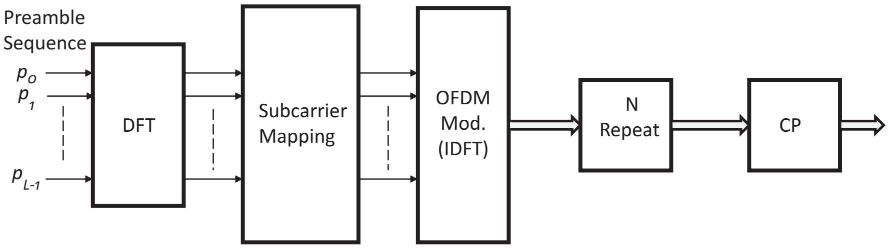

Fig. 10.12 Structure for RACH construction

PBCH transmission is 576 which includes 144 for the Demodulation Reference Signals (DM-RS) required for the coherent demodulation of the PBCH. Resource elements are mapped to the series 4000 antenna port.

Beam sweeping can be applied to SS block transmission. With such beam sweeping, SS blocks are transmitted in different beams in successive times, the set of such blocks within a beam sweep being referred to as an SS burst set.

In the receiver, the processing is the inverse of that in the transmitter as shown in Fig. 10.10.

10.10.3 Physical Random-Access Channel (PRACH) Processing

Shown in Fig. 10.12 is block diagram of the structure used to generate random-access preambles in NR. A preamble is generated by commencing with a length L Zadoff–Chu sequence, such a sequence consisting of complex symbols and exhibiting the property of constant power in both the frequency and time domain. With a prime length Zadoff–Chu sequence of length L, L-1 different sequences can be generated. As shown in Fig. 10.12, the sequence is DFT precoded, mapped to the designated subcarriers, and then applied to an OFDM modulator. Next, the modulator output is repeated N times after which a cyclic prefix is added.

NR specifies two types of preambles, namely, *long preambles* and *short preambles*, based on the length of the preamble sequence.

Long preambles have a sequence length $L = 839$ and a subcarrier spacing of either 1.25 kHz or 5 kHz. Such preambles are used only for operation in the lower frequency bands (FR1). There are four different formats where for each format there is a specified subcarrier spacing, number of repetitions, and cyclic prefix length. The number of repetitions varies between 1 and 4, and the cyclic prefix length varies between 15 and 680 μs.

Short preambles have a sequence length $L = 139$ and a subcarrier spacing of either 15 kHz or 30 kHz for operation in the lower frequency bands (FR1), and a subcarrier spacing of 60 and 120 kHz for operation in the higher frequency bands (FR2). Here, there are nine formats, and the number of repetitions varies between 1 and 12, and the cyclic prefix length varies between 7 and 66.7 μs.

The PRACH preamble configuration to use is provided to the UE in previously received system information.

10.11 Physical Layer Processing of Control Channels

In this section, we examine how control information is processed at the physical layer, namely, by the Physical Downlink Control Channel (PDCCH) and the Physical Uplink Control Channel (PUCCH). As with transport channels covered above, time/frequency resource mapping, which can be quite detailed, is only addressed here in the most general terms. For those interested in a more detailed description, it is well covered in [4].

10.11.1 Physical DL Control Channel (PDCCH) Processing

Shown in Fig. 10.13 is the physical layer processing by the PDCCH. The PDCCH carries *Downlink Control Information* (DCI) from the gNB to the UE. DCI includes information on where in the DL resource assignments have been scheduled; information to enable the UE to properly receive, demodulate, and decode the DL-SCH; and information to the UE about UL scheduling allotments and the specific resources and transport format to use for UL-SCH transmission, power control commands, etc. Starting at the input to the PDCCH processor, we have:

CRC processor: A 24-bit CRC is added to each received transport block. After attachment, the CRC parity bits are scrambled with the corresponding 16-bit *Radio Network Temporary Identifier* (RNTI). The RNTI can contain the identity of an

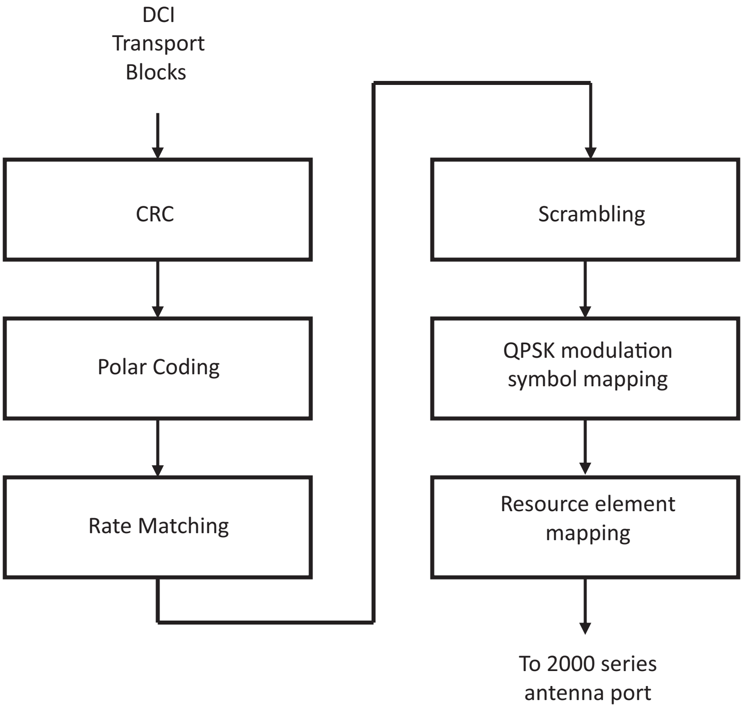

Fig. 10.13 PDCCH processor

intended device, identity of a group of UEs in the case of paging, identity of a group of UEs for which power control is issued, etc.

Coding: Polar coder. See Sect. 6.3.5.

Rate matcher: The rate matching is done per coded block and consists of sub-block interleaving, bit selection, and bit interleaving. It is done to set the number of coded bits to the resources available for PDCCH transmission. In sub-block interleaving, the incoming block is divided into 32 sub-blocks, and these sub-blocks interleaved. Bit selection is the true rate matching. It is achieved by feeding the output of the sub-block interleaver of length N bits into a circular buffer of length N and feeding data out with either no puncturing or puncturing, shortening, or repetition. This selection process is shown in Fig. 10.14. If E coded bits are required for transmission and $E = N$, then no rate matching is applied. If $E < N$, then puncturing or shortening is applied. Puncturing is achieved by selecting from the buffer bits from position $N - E$ to position $N - 1$, while shortening is achieved by selecting bits for transmission from position 0 to position $E - 1$. If $E > N$, then repetition is applied and is achieved by selecting all N bits from the buffer and repeating $E - N$ consecutive bits from the buffer.

Modulation Symbol Mapper: QPSK modulation.

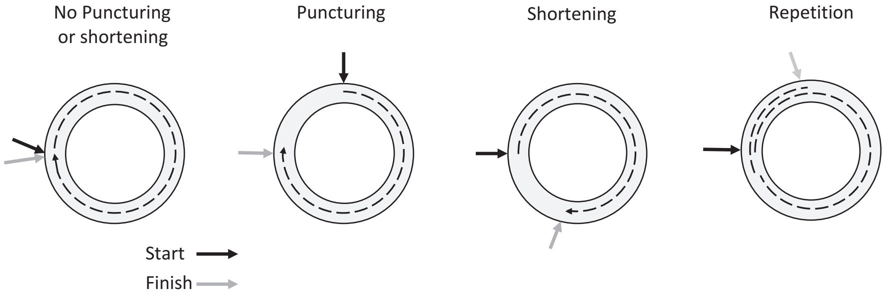

Fig. 10.14 PDCCH rate matching

Resource and physical antenna mapper: The time-frequency resources upon which the PDCCHs are transmitted are referred to as *Control Resource Sets* (CORSETs). In the frequency domain, a CORSET is a set of contiguous or discontinuous physical resource blocks (PRBs), structured in multiples of six PRBs, and can be configured anywhere in the frequency range of the carrier. In the time domain, each is of length one to three OFDM symbols and can be configured at any position within a slot. The PDCCH is transmitted on the 2000 series antenna port.

10.11.2 Physical UL Control Channel (PUCCH) Processing

The PUCCH carries *Uplink Control Information* (UCI) from the UE to the gNB. UCI information includes H-ARQ acknowledgments of received DL-SCH transport blocks, Channel State Information (CSI) regarding the DL channel status, and scheduling requests by the UE to be granted UL resources for UL-SCH transmission. It will be observed on Fig. 10.4 that the UCI is shown directed to both the PUCCH and the PUSCH. This is because NR allows the transmission of the UCI on the PUSCH if the device is transmitting data on the PUSCH. When this is the case, the UCI is multiplexed with the data on the allocated resources. If no data is being transmitted, then the UCI is transmitted on the PUCCH. The UE can beam-form the PUCCH by essentially transmitting it on the same beam as it is used for receiving the corresponding downlink transmission.

Five PUCCH formats are specified in NR. Two formats, Formats 0 and 2, are referred to as Short PUCCH formats. They occupy a maximum of two OFDM symbols that are transmitted in the last one or two symbols of a slot. Format 0 is capable of transmitting 2 or less bits, while Format 2 is capable of transmitting more than 2 bits. They support very fast feedback of HARQ acknowledgments. The other three formats, Formats 1, 3, and 4, are referred to as long PUCCH formats. They occupy from 4 to 14 OFDM symbols and are all capable of transmitting more than 2 bits. The long formats are used when better coverage than that afforded by the short formats is required as the long format results in higher received energy and hence more reliable reception. The following are structures of the various formats.

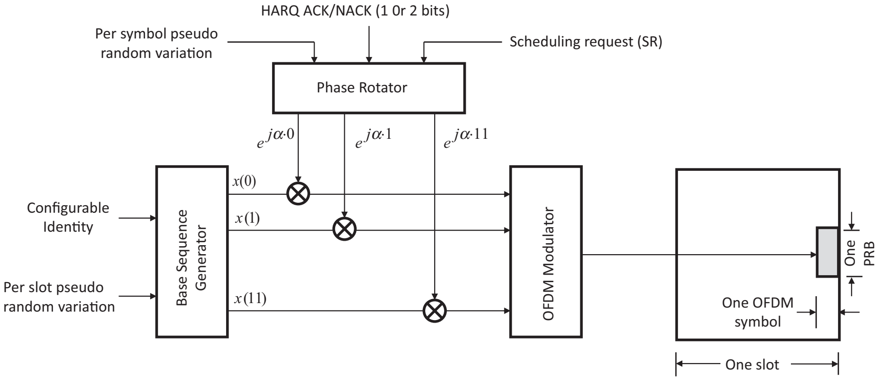

Fig. 10.15 PUCCH Format 0 structure

10.11.2.1 PUCCH Format 0 Structure

Shown in Fig. 10.15 is the PUCCH Format 0 structure. It is used for HARQ acknowledgments and scheduling requests (SRs). Information is transmitted by creating a length 12, low PAPR, Zadoff–Chu base sequence and then applying different linear phase rotations in the frequency domain to the sequence generator output via a phase rotator to create different sequences.

The base sequence used is configured per cell using an identity provided as part of the system information. In addition, it is varied randomly on a per slot basis to randomize interference between different cells.

As shown in Fig. 10.15, one input is HARQ ACK/NACK. If only one DL transport block was transmitting in a TTI (four layers or less), then only one ACK (block successfully decoded) or NACK (block not successfully decoded) needs to be transmitted, i.e., only 1 bit is required. In this case, the phase rotation parameter α is set to either 0 or π. If two DL transport blocks were transmitted, then one of four possible sets of information needs to be transmitted, namely, NACK/NACK, NACK/ACK, ACK/ACK, and ACK/NACK. Thus, 2 bits are required. Here α is set to either 0, π/2, π, or 3π/2.

In the case of a simultaneous scheduling request, the phase rotation α is increased by π/4 for 1-bit acknowledgments. Thus, α is set to either π/4 or 5π/4. For 2-bit acknowledgments, α is increased by π/6. Thus, α is set to either π/6, 3π/2, 7π/6, or 5π/3.

Just as the base sequence can be varied randomly per slot, so too can the phase rotation be varied per symbol.

Also shown in Fig. 10.15 is the structure of the Format 0 PUCCH in the time/frequency resource. In the time domain, it is shown occupying the last symbol but can also occupy the last two symbols. If two symbols are used, the same information is transmitted on both symbols. In the frequency domain, it consists of 12 subcarriers and hence occupies one or two PRBs.

We note that the PUCCH Format 0 is not accompanied by a DM-RS.

10.11.2.2 PUCCH Format 1 Structure

PUCCH Format 1 is, to a certain extent, the long format version of Format 0. It can carry 2 bits using 4 to 14 OFDM symbols, each symbol occupying one resource block in the frequency domain. It is used for HARQ acknowledgments, scheduling requests (SRs), or both. The OFDM symbols are split sequentially between those used for control data and those used as reference signals to facilitate coherent detection. The split between those used for control and those used for reference is typically close to even. Shown in Fig. 10.16 [4] is the PUCCH Format 1 structure. If the UCI data is a single bit, it is BPSK modulated; if it is 2 bits, then it is QPSK modulated, resulting in both cases in a complex valued symbol, $d(0)$. Per modulated symbol out, as with Format 0, a cyclic shift is applied to create a pseudo-random variation. This cyclic shifted variation is multiplied by a 12-bit sequence, of the same type used in the Format 0 creation, resulting in a length 12 sequence of complex valued symbols. Also, as with Format 0, this sequence is configured per cell using an identity provided as part of the system information and varied randomly on a per slot basis to randomize interference between different cells. The complex valued sequence is block-wise spread in the time domain with an orthogonal sequence of length equal to the number of OFDM symbols used for control information. In the example shown in Fig. 10.16, four symbols are used for control information; hence, a length 4 orthogonal sequence is used. By using different orthogonal codes, multiple UEs having the same base sequence and phase rotation and using the same

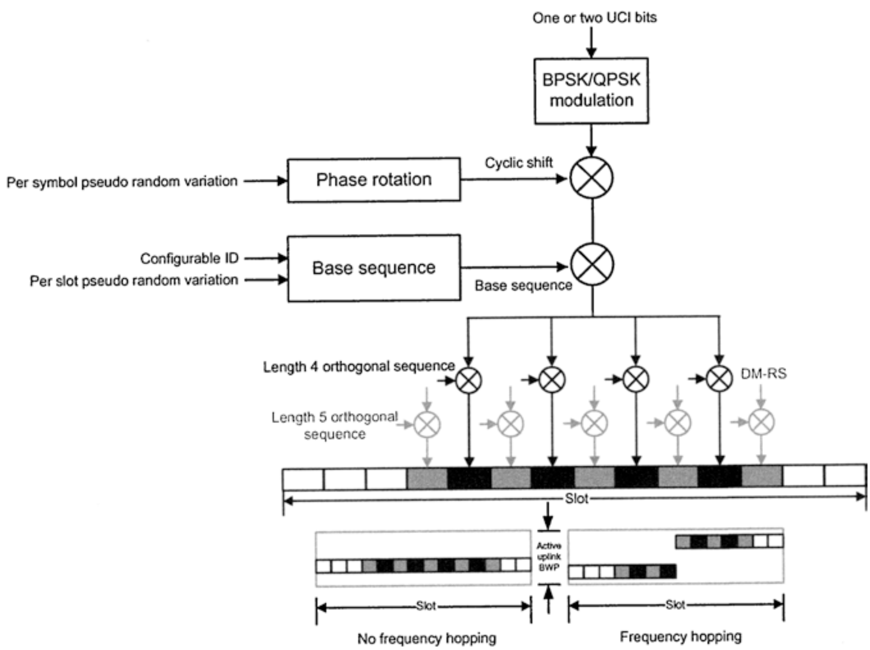

Fig. 10.16 PUCCH Format 1 structure. (From [4], with the permission of Elsevier)

resource can still be differentiated at the gNB, thus increasing the multiplexing capacity. For a PUCCH transmission spanning multiple slots, the complex valued symbol $d(0)$ is repeatedly used in creating the length 12 sequences for those slots.

PUCCH Format 1 can use frequency hopping to achieve a certain degree of frequency diversity, an example of such frequency hopping being shown in Fig. 10.16. The use of frequency hopping is configurable and determined by the gNB as part of the PUCCH resource configuration.

10.11.2.3 PUCCH Format 2 Structure

As indicated above, Format 2 is a short format PUCCH. It occupies a maximum of two OFDM symbols that are transmitted in the last one or two symbols of a slot. It is capable of transmitting more than 2 bits, carrying, for example, simultaneous HARQ acknowledgments and CSI reports, as well as a SR. If the payload is larger than can be accommodated, then HARQ acknowledgments are given priority over CSI reports. Shown in Fig. 10.17 is the PUCCH Format 2 structure. Let the payload size be N_{UCI}. Then, starting at the UCI input point, we have:

CRC processor: If $N_{UCI} \leq 11$, no CRC bits are attached. If $12 \leq N_{UCI} \leq 19$, 6 CRC bits are attached. If $N_{UCI} \geq 20$, 11 CRC bits are attached.

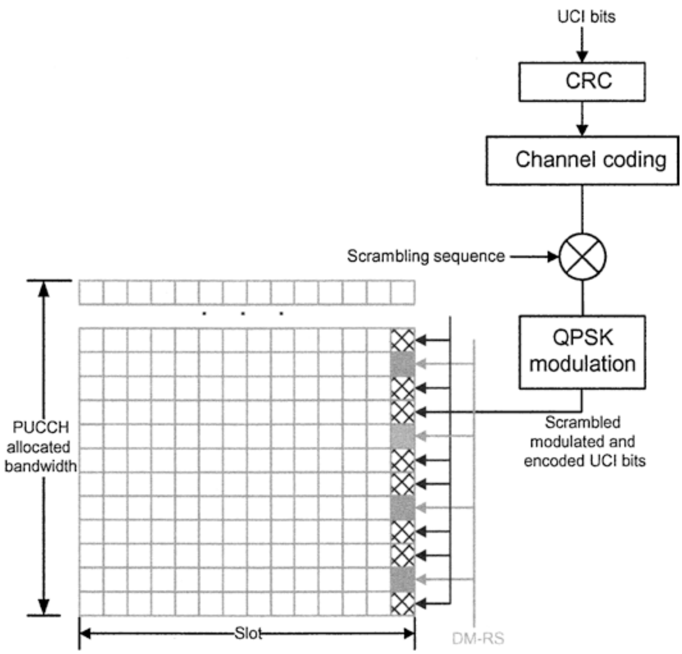

Fig. 10.17 PUCCH Format 2 structure. (From [4], with the permission of Elsevier)

Channel coder: $N_{UCI} \geq 12$, the UCI bits with CRC attachment are encoded with polar coding (Sect. 6.3.5). If $3 \leq N_{UCI} \leq 11$, then the UCI bits are encoded with Reed–Muller coding (Sect. 6.3.6.3). If $N_{UCI} = 2$, simplex coding is used (Sect. 6.3.6.2). If $N_{UCI} = 1$, repetition coding is used (Sect. 6.3.6.1).

Scrambler: The scrambling sequence is a function of both the UE identity and the physical layer cell identity.
Modulation symbol mapping: QPSK
Resource element mapping: The modulated symbols are sequentially mapped to subcarriers reserved for PUCCH transmission and not used by the associated DM-RSs. They are mapped across multiple resource blocks utilizing one or two OFDM symbols. The associated DM-RSs are mapped to every third subcarrier and consist of pseudo-random generated QPSK symbols. PUCCH Format 2 is normally transmitted at the end of a slot as shown in Fig. 10.17 [4], but depending on certain conditions, they may be transmitted in other positions within a slot.

10.11.2.4 PUCCH Format 3 Structure

PUCCH Format 3 is the long version of PUCCH Format 2. Like Format 2, it can transmit more than 2 bits, but here, these can be transmitted over 4 to 14 symbols using multiple resource blocks per symbol. The result is that this is the format with the largest payload capacity. Just as with Format 1, OFDM symbols used are split between those used for control data and those used as reference signals to facilitate coherent detection. Shown in Fig. 10.18 is the PUCCH Format 3 structure. As before, let the payload size be N_{UCI}. Then, starting at the UCI input point, we have:

CRC processor: Same as for Format 2.
Channel coder: Same as for Format 2.
Scrambler: Same as for Format 2.
Modulation symbol mapping: QPSK, with option to use π/2 BPSK.
Resource element mapping: Modulated symbols are divided between the OFDM symbols with $12 \times M_{RB}$ modulated symbols directed to each OFDM symbol in M_{RB} resource blocks, where M_{RB} denotes the number of resource blocks occupied by the PUCCH. These sets of modulated symbols are DFT spread to lower the resulting PAPR prior to OFDM modulation which results in its placement in the resource grid. Operation is configurable with frequency hopping as shown in Fig. 10.18 [4] to exploit frequency diversity, but also configurable without. Reference signal symbol placement is based on whether or not frequency hopping is used and also on the length of the PUCCH transmission as at least one reference signal must be utilized per hop.

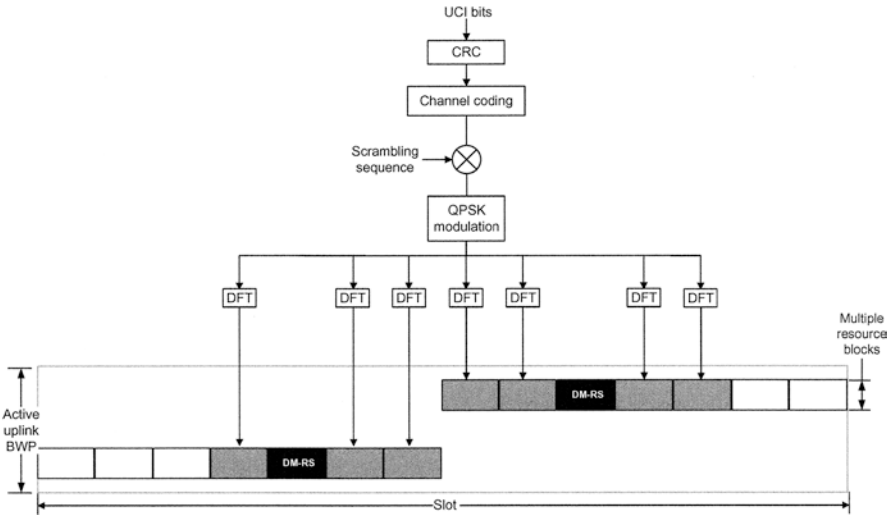

Fig. 10.18 PUCCH Format 3 structure. (From [4], with the permission of Elsevier)

10.11.2.5 PUCCH Format 4 Structure

PUCCH Format 4 is identical to Format 3 up to the point of modulation output. The difference, as shown in Fig. 10.19 [4], is in the modulated symbol mapping. Here, the modulated symbols are divided between (a) the OFDM symbols with $12/N_{SF}$ modulated symbols directed to each OFDM symbol, (b) each $12/N_{SF}$ set spread by a length N_{SF} block spreading sequence ($N_{SF} = 2$ or 4) to create a length 12 modulated symbol output, and (c) the length 12 symbol output DFT spread and OFDM modulated resulting in its placement in the resource grid. Frequency hopping and reference signal placement are similar to Format 3.

10.12 Physical Signal Processing

Physical signal processing is now addressed. Such signals are predefined, contain no information from higher layers, and occupy specified resource elements in the physical resource grid. NR physical signals are generated via three types of sequences, namely:

– *Maximum length* (m) sequences, which are pseudo-random binary sequences. They are generated using a shift register with linear feedback. A length n register produces sequences of length $2^n - 1$. It is possible to select a small subset of sequences created by the same shift register that have relatively low cross correlation.

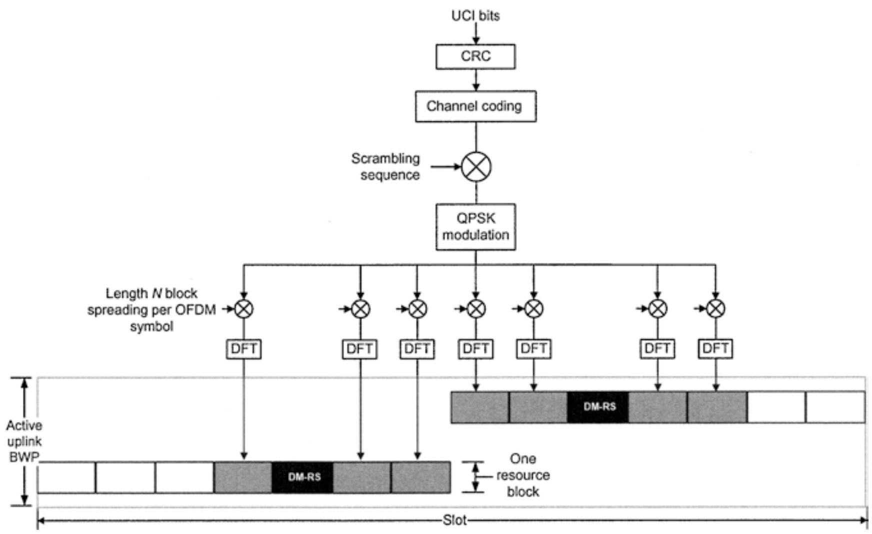

Fig. 10.19 PUCCH Format 4 structure. (From [4], with the permission of Elsevier)

– *Gold sequences*, which are also a type of pseudo-random binary sequences, constructed by adding together in a defined fashion two *m* sequences of the same length. A set of Gold sequences consists of $2^n - 1$ sequences, each one of length $2^n - 1$. Every sequence within a set has a low cross correlation with the other sequences in the set, making it easy in decoding to distinguish between the individual sequences even when corrupted by noise. If properly constructed, gold sequences have better cross-correlation properties than *m* sequences.
– *Zadoff–Chu sequences* (already mentioned above), which are complex value sequences that when used to linearly modulate a signal give rise to a signal of constant amplitude. Further, the DFT of such a sequence is also a Zadoff–Chu sequence. Such sequences exhibit the very useful property that cyclically shifted versions of themselves appear as orthogonal to one another at a receiver, given that each cyclic shift, as seen within the time domain of the signal, is greater than the propagation delay plus multipath delay spread of the signal between the transmitter and receiver.

As indicated above, NR physical signals are Demodulation Reference Signals (DM-RSs), Phase-Tracking Reference signals (PT-RSs), Channel State Information Reference Signals (CSI-RSs), and synchronization signals (SSs). Following is a description of the processing applied to create physical signals. Time/frequency resource mapping, which can be quite detailed, is only addressed here in the most general terms. For those interested in more detailed descriptions, they are well covered in [2, 4]. First DL physical signals are described followed by UL signals.

10.12.1 Downlink Physical Signals

10.12.1.1 Demodulation Reference Signal (DM-RS) for PDSCH

The DM-RS for PDSCH is transmitted together with the PDSCH on Series 1000 antenna ports and subject to the same precoding as the PDSCH. Its point of injection in the PDSCH processing is shown in Fig. 10.9. It is created via the generation of pseudo-random sequences. The generation process is as shown conceptually in Fig. 10.20. First, two slightly different length 2^{31}–1 Gold sequences are created. Both unipolar sequences (0s and 1s) are then converted to bipolar format (-1s and +1s). One of the sequences is then rotated through 90^0 so as to be in quadrature with the other, and both sequences are then summed to create complex QPSK modulation symbols. As with PDSCH modulation symbols, these symbols undergo multi-antenna precoding, resource element mapping, OFDM signal generation, and antenna duplexing. The QPSK sequence is generated across all the resource blocks but transmitted only in the resource blocks used for data transmission as there is no need for having knowledge of the channel outside of the frequency spectrum occupied by the channel.

In NR, to help achieve low latency, a front-loaded structure is used whereby the DM-RSs are located early in the transmission. There are two main mapping structures in the time domain, namely, mapping type A and mapping type B. With type A, the first DM-RS is located in symbol 2 or 3 of the slot and is used in cases where the data uses most of the slot. With type B, the first DM-RS is located in the first symbol where data is allocated. To assist high-speed situations, up to three additional DM-RSs can be configured in a slot for both types A and B.

Two different types of DM-RSs can be configured, type 1 and type 2. With both types, the underlying pseudo-random sequence is mapped to every second subcarrier in the OFDM symbol used for DM-RS transmission. To support multilayer MIMO transmission, multiple orthogonal signals, each directed to a different series 1000 antenna port, can be created by multiplying the underlying sequence with different length 2 orthogonal sequences in the frequency domain. With type 1, up to four orthogonal reference signals can be provided with a single-symbol DM-RS and up to eight with a double-symbol DM-RS. With type 2, up to 6 orthogonal reference signals can be provided with a single-symbol DM-RS and up to 12 with a double-symbol DM-RS, thus supporting a higher degree of multiuser MIMO than with type 1.

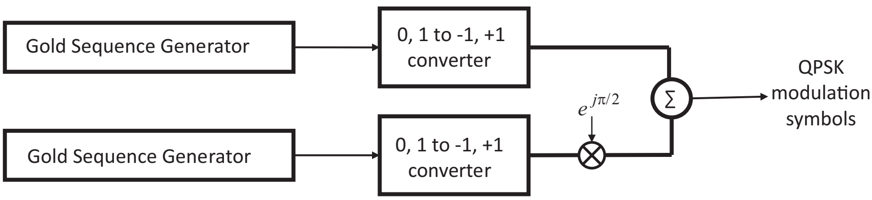

Fig. 10.20 PDSCH-associated DM-RS generation

10.12.1.2 DM-RS for PDCCH

The DM-RS for the PDCCH is transmitted together with the PDCCH. It is generated in a similar fashion to the DM-RS for PDSCH where a pair of different length 2^{31}–1 Gold sequences is used to derive a QPSK modulated symbol stream. The QPSK sequence is generated across all the resource blocks but transmitted only in the resource blocks used for PDCCH transmission, being mapped onto every fourth subcarrier in an occupied resource group, and directed to the single antenna port used for PDCCH transmission, namely, antenna port 2000.

10.12.1.3 DM-RS for PBCH

The DM-RS for the PBCH is transmitted together with the PBCH. It is generated in a similar fashion to the DM-RS for PDSCH where a pair of different length 2^{31}–1 Gold sequences is used to derive a QPSK modulated symbol stream. The QPSK sequence is generated across all the resource blocks but transmitted only in the resource blocks used for PBCH transmission, being mapped onto every fourth subcarrier in an occupied resource group, and directed to the single antenna port used for PBCH transmission, namely, antenna port 4000.

10.12.1.4 Downlink Phase-Tracking Reference Signal (PT-RS)

The DL Phase-Tracking Reference Signal (PT-RS), like the DL DM-RS for the PDSCH, is transmitted together with the PDSCH, its point of injection in the PDSCH processing being as shown in Fig. 10.9. It is present only if it is explicitly configured. It may not be configured, for example, in low-frequency transmission where common phase error may be negligible. It is generated in a similar fashion to the DM-RS for PDSCH where a pair of different length 2^{31}–1 Gold sequences is used to derive a QPSK modulated symbol stream. The QPSK sequence is generated across all the resource blocks but transmitted only in the resource blocks used for PDSCH transmission. Its mapping is designed to have low density in the frequency domain and high density in the time domain. This is because the phase rotation caused by common phase error is the same for all subcarriers in an OFDM symbol, whereas it has low correlation across OFDM symbols.

In the frequency domain, the PT-RS is transmitted in every second or fourth resource block. In the time domain, the first PT-RS is repeated every first, second, or fourth symbol starting with the first PT-RS symbol in the allocation.

The antenna port used for PT-RS transmission is the lowest numbered 1000 series port used in the DM-RS antenna port group.

10.12.1.5 Channel State Information Reference Signal (CSI-RS)

The CSI-RS is generated in a similar fashion to the DM-RS for PDSCH where a pair of different length 2^{31}–1 Gold sequences is used to derive a QPSK modulation symbol stream. The QPSK sequence is generated across all the resource blocks but transmitted only in the resource blocks used for PDSCH transmission.

A configured CSI-RS can correspond to up to 32 antenna ports and is transmitted on the Series 3000 antenna ports. In the time domain, it may commence at any OFDM symbol of a slot and extend over 1, 2, or 4 OFDM symbols depending on the number of antenna ports configured. It is always configured on a per UE basis. However, this does not necessarily mean that a configured CSI-RS can only be used by a single UE, but rather that the same set of CSI-RS resources can be configured separately for several devices. A single-port CSI-RS occupies a single resource element within one resource block in the frequency domain and one slot in the time domain and can be configured to occupy any position within the resource block that doesn't collide with other DL physical channels and signals.

A multi-port CSI-RS can be regarded as multiple CSI-RSs, orthogonal to each other per antenna port, sharing the same set of resource elements specified for transmission of the configured multi-port CSI-RS. This resource sharing is accomplished, in general, via a combination of:

– Code-domain sharing, where different per antenna port CSI-RS are transmitted using the same set of resource elements, differentiation achieved by spreading each CSI-RS with different orthogonal codes
– Frequency-domain sharing, where different antenna port CSI-RS are transmitted on different subcarriers over an OFDM symbol
– Time-domain sharing, where different antenna port CSI-RS are transmitted on different OFDM symbols within a slot

Unlike the DM-RS, the CSI-RS does not undergo multi-antenna precoding that's applied to user data. Rather, the CSI-RS ports are mapped directly to either the gNBs physical antennas or via a spatial filter that maps M CSI-RS ports to N physical antennas. When such a filter is used, it is seen by the UE as an integrated part of the overall channel, and thus the channel being sounded is not the actual, physical channel. The mapping via the spatial filter allows different CSI-RSs to be beamformed in different directions.

10.12.1.6 Tracking Reference Signal (TRS)

The TRS is not a CSI-RS per say. Rather, it is a resource set consisting of multiple periodic NZP CSI-RSs and transmitted on the Series 3000 antenna ports. As shown in Fig. 10.21 [2], the resource set is configured within one resource block in the frequency domain and two consecutive slots in the time domain. The set consists of four single-port CSI-RSs, each occupying three resource elements in an OFDM symbol, with the separation between two CSI-RSs within a slot always being four

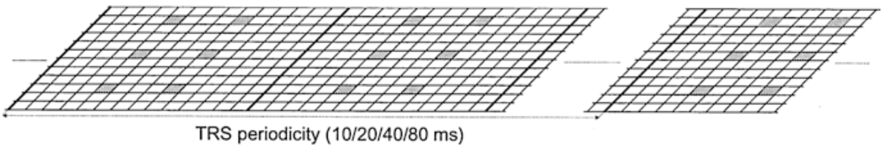

Fig. 10.21 TRS resource plane structure. (From [2], with the permission of Elsevier)

symbols. The set can be configured with a periodicity of 10, 20, 40, or 80 ms. An alternative structure is defined with the same per slot structure as shown Fig. 10.21 but with the set consisting of two CSI-RSs in only one slot.

10.12.1.7 Primary Synchronization Signal (PSS)

The PSS is generated from a length 127 m sequence shift register. By applying different shifts to the basic *m*-sequence, three different PSSs are generated. The 0s and 1s outputted from the sequence generating shift register are used to create BPSK modulation symbols of value -1 and +1. These symbols are fed to the OFDM processor which resource maps them to 127 subcarriers as part of an SS block and as shown in Fig. 10.11. They are mapped to antenna port 4000 along with the PBCH.

10.12.1.8 Secondary Synchronization Signal (SSS)

The SSS is generated from a length 127 Gold sequence generator. A total of 336 different sequences are defined. The 0s and 1s outputted from the sequence generator are used to create BPSK modulation symbols of value -1 and +1. As with the PSS, these symbols are fed to the OFDM processor which resource maps them to 127 subcarriers as part of an SS block and as shown in Fig. 10.11. They are mapped to antenna port 4000 along with the PBCH.

As the PSS can be one of three different sequences and the SSS one of 336 different sequences, together they are able to convey 1008 (3 × 336) unique physical layer cell identities.

10.12.2 Uplink Physical Signals

10.12.2.1 DM-RS for CP-OFDM PUSCH

In NR, the same DM-RS as used in the DL, i.e., Gold sequence-derived QPSK modulation symbols, is used in the UL for the CP-OFDM case. Thus, the description given in Sect. 10.12.1.1 above is equally applicable here if we substitute PUSCH for PDCSH and antenna port series starting with 0 for that starting with 1000.

10.12.2.2 DM-RS for DFTS-OFDM PUSCH

DFTS-OFDM is configured for single-layer transmission only and is designed to be used primarily in situations where coverage is challenging. Here, the DM-RS is based on Zadoff–Chu sequences and supports continuous allocations. Multiple sequences are generated from a single-base sequence by applying different linear phase shifts in the frequency domain as shown in Fig. 10.22. Defined OFDM symbols within a slot are assigned exclusively for DM-RS transmission, following the same mapping as configuration of type 1 defined in Sect. 10.12.1.1 for the DL DM-RS. Type 2 mapping is not supported as there is no requirement for handling a high degree of multi-used MIMO.

10.12.2.3 DM-RS for PUCCH Format 1

The DM-RS for PUCCH Format 1 is created from a length 12 unmodulated Zadoff–Chu sequence. It is inserted in the time domain as shown in Fig. 10.16, the sequence being block-wise spread in the time domain with an orthogonal sequence of length equal to the number of OFDM symbols used for reference signals. In Fig. 10.16, five OFDM symbols are used for reference signals, hence the length 5 orthogonal sequence used for block-wise spreading. Resource mapping is to antenna port 2000 series.

10.12.2.4 DM-RS for PUCCH Format 2

The DM-RS for PUCCH Format 2 has the same structure as that used for the PDSCH and is thus gold sequence-derived QPSK modulation symbols. These symbols are mapped to every third subcarrier in each OFDM symbol as shown in Fig. 10.17. Resource mapping is to antenna port 2000 series.

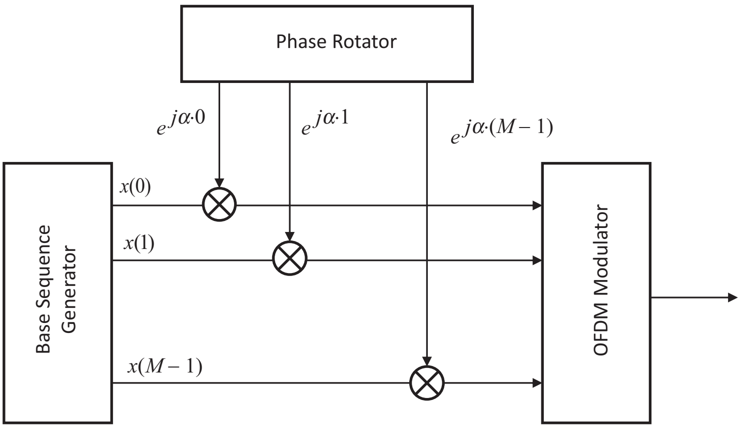

Fig. 10.22 Generation of DFTS-OFDM-associated UL DM-RS

10.12.2.5 DM-RS for PUCCH Formats 3 and 4

The DM-RS for PUCCH Formats 3 and 4 is generated in the same way as for DFTS-OFDM PUSCH transmissions described in Sect. 10.12.2.2. The placement of reference signal symbols depends on whether or not frequency hopping is used as well as the length of the PUCCH transmission given that there must be at least one reference signal per frequency hop. Resource mapping is to antenna port 2000 series.

10.12.2.6 Uplink PT-RS for CP-OFDM

In NR, the same PT-RS structure as used in the DL is used in the UL for the CP-OFDM case. Thus, the description given in Sect. 10.12.1.4 above is equally applicable here if we substitute PUSCH for PDCSH and antenna port series starting with 0 for that starting with 1000.

10.12.2.7 Uplink PT-RS for DFTS-OFDM

The UL PT-RS for DFTS-OFDM is generated via Gold sequences to create a QPSK modulation symbol stream. These symbols are inserted prior to DFT precoding, and the time domain mapping is the same as the CP-OFDM case.

10.12.2.8 Sounding Reference Signal (SRS)

The SRS is configured to allow the gNB to estimate the UL channel in much the same way that the CSI-RS is configured to allow the UE to estimate the DL channel. The SRS is designed to have low PAPR, this being achieved by employing sequences partly based on Zadoff–Chu sequences. It can, in general, cover one, two, or four consecutive OFDM symbols, being located within the last six symbols of a slot. In the frequency domain, it has a comb-like structure, where it is transmitted on every second or fourth subcarrier, this being referred to as comb-2 or comb-4, respectively. Transmission from different UEs can take place within the same frequency range by each using a comb pattern that corresponds to different subcarriers than those used by the rest. With comb-2, two SRSs can be so accommodated, while for comb-4, up to four SRSs can be accommodated.

When an SRS supports more than one antenna port, the different ports utilize the same basic SRS sequence and the same set of resource elements. However, the signals transmitted from different ports are differentiated from each other by applying different phase rotations to the symbols on the different ports.

Up to four Series 1000 antenna ports are supported, being configurable to either 1, 2, or 4 ports. Unlike the DM-RS, the SRS does not undergo multi-antenna precoding that's applied to user data. Rather, similar to CSI-RS, the SRS ports are either mapped directly to the UEs physical antennas or via some spatial filter that maps *M* SRS ports to *N* physical antennas. This mapping via the spatial filter allows beamforming of the transmitted signal. When such a filter is used, it is seen by the gNB as an integrated part of the overall channel.

10.13 Initial Access

Initial access are the steps executed to allow a UE that is just powered on to find a cell to camp on and, having found such a cell, the steps that the UE, in the idle or inactive RRC state, uses to access the network and, via a random-access procedure, establishes normal communication, i.e., go to the RRC-CONNECTED state [2]. A random-access procedure can be contention-based or contention-free, the latter used only if the UE is already known to the network and has been allocated preambles. Shown in Fig. 10.23 and described below is a simplified version of the initial access steps where the random access is contention-based:

1. The gNB periodically transmits SS blocks on the PBCH and SIBs on the PDSCH.
2. UE performs initial cell selection and DL synchronization. Initial cell selection involves finding a strong PBCH signal (if the PBCH is being beam swept, then this means finding the strongest beam). It does this by scanning all RF channels within its capability. On each carrier frequency, it searches for the strongest cell. Once a suitable cell is found, this cell is selected. It uses the PBCH signal of this cell for establishing a connection, obtaining an estimate of frame timing, obtaining cell identification, and finding the PSS and SSS necessary for the coherent demodulation of the PCBH and PDCCH. First the UE tries to find and demodulate the PSS so as to obtain symbol and half-frame timing. If successful, it then

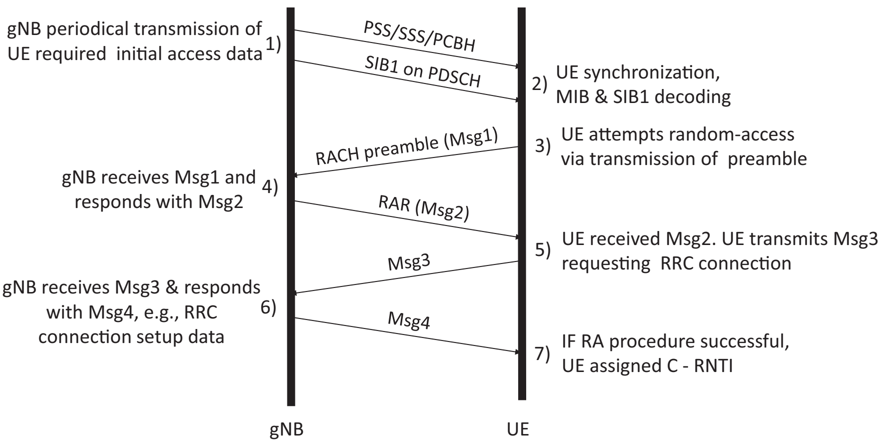

Fig. 10.23 Initial access procedure

attempts to demodulate the SSS so as to determine the cyclic prefix length, the duplexing scheme, and the exact frame timing. It then detects the physical cell identity from the sequences used on the PSS and SSS and decodes the PBCH to obtain the MIB. With the aid of the MIB, it is able to receive the PDSCH and PDCCH. It decodes the PDSCH and obtains SIB1 which allows it to commence access to the system.
3. UE attempts random access by transmitting a preamble (Msg1) on the PRACH channel. If no response is received from the network within a defined window, the UE assumes that the preamble was not correctly received and resends it at a higher transmit power level and continues this process until a response is received or a maximum number of tries is reached.
4. The gNB receives Msg1 and responds with Msg2, a Random-Access Response (RAR) via the PDCCH/PDSCH. Included in the response is a scheduling grant indicating resources the UE can use for the transmission of its next message, Msg3, a timing correction determined by the network based on the timing of the received preamble (Sect. 10.14), and a temporary identity, the Temporary Cell Radio Network Temporary Identifier (TC-RNTI). Had the network received random-access attempts from several devices, the individual response messages are combined in a single transmission. Had all the devices used different preambles, then resources allocated for upcoming uplink transmission would be different, and no collision would occur. However, should multiple devices have used the same preamble, then a collision or collisions would occur. The following steps resolve this collision(s) dilemma.
5. The UE receives Msg2 and adjusts its uplink transmission timing. For the UE to be able to transmit user data, i.e., become RRC-CONNECTED, it needs to be assigned a unique identity within the cell, the C-RNTI. The UE transmits the necessary information to make this possible, including a UE identity, over the UL-SCH, Msg3, in the resources assigned in the DL Msg2.
6. The gNB receives Msg3 and responds with a contention resolution message, Msg4, addressed using the TC-RNTI and intended to ensure that a UE does not incorporate another UE identity and, if so, become RRC-CONNECTED.
7. The UE compares the identity in the message with the identity transmitted in Msg3. If these identities match, the random-access procedure is declared successful, the TC-RNTI is redefined as the C-RNTI, and the UE becomes RRC-CONNECTED.

10.14 Uplink Timing Correction

As indicated in Sect. 10.13 above, the gNB, during the random-access procedure, sends a timing correction to the UE determined by the gNB based on the timing of the received preamble. This timing feedback by the gNB is then maintained during the connection as frequently as needed. The purpose of timing alignment is to assure as best as possible that the uplink slot boundaries for a given numerology from

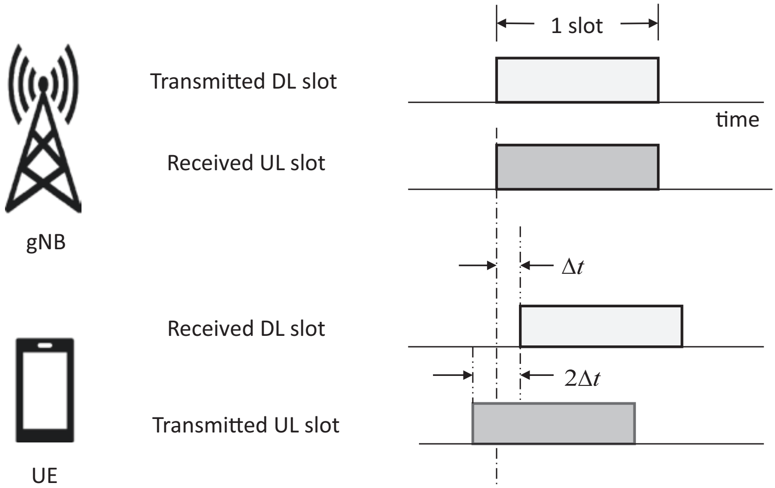

Fig. 10.24 Uplink timing advance

different UEs align at the gNB in order to guarantee uplink orthogonality. This is assured if any offset after adjustment falls within the cyclic prefix. Following initial access, the gNB can use the sounding reference signals to estimate the required offset but in principle can use any signal received from the UE. If a timing correction is required, the gNB transmits to the UE a timing advance command as a MAC control element on the DL-SCH.

As shown in Fig. 10.24, timing advance is a negative time offset, at the UE, to the start of an uplink slot relative to the start of the most recently downlink slot as observed by the UE. As shown in the figure, the downlink slot, when received by the UE, suffers a propagation delay of Δt relative to its start time at the gNB. Given that the uplink slot will suffer the same propagation delay, then, in order for it to arrive at the gNB time aligned, it needs to be offset relative to the received time of the downlink slot by $2\Delta t$. Clearly timing advance is a function of propagation delay and hence the distance between the gNB and the UE. The timing advance is conveyed in incremental steps, the individual step being a function of the subcarrier spacing. The larger the subcarrier spacing, the shorter the slot, and the smaller the timing advance step.

10.15 Uplink Power Control

The primary purpose of uplink power control is to minimize interference to cells other than the UE's cell. Interference within the UE's cell is a lesser issue given that transmissions within the same cell are ideally orthogonal. It also serves the purpose of minimizing UE power consumption. Uplink power control is the procedure whereby the transmit power at the UE of different physical channels and signals is adjusted so as to be received by its target gNB at a level necessary, but not excessively above, for proper decoding while at the same time being such that they are received throughout the surrounding cells at levels that don't result in undue interference. By a received level necessary for proper decoding, we mean a level resulting in the signal to noise necessary to meet the required BER. This level is a function of modulation complexity, higher complexity requiring a higher signal-to-noise ratio for the same BER. In achieving the purposes stated above, power control in a mobile environment has to continually react to the varying characteristics of the propagation channel, including distance-induced path loss, shadowing, fast fading, and interference from other users both within the cell and neighboring cells.

In NR, there are two types of uplink power control operations, open-loop and closed-loop:

– Open-loop power control, where the power is set by the UE based on the downlink received signal strength. This form of control provides an initial power setting for the UE prior to it establishing an RRC connection with the gNB, after which closed-loop operation can commence.
– Closed-loop power control, where the power is set based on explicit signaling from the gNB. The power commands contained in the signaling are based on the gNB measurements of the received uplink power, hence the "closed-loop" nomenclature. This form of power control provides finer granularity and is normally more effective than open-loop control, particularly given that it is based on actual measured uplink loss, not an estimate based on the assumption that uplink loss is similar to the downlink loss, an assumption that may not necessarily be correct in the case of FDD.

10.16 Low Latency

A key feature of NR is its ability to be configured for low latency [2, 3]. This feature is of particular importance in ultra-reliable low-latency communication (URLLC) service, an NR service category to support latency-sensitive services such as remote control and autonomous driving. In 3GPP TR 38.913, user plane latency is defined as "the time to successfully deliver an application layer packet/message from the radio protocol layer 2/3 SDU ingress point to the radio protocol layer 2/3 SDU egress point via the radio interface in both uplink and downlink directions, where neither device nor the Base Station reception is restricted by DRX." In simpler language, this means the time from when IP packets enter the RAN to when they leave, in the UL and DL directions, given no discontinuous reception. Physical layer latency results from the addition of many contributing components:

– The processing time at the transmit end to get ready for transmission, including encoding, etc.
– Average buffering time in the transmission buffer until the next TTI starts. This time is half the TTI time

Table 10.3 NR DL and UL user plane latency

Delay component	Downlink latency (ms)	Uplink latency (ms)
TTI	0.125	0.125
Frame alignment	0.063	0.063
Transmit side processing	0.160	0.320
Receive side processing	0.170	0.090
Total time	**0.518**	**0.598**

– The TTI, the time to transmit a packet
– The signal propagation time from transmitter to receiver
– The processing time at the receive end to decode received data
– Retransmission delay, if any

For eMBB, the user plane latency target is 4 ms for UL and 4 ms for DL. For URLLC, the target is 0.5 ms for UL and 0.5 ms for DL. As way of comparison, the lowest equivalent latency for 4G LTE is approximately 5 ms. Thus, NR seeks to improve latency by a factor of about 10 in the extreme case.

The primary way in which NR addresses the low-latency requirement is, as was described in Sect. 10.7, to allow a minimum standard slot length and hence TTI of 0.125 ms. Further, by the use of mini-slots, this number can be reduced further. For 4G LTE, the TTI is 1 ms. The shorter minimum TTI in NR means the transmission time per packet is shorter which in turn reduces the buffering time in the transmission buffer and shortens the processing time. Shorter processing time is further aided by the use of "front-loaded" reference signals as was described in Sect. 10.12 and control signaling that conveys scheduling information at the beginning of the slot, allowing the receiving device to start processing the received data without prior buffering. Using data from [3], Table 10.3 has been constructed to show DL and UL NR RAN user plane latency, with the UE configured for low-latency services, and 120 kHz subcarrier spacing and hence a TTI of 0.125 ms. Note that the signal propagation time from transmitter to receiver, which is a function of transmission distance, is not shown in the table. For a 100-meter path, this propagation time is only 0.33 μs, and for a 10 km path, it is 0.033 ms.

10.17 Scheduling

Scheduling is key function of NR (and any multiuser mobile system) and to a significant extent determines the overall behavior of the system. Stated in simple terms, the scheduler determines, for each TTI, to which users the shared time-frequency resource is to be assigned and what data rate to apply in transmission. Both uplink and downlink transmissions are scheduled, and thus in the gNB, there is both a DL and UL scheduler. The scheduler is a part of the MAC layer, controlling the

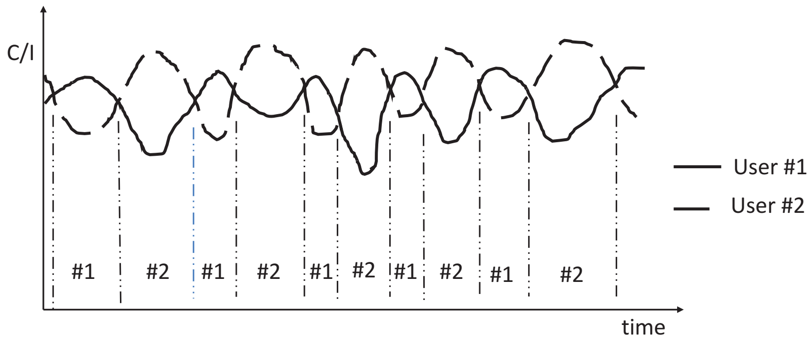

Fig. 10.25 Max C/I allocation

assignment of transmission resources in the form of resource blocks in the frequency domain and OFDM symbols in the time domain.

NR scheduling design is not specified by 3GPP, leaving this up to the system provider. The goal of such design is to factor in the instantaneous channel variations experienced by the various user channels in both the downlink and uplink directions and schedule transmissions that take advantage of these conditions in such a way as to assign resources to the channels that can best use them, such scheduling being referred to as *channel-dependent scheduling*.

Before looking, at a high level, as to how scheduling is implemented in NR, a review of three fundamental scheduling schemes is in order. These such schemes are round-robin (RR), maximum Carrier-to-interference ratio (Max C/I), and proportional fair (PF).

With *round-robin* (RR), radio resources are sequentially allocated among users. It is thus very fair but at the expense of overall system throughput as it may allocate resources to a user or users whose channel conditions do not permit effective data transfer. It most certainly is not channel-dependent scheduling.

With *maximum carrier-to-interference ratio* (Max C/I), users with the highest C/I and hence highest instantaneous achievable data rate are scheduled during the current scheduling decision interval. This results in the highest system throughput but does not provide any kind of fairness among users. Figure 10.25 shows a case of two users experiencing differing C/I and the time allocation to these users under the Max C/I scheme.

Proportional fair (PF) provides a good trade-off between RR and Max C/I, and its features, or slight variations thereof, are thus often implemented. Here users are scheduled according to the ratio of their C/I to their average served data rate. Thus, it schedules not only users with the best channel conditions but also those experiencing a low average data rate because of their channel conditions. This results in all users having a relatively equal probability of being served even though they experience very different instantaneous throughput rates. Those users experiencing a high C/I will be scheduled with a high coding rate and modulation complexity resulting in a high-throughput rate, whereas those experiencing a low C/I will be scheduled with a low coding rate and modulation complexity resulting in a low, but reliable, throughput rate. In addition to C/I and average served data rate, other factors, however, may need to be considered in scheduling such as quality of service guarantees granted to various users.

In NR, downlink and uplink scheduling are separate functions, and their scheduling decisions can be taken independently of each other.

The downlink scheduler dynamically controls the UEs to be transmitted to. Each scheduled UE is informed via the PDCCH of the time-frequency resources upon which the UE's DL-SCH is to be transmitted and the associated transport format, i.e., transport block size, the modulation and coding scheme, and antenna mapping. Downlink scheduling relies on knowledge of the downlink channel obtained via the downlink CSI-RS and reported back up to the gNB.

The uplink scheduler is similar to the downlink one in that it too dynamically controls which UEs are permitted to transmit on their UL-SCH. Each scheduled UE is provided by the PDCCH with a scheduling grant informing it of the time-frequency resources upon which the UE's UL-SCH is to be transmitted and the associated transport format. Uplink scheduling relies on knowledge of the uplink channel obtained from the uplink SRS. There is one difference, however. Here, the gNB does not explicitly schedule a given logical channel, but only schedules broadly the device, leaving it to the device to select from which radio bearers data will be transferred using a defined set of rules, designed to ensure that the UE satisfies the QoS of each radio bearer in an optimal way.

Though dynamic scheduling is the normal mode of scheduling, in certain cases, it is possible to configure transmission without a dynamic grant, in order to reduce control signal overhead.

In the downlink, semi-persistent scheduling is supported. Here, a semi-static scheduling pattern is communicated in advance to the UE via RRC signaling. A downlink assignment is provided by the PDCCH indicating required information such as time-frequency resources and transport format and activated/deactivated via Layer 1 signaling. Upon activation, the UE receives downlink data transmission according to the RRC communicated scheduling pattern.

In the uplink, two types of transmission without a dynamic grant are supported, namely:

– Grant type 1, where the grant, including all necessary transmission parameters, is provided by the RRC which also activates/deactivates the uplink transmission.
– Grant type 2, where the periodicity of transmission is provided by the RRC, but Layer 1/Layer 2 control signaling is used to activate/deactivate the transmission. As with downlink persistent scheduling, receiving the activation command the UE transmits according to the periodicity provided by the RRC if there is data in the buffer.

Both schemes reduce control signaling overhead as well as latency given that once a UE has data to send, it can immediately commence transmission as there is now no need to first send a scheduling request and await a grant.

10.18 Spectrum for 5G

5G NR is conceived to operate ultimately in spectrum ranging from approximately 400 MHz up to about 90 GHz. It is being designed to be able to operate in licensed, unlicensed, and shared frequency bands, shared bands being those where 5G shares the spectrum with nonmobile service providers. It operates in both the FDD and TDD modes.

3GPP defines two frequency ranges for NR, namely, FR1 and FR2. FR1 was specified in 3GPP TS 38.104 V15.1.0 to cover the range 410 to 6000 MHz. However, in V15.5.0, the range was extended covering 410 to 7125 MHz to allow it to be used in unlicensed bands in the 6 GHz region. FR2 was specified in 3GPP 38.104 V15.1.0 to cover the range 24.25 to 52.6 GHz, frequency bands in this range being referred to as millimeter wave (mmWave) bands. Spectrum wise, the FR2 range is the distinguishing feature of 5G relative to 4G. It allows for transmission in bands with much more spectrum than available in FR1 enabling high capacity and data rates.

The lowest frequency bands currently specified in FR1 is for FDD operation, covering approximately 600–800 MHz. Total spectrum in these bands is on the order of 15 to 40 MHz per direction and thus, as we shall see below, result in much lower capacity relative to FR2 bands. The big advantage of such bands, however, is wide area coverage, typically on the order of tens of kilometers, and low outdoor-to-indoor penetration loss.

In the 1.4–2.7 GHz region of FR1, the operation is mostly FDD with typical maximum spectrum per operating frequency being 20 MHz per direction. With carrier aggregation, up to 16 such carriers can be aggregated. Such a large number of aggregated carriers is unlikely, however, with a typical maximum more likely to be about five for a total aggregated bandwidth of about 100 MHz. Bands available in this range have been widely used by 3G and 4G networks.

In the 3.3–4.2 GHz region of FR1, often referred to as the C-Band or the 3.5 GHz band, the operation is TDD, and the maximum spectrum per operator is typically about 100 MHz. This band is likely to be highly used in 5G networks as the amount of spectrum is relatively large yet propagation loss only on the order of about 5 dB more than those bands in the 2 GHz range. Further, this additional path loss can be overcome by utilizing high-gain beamforming antennas leading to coverage similar to that in the 2 GHz bands.

In the 5–6 GHz region of FR1, bands are unlicensed. Such bands are not supported in 3GPP Rel.15 but are expected in the future. They will all offer large amounts of spectrum ranging from about 300 MHz to 800 MHz. Because such bands are available to all, they are accessed on a first come first serve basis, where a potential user must first listen to check if the band is free and only access it if it is. Thus, though the large spectrum potentially affords high data rates, operation is subject to high latency which can lower the average achieved data rate.

In the FR2 range, all operation is TDD, and available spectrum undergoes a large increase relative to FR1. In order of increasing frequency, the first 3GPP specified band covers 24.25–27.5 GHz and is referred to as the 26 GHz band. Next is the

28 GHz band which in fact refers to two bands, namely, one that covers 26.5–29.5 GHz and a second narrower one that covers 27.5–28.35 GHz. Finally, among the specified bands is the 39 GHz band which covers 37.0–40.0 GHz. In these bands, the spectrum per operator range typically up to 400 MHz which supports user rates of up to 5 Gbps. The price paid for these high data rates is much reduced coverage, typically on the order of hundreds of meters or less. Further, this low-coverage problem is made more difficult due to achievable base station output power being generally lower than that in FR1 bands and penetration loss through walls, windows, and doors being much higher than in the lower bands, as discussed in Sect.4.5.6.

Several bands above 40 GHz are under consideration which offer even more spectrum and even higher user data rates, but they each bring their own challenges over and above that of ever-increasing propagation loss. For example, as can be seen in Fig. 4.12, in the 60 GHz region, there is very large oxygen absorption.

Shown in Table 10.4 are operating bands in the 600–800 MHz portion of FR1. It will be seen that the maximum per carrier bandwidth is 20 MHz. Shown in Table 10.5 are those NR operating bands in the mid to high portions of FR1 that afford the largest channel bandwidths per carrier (50 and 100 MHz). These bands are therefore likely to be very desirable. Finally, shown in Table 10.6 are NR operating bands in FR2 where the maximum per carrier bandwidth is 400 MHz.

10.19 5G Data Rates

Extremely high data rates are one of the major features of 5G NR. For NR, the approximate maximum data rate per carrier can be computed as follows [5]:

$$\text{Maximum data rate}\left(\text{Mb/s}\right) = \frac{1}{T_s^{\mu}} \times N_{PRB}^{BW,\mu} \times 12 \times \left(1 - OH\right) \times Q \times R_{\max} \times f \times v \times 10^{-6} \tag{10.1}$$

where:

- $T_s^{\mu} = 10^{-3} / 14 \times 2^{\mu}$ is the average OFDM symbol duration in a subframe of numerology μ, assuming normal cyclic prefix.
- $N_{PRB}^{BW,\propto}$ is the maximum RB allocation in the available system bandwidth BW with numerology μ.
- *OH* is the time-frequency resource overhead. Simply stated, this is the average ratio of all the resource elements not used by the PDSCH or the PUSCH to the total number of available resource elements. It takes the following values:
- 0.14 for FR1 DL
- 0.18 for FR2 DL
- 0.08 for FR1 UL

Table 10.4 NR operating bands in 600–800 MHz portion of FR1

NR band #	Uplink frequency range (MHz)	Downlink frequency range (MHz)	Duplex mode	Maximum channel bandwidth (MHz)
n12	699–716	729–746	FDD	15
n14	788–798	758–768	FDD	10
n28	703–748	758–803	FDD	40
n71	663–698	617–652	FDD	20

Table 10.5 Selected NR operating bands in mid to high portions of FR1

NR band #	Uplink frequency range (MHz)	Downlink frequency range (MHz)	Duplex mode	Maximum channel bandwidth (MHz)
n7	2500–2570	2620–2690	FDD	50
n40	2300–2400	2300–2400	TDD	80
n41	2496–2690	2496–2690	TDD	100
n48	3350–3700	3350–3700	TDD	100
n50	1432–1517	1432–1517	TDD	80
n77	3300–4200	3300–4200	TDD	100
n78	3300–3800	3300–3800	TDD	100
n79	4400–5000	4400–5000	TDD	100

Table 10.6 NR operating bands in FR2

NR band #	Uplink frequency range (MHz)	Downlink frequency range (MHz)	Duplex mode	Maximum channel bandwidth (MHz)
n257	26,500–29,500	26,500–29,500	TDD	400
n258	24,250–27,500	24,250–27,500	TDD	400
n260	37,000–40,000	37,000–40,000	TDD	400
n261	27,500–28,350	27,500–28,350	TDD	400

– 0.10 for FR2 UL
– Q is the bits per symbol for the applied modulation scheme, being 8 for 256-QAM, the highest-order modulation supported.
– R_{max} is the maximum code rate. In 5G NR, it is 948/1024.
– f is a scaling factor used to reflect the capability mismatch between baseband and RF capability of the UE. It is signaled per band and can take the values 1, 0.8, 0.75, and 0.4.
– v is the maximum of layers.

We now consider why the above equation gives maximum data rate.

– $T_s^{\propto}$: The subframe (slot) duration, as indicated in Sect. 10.7, is given by $\frac{10^{-3}}{2^{\mu}}$ secs. Thus, assuming 14 OFDM symbols per slot, the OFDM symbol duration is given by $\frac{10^{-3}}{14 \times 2^{\mu}}$, and $1/\,T_s^{\propto}$ represents the OFDM symbol rate.

- One OFDM symbol is composed of $N_{PRB}^{BW,\mu} \times 12$ modulated subcarriers. Thus, $\frac{1}{T_s^{\mu}} \times N_{PRB}^{BW,\mu} \times 12$ represents the total modulating *symbol* rate, SR_T say.
- $(1 - OH)$ represents the fraction of REs available for user data transmission. Thus, $SR_T \times (1 - OH)$ represents the fraction of modulating symbols available for user data.
- The user bits per OFDM modulating symbol equals $Q \times R_{\max}$. Thus, the user *bit* rate is $SR_T \times (1 - OH) \times Q \times R_{\max}$.
- v: Data rate clearly proportional to the number of layers hence the multiplication by v.
- f: This scaling factor allows the introduction of a practical baseband to RF limitation to the calculation.

The rate computed by Eq. 10.1 is for a single-component carrier. For the case of carrier aggregation, then the maximum rate is the sum of the rates of each component carrier. As indicated in Sect. 9.8, NR supports up to 16 component carriers in both FR1 and FR2.

The maximum downlink user data rate attainable with NR with a single-component carrier occurs for:

- A subcarrier spacing of 120 kHz in FR2 with the maximum resource block allocation of 264 (Table 10.1) and hence a maximum bandwidth of approximately 400 MHz
- 8 layers
- $R_{\max}$ of 948/1024
- Overhead OH of 0.18
- 256-QAM modulation resulting in 8 bits per modulating symbol
- A scaling factor f of 1

Applying this data to Eq. 10.1 gives a maximum component carrier DL data rate of 17.24 Gb/s.

For the uplink, all the above data remains the same with the exception of the maximum number of layers which is now 4 and the overhead which is now 0.1. Applying this data to Eq. 10.1, we get a maximum component carrier UL data rate of 9.46 Gb/s.

As the maximum number of component carriers supported is 16, then if all were of 400 MHz bandwidth, the maximum DL data rate would be 276 Gb/s!! Such a scenario is highly unlikely, but it does demonstrate the ultimate DL data rate capability of NR. A more likely scenario is two 400 MHz component carriers leading to a maximum DL data rate of 34.5 Gb/s which handily meets the IMT-2020 requirement of a minimum DL peak data rate of 20 Gb/s. Table 10.7 shows maximum NR data rates per layer per component carrier.

All FR2 bands are TDD. Thus, with operation in such bands, the maximum downlink and uplink rates are not mutually exclusive. In typical scenarios, there is more downlink demand than uplink demand. Such a scenario could be, for each slot,

Table 10.7 Maximum NR data rates per layer per component carrier

Frequency range	Subcarrier spacing (kHz)	Bandwidth (MHz)	Downlink rate (Mb/s)	Uplink rate (Mb/s)
FR1	15	20	113	121
FR1	15	50	290	309
FR1	30	100	584	625
FR1	60	100	578	618
FR2	60	200	1080	1180
FR2	120	400	2150	2370

ten ODFM symbols for the downlink, one symbol for transition, and three symbols for the uplink. This would result in peak downlink rates of 71% of the maximum available and peak uplink rates of 21% of maximum available.

It is important to not lose sight of the fact that the data rates shown in Table 10.7 represent the very best achievable and that in the real world will only likely be achieved by a small percentage of UEs. These rates assume that the modulation is 256-QAM and the coding rate the maximum of 984/1024. This will only be the case for high SINR situations and thus likely only for UEs located close to the base station. As UE distance from the base station increases, both the modulation order and coding rate will decrease to the point where at the cell edge the modulation is likely to be QPSK and the coding rate less than maximum. Even if the coding rate stays at the maximum, the peak rate attainable with QPSK would only be one fourth that attainable with 256-QAM. As a result, very roughly speaking, the average maximum data rate per layer attainable by a UE in both the DL and UL will likely be about 50% of the maximum rates shown in Table 10.7.

Clearly, in a multiuser system, where total capacity has to be shared by all users, practical user average data rates will be less than the maximum rates discussed above. In such situations, the average user data rate decreases as the number of users increases but is also dependent on bandwidth, received signal levels, and intercell interference levels.

10.20 Transmitter Output Power and Receiver Reference Sensitivity

There are hundreds of individual specifications that define the parameters of 5G base stations and UEs, all precisely stated in the 3GPP 5G NR technical specifications. Two performance parameters, however, have a large impact on coverage, these being transmitter output power and receiver sensitivity. The difference between the two, plus transmitter and receiver antenna gain, defines the maximum path loss tolerable while maintaining packet error rate above defined minimum. In an ideal world, one thus seeks, within limits, to maximize transmitter output power and minimize the value of receiver sensitivity. This section will review at a high-level NR specified base station and UE transmitter output power and receiver reference sensitivity, the latter being a specified receiver sensitivity in a defined reference channel.

10.20.1 Base Station Transmitter Output Power

3GPP base station specifications apply to wide area base stations, medium-range base stations, and local area base stations, and four classes of base stations are defined [6]:

– BS typed *1–0* and *2–0*: Wide area, medium-range, and local area base stations with a BS to UE minimum distance along the ground of 33, 5, and 2 meters, respectively. Types *1–0* and *2–0* have integrated AASs and operate in FR1 and FR2, respectively.
– BS type *1-C* and *1-H*: Wide area, medium range, and local area base stations with a BS to UE minimum coupling loss of 70, 53, and 45 dB. Both types operate in FR1, with type *1-C* connected to antennas via coaxial cables and type *1-H* having an integrated AAS.

For each base station type, the specified maximum base station output power, per carrier and per antenna, is as follows:

– Types *1-C* and *1-H* Local Area BS: 24 dBm
– Types *1-C* and *1-H* Medium-Range BS: 38 dBm
– Types *1-C* and *1-H* Wide Area BS: No upper limit
– Type *1–0* Local Area BS: 33 dBm
– Type *1–0* Medium Range BS: 47 dBm
– Type *1–0* Wide Area Bs: No upper limit
– Type *2–0*: No upper limit

10.20.2 Base Station Receiver Reference Sensitivity

The reference sensitivity level is the minimum received signal level at which there is a sufficient SINR for the specified reference measurement channel to achieve a throughput that's 95% of the maximum possible.

For base station types *1-C*,*1-H*, and *1–0*, all of which operate in FR1, reference sensitivity is specified for reference measurement channels with a QPSK modulated signal and a coding rate of 1/3 [6]. Not all sensitivities specified will be shown here, but to give a sense of the values, those for the channel with the lowest total

Table 10.8 Some BS type *1-C*, *1-H*, and *1–0* reference sensitivities

Ref. measurement channel	Wide area BS ref. sens. (dBm)	Medium area BS ref. sens. (dBm)	Local area BS ref. sens. (dBm)
G-FR1-A1–1	−101.7	−96.7	−93.7
G-FR1-A1–4	−95.3	−90.3	−87.3

subcarrier bandwidth (4.5 MHz), Channel G-FR1–1-A1–1, and the channel with the highest total subcarrier bandwidth (192 MHz), Channel G-FR-1-A1–4, are indicated in Table 10.8.

For base station type *2–0*, which operates in FR2, reference sensitivity, as with the FR1 base stations, is specified for reference measurement channels with a QPSK modulated signal and a coding rate of 1/3. Here, the specified reference sensitivity is declared by the vendor but must lie within a defined range.

For a wide area BS, the range is −96 to −119 dBm for a reference channel with a 50 MHz bandwidth and shifted upward by 3.15 dB for a reference channel with a 100 MHz bandwidth.

For a medium area BS, the range is −91 to −114 dBm for a reference channel with a 50 MHz bandwidth and shifted upward by 3.15 dB for a reference channel with a 100 MHz bandwidth.

For a local area BS, the range is −86 to −109 dBm for a reference channel with a 50 MHz bandwidth and shifted upward by 3.15 dB for a reference channel with a 100 MHz bandwidth.

10.20.3 UE Transmitter Output Power

For FR1, UE maximum output is specified per power class [7]. Power Class 3, the default power class, covers all bands and specifies a maximum output power of 23 dBm. For bands n41, n77, n78, and n79, Power Class 2 is also applicable, with a maximum output power of 26 dBm.

For FR2, four power classes are specified [8]:

- Power Class 1: Fixed wireless access (FWA)
- Power Class 2: Vehicular UE
- Power Class 3: Handheld UE
- Power Class 4: High-power non-handheld UE

For Power Class 1, maximum output power is 35 dBm.

For Power Classes 2, 3, and 4, maximum output power is 23 dBm.

For all power classes, a maximum EIRP is specified as 20 dB above the maximum output power. This implies a maximum antenna gain of 20 dB.

10.20.4 UE Receiver Reference Sensitivity

In FR1, the UE is required to be equipped with a minimum of two receive antenna ports in certain operating bands and with a minimum of four receive antenna ports in others. Here, the reference sensitivity power level is the minimum mean power applied to each one of the UE antenna ports at which there is a sufficient SINR for the specified reference measurement channel to achieve a throughput that's 95% of the maximum possible. For UEs equipped with two receive antenna ports, reference sensitivity is specified for all operating bands and varying channel bandwidths, with QPSK modulated and 1/6 coding rate reference measurement channels [7]. To give a sense of the sensitivity power levels, the lowest level indicated is −100 dBm and occurs in the lowest channel bandwidth (5 MHz) in certain operating bands. The highest level indicated is −84.7 dBm and occurs in the highest channel bandwidth in a specific operating band.

In FR2, as with FR1, reference sensitivity is specified for reference measurement channels with a QPSK modulated signal and a coding rate of 1/6, with levels given are per power class for the various operating bands and channel bandwidths [8]. The lowest level indicated is −97.5 dBm in Power Class 1 (FWA) for a 50 MHz channel bandwidth. The highest level indicated is −79.3 dBm in Power Class 3 (handheld UE) for a 400 MHz channel bandwidth.

10.21 Key 3GPP 5G NR Physical Layer-Related Technical Specifications (TSs) and Reports (TRs)

This text addresses the 5G NR physical layer key technologies at a level believed to be deep enough to impart a solid understanding but not so deep as to be fully comprehensive. Such an understanding requires strong familiarity with the relevant 3GPP specifications and reports. There are tens of 3GPP 5G NR technical specifications (TS) and technical reports (TR). However, some have a larger bearing on the physical layer than others. As this text is largely focused on the physical layer, a listing of those specifications and reports that largely relate to the physical layer is deemed in order and shown in Table 10.9.

10.22 Summary

In earlier chapters, key 5G physical layer technologies were introduced. In this chapter, where and how in NR these technologies are applied have been shown. The NR physical layer is the most complex and most flexible point-to-multipoint radio-access system introduced to date, offering the potential of unprecedented data rates and low latency. It is largely due to the technologies introduced in earlier chapters of this text and the way in which they have been applied, as shown in this chapter,

Table 10.9 Selected 3GPP technical specifications and technical reports

Spec. #	Title
TS 38.101–1	NR, user equipment (UE) radio transmission and reception, part 1: Range 1 standalone
TS 38.101–2	NR, user equipment (UE) radio transmission and reception, part 2: Range 2 standalone
TS 38.104	NR, base station (BS) radio transmission and reception
TS 38.201	NR, physical layer, general description
TS 38.202	NR, services provided by the physical layer
TS 38.211	NR, physical channels and modulation
TS 38.212	NR, multiplexing and channel coding
TS 38.213	NR, physical layer procedures for control
TS 38.214	NR, physical layer procedures for data
TS 38.215	NR, physical layer measurements
TS 38.300	NR, overall description; Stage-2
TS 38.306	NR, user equipment (UE) radio access capabilities
TR 38.801	Study on new radio access technology: radio access architecture and interfaces
TR 38.202	Study on new radio access technology, physical layer aspects
TR 38.817–01	General aspects for user equipment (UE) radio frequency (RF) for NR
TR 38.817–02	General aspects for base station (BS) radio frequency (RF) for NR
TR 38.912	Study on new radio (NR) access technology

that has made this progress possible. Though 5G NR has been developed primarily for mobile access, it is equally applicable to fixed wireless access. In fact, as indicated in Sects. 10.20.3 and 10.20.4, a UE FWA transmitter power class along with FWA receiver reference sensitivity for this class has been specified for FR2. Though not explicitly indicated, FWA is equally applicable to FR1. In the concluding chapter that follows, we will explore the application of NR to FWA.

References

1. 3GPP TR 38.912 (2018) Study on New Radio access technology, Rel. 15, version 15.0.0
2. Dahlman E et al (2018) 5G NR: the next generation wireless access technology. Academic Press, London
3. Holma H et al (eds) (2020) 5G technology: 3GPP new radio. John Wiley & Sons Ltd., Hoboken
4. Ahmadi S (2019) 5G NR: architecture, technology, implementation, and operation of 3GPP new radio standards. Academic Press, London
5. 3GPP TS 38.306 (2020) NR; User Equipment (UE) radio access capabilities, Rel. 16, version 16.0.0
6. 3GPP TS 38.104 (2020) NR; Base Station (BS) radio transmission and reception, Rel. 16, version 16.3.0
7. 3GPP TS 38.101-1 (2020) NR; User Equipment (UE) radio transmission and reception; Part 1; Range 1 Standalone, Rel.16, version 16.3.0
8. 3GPP TS 38.101-2 (2020) NR; User Equipment (UE) radio transmission and reception; Part 2; Range 2 Standalone, Rel. 16, version 16.3.0