

Uplink Tx diversity is not explicitly specified, but open-loop Tx diversity can be performed transparently by, e.g., frequency hopping. Furthermore, Tx antenna selection diversity can be performed by utilizing SRI or CSI-RS.

7.3.4 BEAM MANAGEMENT

One of the main new features in NR is the support for analog beam-forming, which is foreseen to be prevailing at millimeter-wave frequencies. For this purpose a new framework called beam management has been developed in order to support analog beam-forming at both the BS and the UE side. Beam management has been defined in 3GPP as a set of Layer 1/2 procedures to acquire and maintain a set of BS and/or UE beams²⁶ that can be used for downlink and uplink transmission/reception [1]. It includes a number of features, such as:

- Sweeping. Covering an angular sector by sweeping analog beams over the sector.
- Measurement. Measuring the quality of different beams.
- Reporting. Reporting beam information such as which beams are best and their measured qualities.
- Determination. Selecting one or a few beams out of a number of candidate beams.
- Indication. Indicating which beam or beams has been or have been selected for data transmission.
- Switching. Switching to another beam if another beam gets higher quality than the current beam.
- Recovery. Finding a new beam if the current beam cannot maintain a communication link due to, e.g., blockage.

If analog beam-forming is used in both the BS and the UE, beams at both sides need to be found so that beam pairs rather than just beams need to be acquired and maintained. Beam management is used for both data and control channels, i.e., the physical downlink shared channel (PDSCH) and physical downlink control channel (PDCCH) (see Chapter 2). Multiple beam pair links can be established for robustness, joint transmission or spatial multiplexing.

Beam management is intended to be a fast process by involving only Layer 1 and 2 signaling. It is aimed at maintaining beam pair links between a UE and beams from one BS or several tightly synchronized BSs. Mobility between unsynchronized BSs is handled by other procedures involving slower Layer 3 signaling. The beam management framework has also been designed for flexibility so it can be adapted to different use cases. For example, for short data sessions, when only small amount of data should be transmitted/received, a quick beam management procedure using a coarse level of accuracy could be used, while a more refined procedure can be invoked for longer sessions.

A general principle of beam finding is to transmit reference signals in a number of candidate beams and estimate the quality of the received reference signal at the receiver for each candidate Tx beam. In NR, two different reference signal types can be used for downlink beam management: SSB and CSI-RS. For uplink beam management, SRS can be used. A reference to the beam(s) with highest quality is then reported back to the transmitter so that the transmitter knows which beam is best to use in subsequent data transmissions. For NR downlink, an index to a previously transmitted reference signal (SSB or CSI-RS) is used as a reference to the best beam, so-called SSB resource indicator

²⁶The NR specifications seldom use the term “beam”. What we mean by beam herein is what in the NR specifications is referred to as *spatial domain Tx filter* or *spatial domain Rx filter*.

(SSBRI) or CRI. Similarly, the receiver can find its best Rx beam by estimating the quality of the received reference signal for each candidate Rx beam. During the Rx beam finding procedure, the transmitter should repeat the transmission of the reference signal in the same Tx beam so that the receiver can make a fair comparison of the candidate Rx beams. These procedures should be repeated whenever needed to track UE movements and changes in the radio environment.

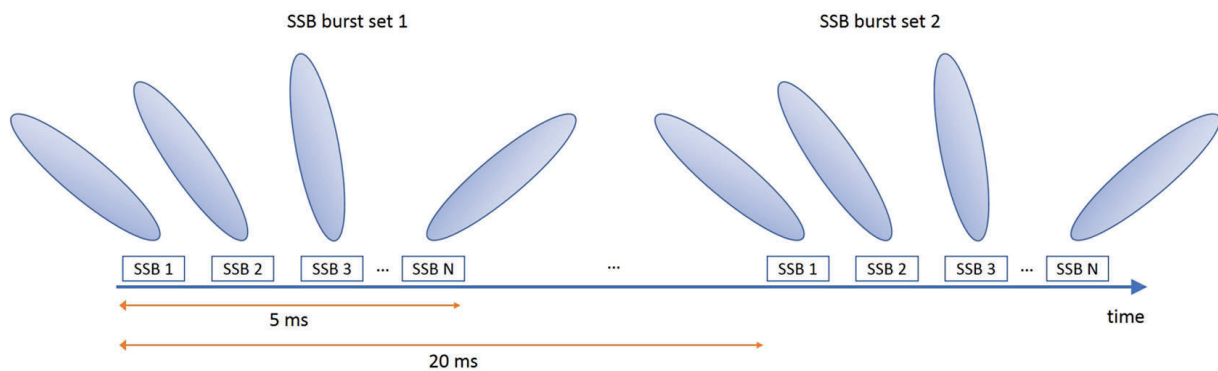
The main focus of this section is on downlink beam management. In many cases uplink beam management is not needed. If the BS and UE has so-called *beam correspondence*, the beam pair links established by downlink beam management procedures can also be used for uplink without any need for separate uplink procedures. Beam correspondence is a kind of reciprocity that implies that a BS or UE can determine its Tx beam based on measurements on its Rx beams, or vice versa. For example, if a set of beam-forming weights gives the same beam pattern for the Tx and the Rx, beam correspondence holds. If beam correspondence does not hold, separate uplink beam management procedures can be invoked. These are similar to the corresponding downlink procedures. Uplink specific beam management aspects are summarized at the end of this section.

7.3.4.1 Beam Acquisition During Initial Access

Before a UE enters the network, a BS does not have any information about the direction to the UE. The signals in the initial access procedure therefore need to be transmitted and received without any prior knowledge about the direction to the UE. At low frequencies, these signals can be transmitted with a wide beam that covers the entire angular sector of the cell. At millimeter-wave frequencies, however, the antenna gain with such a wide beam may not be sufficient to achieve the desired coverage. Therefore, NR supports beam-forming also of initial access signals. Since the direction to a UE is not known a priori, the set of beams carrying the initial access signals must cover the entire sector, e.g., by beam sweeping.

The first signals that the BS transmits to aid a UE to access the network are contained in the SSB. The purpose with SSB is to provide coarse time and frequency synchronization and basic system information such as how the UE shall access the system, indication of the physical cell identity (ID), and where to find the remaining configurations. The SSB is transmitted periodically and can be beam-formed in order to ensure sufficient cell coverage. The SSB is then transmitted repeatedly in different beams, a so-called SSB burst set. For carrier frequencies above 6 GHz, up to 64 beams can be used within an SSB burst set of 5 ms. This burst set is then repeated with a specified periodicity. For initial cell selection the default value of the SSB burst set periodicity is 20 ms. See Fig. 7.13 for an illustration of beam-forming of SSB.

During initial access, the UE measures the different SSBs in an SSB burst set in order to determine the best BS Tx beam. If the UE has analog beam-forming it can measure multiple SSB burst sets to find a suitable Rx beam, or it can use a wide beam. The UE then transmits the physical random access channel (PRACH) preamble in a resource that is associated to the SSB that corresponds to the determined BS Tx beam. If the UE has analog beam-forming and beam correspondence, the UE can transmit the preamble with a Tx beam that corresponds to the Rx beam it used when receiving the SSB from the best BS Tx beam. The BS can receive the PRACH preamble using a wide beam or the same beams it used for transmitting the SSB burst set. When the BS has received the PRACH preamble it can deduce from the corresponding PRACH resource which BS Tx beam was best for that UE. The BS can then use the identified Tx beam in possible beam refinement procedures or subsequent data and control transmission.

**FIGURE 7.13**

SSB burst. Two SSB burst sets with N beams in each set.

Since the SSB can be transmitted with a periodic beam sweep, it is also useful for beam management purposes. SSB can be used not only for initial beam acquisition but also for other purposes such as input to beam refinement procedures and discovery of new beams when a UE moves or if changes occur in the radio environment. In some cases, SSB may be all that is needed in beam management. The SSB transmissions can be used both by the BS and the UE to find suitable Tx and Rx beams.

To provide robustness against mobility and blockage and to reduce signaling overhead, the SSB beams can be relatively wide so that the cell can be covered by a few beams. For UEs with low data rate requirements or good channel conditions it may be sufficient to use the wide beams acquired during initial access also for the data transmission. In other cases, narrower beams with higher gain can be acquired by subsequent beam refinement procedures. The beams found during initial access can then be useful as input to the beam refinement procedures, e.g., to trigger a beam sweep with narrow beams centered around the direction estimated during initial access.

7.3.4.2 Beam Management Procedures

Although not explicitly stated in the specifications, downlink beam management has been divided into three procedures [1]:

- P-1. The purpose of P-1 is to find initial BS Tx beam(s) and possibly also UE Rx beam(s) by performing a beam sweep over a relatively wide angular sector.
- P-2. This is used for beam refinement of the BS Tx beam(s) by performing a beam sweep in a more narrow angular sector than in P1.
- P-3. This is used for performing an Rx beam sweep at the UE. In P-3, the BS Tx beam is fixed during the UE Rx beam sweep.

There are similarities between the procedures and not all procedures are needed. Furthermore, P-2 can be a special case of P-1. An example of how the P-1, P-2, and P-3 procedures can be performed is schematically illustrated in Fig. 7.14. In P-1, the BS performs a beam sweep over an angular sector that covers the entire cell by transmitting a unique reference signal in each beam. To limit the number of beams in such a wide beam sweep the beams could be relatively wide to give an initial, coarse estimate of the best beam direction. The reference signal could be, e.g., the SSBs during initial access

or a periodic CSI-RS transmission that has been configured for beam management. The UE measures the power of the received reference signals from all BS Tx beams using a wide Rx beam and reports to the BS which beam has the highest received power. In P-2, the BS performs beam refinement by an aperiodic CSI-RS transmission using narrower beams in an angular sector around the best beam reported by the UE in P-1. The UE measures the power of the received CSI-RSs from these BS Tx beams, still using a wide Rx beam, and it reports to the BS which of the narrow beams has the highest received power. In P-3, the BS transmits CSI-RS repeatedly in the best narrow beam reported by the UE in P-2 so that the UE can perform an Rx beam sweep to find its best Rx beam by measuring the power of the received CSI-RS in each Rx beam. In the data transmission, the BS uses the best BS Tx beam found during P-2 and the UE uses the best UE Rx beam found during P-3.

Note that this is just one example of how to perform beam management and other ways are possible. For example, P-1 could be a joint BS Tx/UE Rx beam sweep in which the UE sweeps its Rx beams for each BS Tx beam. The BS then has to repeat the reference signal transmissions in each BS Tx beam so that the UE can evaluate different Rx beams for every BS Tx beam. Therefore, this approach is more costly in terms of reference signaling overhead and beam acquisition time.

To provide robustness against blocking, a UE can be configured to monitor PDCCH on multiple beam pair links. For example, while data transmission is being performed on an active beam pair link, the UE can monitor PDCCH on another beam pair as a backup link for swift fallback if there should be a sudden blockage of the active link.

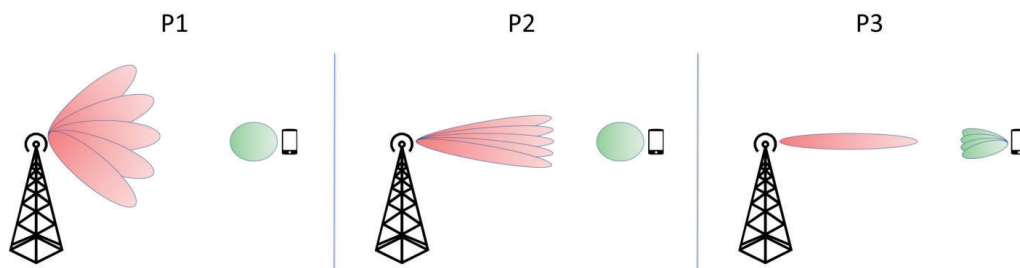


FIGURE 7.14

Beam management procedures. Schematic illustration of the beam management procedures P-1, P-2, and P-3.

Which beam management process to use is not explicitly configured. Instead, the CSI acquisition framework described previously is used to configure a UE with the pertinent report and resource settings. For beam management, a report setting can for example contain information regarding the number of beams to report on, which CSI parameters to report (e.g., L1-RSRP), time-domain behavior, and frequency granularity of the reporting. A resource setting can contain information regarding which reference signal type (CSI-RS, SSB) the UE shall measure on, time-domain behavior, and one or multiple resource sets each containing multiple CSI-RS resources.

7.3.4.3 Beam Measurement and Reporting

A central component in beam management is measuring and reporting of the quality of different candidate beams. In downlink beam management, measurements are performed by the UE on SSB or CSI-RS transmitted by the BS. Beam sweeping using CSI-RS is more flexible than SSB since there is more freedom in configuring the CSI-RS transmission. For example, the number of beams and their

coverage area can be configured more flexibly. A CSI-RS beam sweep is UE-specifically configured, so it is possible to tailor a beam sweep for a particular UE. For example, the beams in a CSI-RS beam sweep can be narrow beams centered around a wide beam acquired from SSB during initial access. A CSI-RS beam sweep does not need to be transmitted periodically but can be triggered whenever needed. Other advantages with CSI-RS over SSB for beam management are that CSI-RS is more wide-band and can use two antenna ports, which potentially could lead to more reliable beam measurements. For example, two antenna ports can be used for transmitting with two orthogonal polarizations in order to reduce polarization mismatch losses. An advantage with using SSB for beam management is that SSB is transmitted for initial access purpose anyway, so using it also for beam management incurs no additional reference signal overhead.

CSI-RS for beam management can be periodic, semi-persistent, or aperiodic. Periodic and semi-persistent CSI-RS can be efficient in wide beam sweeps for acquiring coarse beams when there is little a priori information as regards the directions to UEs, or for beam sweeps at high load for covering a service area containing multiple UEs. To save signaling overhead, aperiodic CSI-RS can be more efficient at low load to perform a refined, UE-specific beam sweep when there is some prior knowledge about the direction to the UE.

The quality metric used for beam management is L1-RSRP for both SSB and CSI-RS. In the case that CSI-RS resources with two ports have been configured, the UE shall report the linear average over both ports. Information as regards which beam the UE has selected is conveyed by reporting an index to the corresponding CSI-RS resource, i.e. CRI. The UE can be configured by the network to report a number of strongest beams.

NR also supports that the UE can report a group of BS Tx beams that can be received simultaneously by the UE, e.g., by using different antenna panels. This can be useful for establishing multiple beam pair links for robustness, joint transmission or spatial multiplexing. The grouping-based reporting can be turned on and off on a per UE basis.

For non-grouping-based reporting, the UE can be configured to measure on up to 64 beams in a beam sweep and report up to four beams in an reporting instance. The reports are transmitted on physical uplink control channel (PUCCH) or PUSCH depending on the reporting being periodic, semi-persistent, or aperiodic.

Measurement and reporting for beam management and CSI acquisition have a common framework so it can be used for both purposes by different configurations. At low frequencies with digital beam-forming, beam management is typically not needed. In this case, the UE can be configured to measure on a number of CSI-RS resources, select the best resources, and report CSI for these resources. At high frequencies with hybrid beam-forming the optimal approach would be to jointly optimize the analog beam-formers and the digital precoder. However, this may not be feasible due to high computational complexity and signaling overhead. A more practical approach is to determine the analog beam-formers and digital precoder separately. The analog beam-formers can then be determined using a beam management procedure followed by design of the digital precoder using the selected analog beams and the CSI acquisition framework. For example, the BS can first transmit SSB or CSI-RS configured for beam management in order to select beam pair links. The UE measures and reports indices and qualities of the preferred beam pair links. The BS can then transmit CSI-RS for CSI acquisition using the selected beam pair links. The UE measures on the CSI-RSs and reports CQI, PMI, and RI. The BS then transmits data using the selected analog beams and digital precoder.

7.3.4.4 Beam Indication

For downlink, NR supports beam management with and without beam indication. Beam indication is an indication to the UE which Tx beam the BS will use in coming data or control transmission so that the UE can update its Rx beam accordingly. This may be needed when analog beam-forming is used in the UE. If the BS changes its Tx beam the UE may need to change its Rx beam at the same time so that it matches the new BS Tx beam, or the link may be lost. It can also be needed for the UE to know which beam to receive a CSI-RS beam sweep in a limited sector (a P-2 procedure). To be able to do this with analog beam-forming, the UE needs to know beforehand which Rx beam to use so that it can switch to that beam when the data is transmitted with the new Tx beam. The downlink beam indication provides information to the UE so that it can use an Rx beam that is suitable for the new Tx beam.

Beam indication is needed when several beam pair links are maintained, e.g., when monitoring multiple PDCCHs, or when a joint BS Tx and UE Rx beam sweep is performed. In that case all combinations of BS Tx beams and UE Rx beams are evaluated, the UE reports the best beam pairs and the BS selects which Tx beam to use. A beam indication is then needed since the UE needs to know which Tx beam the BS uses in order to set its Rx beam.

In other cases, explicit beam indication may not be needed. For example, if a single beam pair link is used for PDSCH and PDCCH transmission and the UE reports only the best BS Tx beam in a beam sweep, the UE does not need any beam indication provided that the BS uses the reported beam in the next transmission. When operating without beam indication, different BS Tx beams are evaluated under the assumption that the UE holds its Rx beam fixed and different UE Rx beams are evaluated under the assumption that the BS holds its Tx beam fixed. The UE can be informed that the BS keeps its Tx beam fixed by a CSI-RS resource set configuration that contains an information element which indicates that the BS repeats its CSI-RS transmission in the same Tx beam. For example, in a sequential P-2/P-3 procedure the BS first performs a P-2 Tx beam sweep and the UE keeps its Rx beam fixed (or uses a wide beam). After the P-2 procedure, the BS updates its Tx beam without informing the UE and triggers a P-3 UE Rx beam sweep. After the P-3 procedure, the UE updates its Rx beam without informing the BS. The UE only needs to remember which Rx beam was best and use that information in the next reception without any beam indication from the BS. An advantage with operation without beam indication is the reduced signaling overhead and delay associated with beam indication.

A beam indication is a reference to a previously transmitted reference signal in the form of a spatial QCL relation. More specifically, it is an indication that the UE can assume that the DM-RS of PDSCH/PDCCH is spatially quasi-co-located with a previously transmitted downlink reference signal, e.g., a particular SSB or CSI-RS resource. This means that the UE can use the same Rx beam for the coming data/control transmission as it used when it received the indicated reference signal. Beam indication can also be used to provide a spatial QCL relation between different reference signals, e.g., that an aperiodic CSI-RS is spatially quasi-co-located with an SSB. This could be used for, e.g., beam refinement when an aperiodic CSI-RS beam sweep should be performed with narrow beams within a wide SSB beam.

Downlink beam indication is performed by signaling a so-called transmission configuration indicator (TCI) to the UE which provides a spatial QCL reference that the UE can use to set its Rx beam. This is similar to the PDSCH rate matching and quasicollocation indicator (PQI) used for CoMP operation in LTE. The UE is configured with a list of TCI states by higher-layer signaling, where each TCI state is configured with a set of CSI-RS or SSB IDs. In each state, one CSI-RS or SSB ID that should be used as spatial QCL reference is selected. The beam indication is then performed by signaling a selected

TCI state to the UE that it should use for obtaining a QCL reference for the coming PDSCH/PDCCH transmission. By maintaining multiple TCI states, the BS can dynamically switch between different Tx beams. Before the TCI states have been configured and activated, the UE can assume that the DM-RS of PDSCH is spatially quasi-co-located with the SSB determined in the initial access procedure. This means that the UE can use the same Rx beam it used when it received the SSB during initial access. A TCI state can also contain a QCL reference for time/frequency parameters such as delay and Doppler spread. Reference signal for time/frequency QCL can be, e.g., a TRS (see Section 2.5). For frequencies below 6 GHz, a TCI state may contain only a time/frequency QCL reference and no spatial QCL information.

7.3.4.5 Beam Recovery

Narrow-beam transmission and reception is useful for improving the link budget at millimeter-wave frequencies but this may become susceptible to so-called beam failure. A beam failure means that the quality of the beam pair links for all control channels becomes too low for maintaining communication. This can be caused by, e.g., sudden blockage or failure in a beam management process. This may lead to radio link failure (RLF) wherein a costly higher-layer reconnection procedure is needed. To avoid this, NR supports a faster procedure using lower layer signaling to recover from beam failure, referred to as beam recovery. For example, instead of initiating a cell reselection when a beam pair link quality becomes too low, a beam pair reselection within the cell can be performed.

Beam failure is detected by monitoring a beam failure detection reference signal and assessing if a beam failure trigger condition has been met. The beam failure detection reference signal can be SSB or a periodic CSI-RS configured for beam management and is associated with the Tx beam with which a control channel is transmitted. Beam failure detection is triggered if a number of consecutive detected beam failure instances exceeds a maximum value, where a beam failure instance occurs if the block error rate (BLER) of a hypothetical PDCCH transmission is above a threshold.

To find candidate new beams, the UE monitors a beam identification reference signal, which can be SSB or a periodic CSI-RS configured for beam management. These reference signals can be transmitted with wider beams than the ones used for data. They can be used both for finding a new BS Tx beam and a new UE Rx beam. When a UE has declared beam failure and found a new beam it transmits a beam recovery request message to the serving BS. The BS responds to the request by transmitting a recovery response over PDCCH to the UE and the UE monitors the control channel for the response. If the response is received successfully, the beam recovery is completed and a new beam pair link has thus been established. If the UE cannot detect any response within a specified time, the UE may perform a retransmission of the request. If the UE cannot detect any response after a specified number of retransmissions, then it notifies higher layers, potentially leading to RLF and cell reselection.

7.3.4.6 Uplink Beam Management

As mentioned previously, if the BS and UE have beam correspondence, separate procedures for uplink beam management are not needed, since the beam determination can be based on downlink procedures. If beam correspondence does not hold, separate uplink procedures similar to P-1, P-2, and P-3 described previously can be used. These are referred to as U-1, U-2, and U-3, and are based on uplink reference signals.

In uplink beam management, SRS is used in a similar way as CSI-RS is used in downlink beam management. An uplink beam management procedure is initiated by the BS by triggering one or several

SRS resource sets, where each resource set contains a number of SRS resources. One SRS resource is transmitted per UE Tx beam, the BS measures the received SRSs, determines a preferred UE Tx beam, and reports an SRI to the UE. The UE may also repeat an SRS transmission in a Tx beam so that the BS can find a suitable Rx beam. Multiple SRS resource sets can be used to train multiple beam-formers in parallel by triggering one set per beam-former.

Uplink beam indication may be needed so that the UE can adjust its Tx beam based on the Rx beam used by the BS. This may be needed also when beam correspondence holds and no separate uplink beam management is performed. Beam indication for uplink is supported for PUCCH and SRS. PUCCH can be spatially related to SRS or downlink signals such as SSB or CSI-RS. Such spatial relations can be used if the UE has beam correspondence. For example, if a UE receives a downlink RS in an Rx beam pointing in a certain direction it can perform an uplink transmission in a Tx beam pointing in the same direction. If the UE does not have beam correspondence, it can use a previously transmitted SRS for determining its Tx beam.

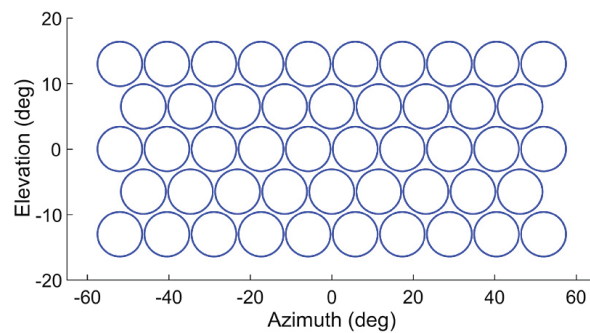
7.4 EXPERIMENTAL RESULTS

In order to lend some empirical support to the theoretical descriptions in previous sections, this section shows results from two different measurement campaigns that demonstrate the effectiveness of beam-forming at millimeter-wave frequencies. In the first campaign, an assessment of the achievable beam-forming gain using analog beam-forming in different environments was made and in the second campaign, successful beam tracking of a high speed UE was demonstrated. Furthermore, we present some results from system simulations using 3GPP models that illustrate what SINR levels and beam-forming gains can be expected in an urban macro scenario at 30 GHz for different antenna sizes at the BS and the UE.

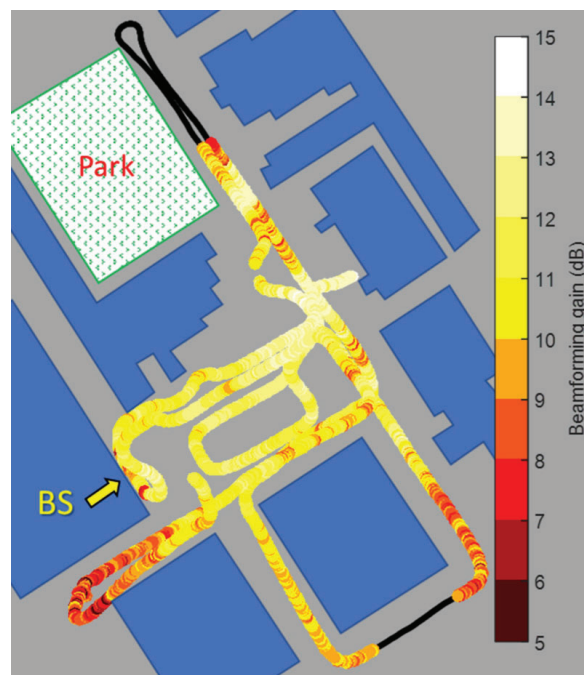
7.4.1 BEAM-FORMING GAIN

In [42], measurements of beam-forming gain using analog beam-forming at 15 GHz in different environments were reported. The measurements were performed with a 5G test-bed using a planar antenna array with 8×8 antenna elements divided into 32 subarrays of 2×1 antenna elements each. Analog beam-forming using a 2-D GoB with 48 beams was applied over the subarrays to generate beams with an azimuth and elevation half-power beam width (HPBW) of 14° . Reference signals were transmitted sequentially in each beam and a UE prototype measured and logged average received power of the reference signals for all beams every 100 ms along a drive route. The analog beam grid is depicted in Fig. 7.15.

Fig. 7.16 shows estimated beam-forming gain along a drive route in an outdoor environment consisting of a square with surrounding buildings and streets. Buildings and the BS position are indicated in the sketch of the environment. The BS is mounted 7 m above the ground and faces the $110 \text{ m} \times 60 \text{ m}$ square, which is surrounded by buildings of two to eight storeys. The measurement positions on the square are in LoS and the positions on the four streets out from the square are mainly in NLoS from the BS. The beam-forming gain was estimated by comparing the received power in the strongest beam with the received power averaged over all beams. The brightness of each dot on the route shows

**FIGURE 7.15**

GoB. Beam grid used in the measurements.

**FIGURE 7.16**

Beam-forming gain. Estimated beam-forming gain along an outdoor drive route.

the estimated beam-forming gain for that position.²⁷ Since the analog beam-forming is performed over 32 subarrays, the maximum beam-forming gain is 15 dB. It is clear from this figure that the LoS positions on the open square in front of the BS have higher beam-forming gain than the NLoS positions on the streets behind buildings. It can also be seen that the beam-forming gain is somewhat higher in

²⁷The thinner black parts of the route correspond to lost connections.

the center of the square than in the positions close to the surrounding buildings where reflections have a negative impact on the beam-forming gain.

Fig. 7.17 shows cumulative distribution functions (CDFs) of the estimated beam-forming gain from the drive route shown in Fig. 7.16. The measurement results have been separated into LoS and NLoS positions along the route. The results show that the beam-forming gain is high in the LoS positions, mainly in the range 10–13 dB. As expected, the beam-forming gain in the NLoS positions is lower, around 7–12 dB. Results of measurements in an indoor environment with rich multipath were also reported in [42]. In this environment, the beam-forming gain was around 6–11 dB. Although the gain with analog beam-forming is substantial also in the reflective environments, there is potential for higher gain with digital precoding that can utilize several different propagation paths. The potential gain with hybrid beam-forming was also assessed in [42] by assuming a perfect coherent combining of the best analog beams. A substantial improvement was found by combining only a few beams. To achieve a certain total beam-forming gain, more beams were required in the reflective environments.

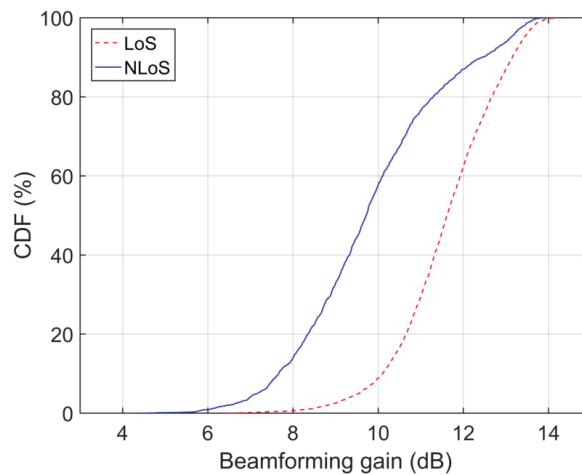
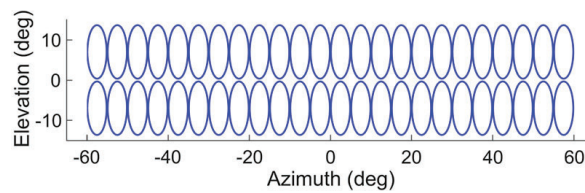


FIGURE 7.17

CDF of beam-forming gain. CDF of estimated beam-forming gain.

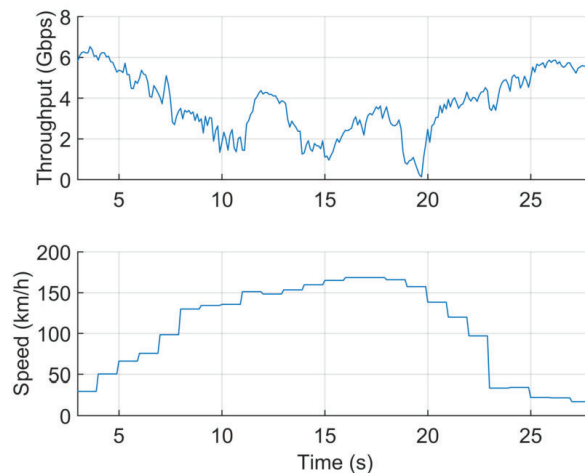
7.4.2 BEAM TRACKING

NR should support use cases with high mobility such as users traveling with high speed in, e.g., trains and cars. At millimeter-wave frequencies it may be essential to track moving users with narrow beams using analog beam-forming. Experimental results of beam tracking at 28 GHz of a car moving at high speed have been reported in [26]. The measurements were performed with a test system consisting of four transmission points deployed along a race track in Republic of Korea and one UE mounted on top of a car. Each transmission point had a rectangular antenna array with $16 \times 4 = 64$ antenna elements per polarization. Analog beam-forming was performed by using a GoB shown in Fig. 7.18. The HPBW of the array was 6° in azimuth and 24° in elevation. The UE had four dual-polarized, directive antennas covering different angular sectors. The bandwidth of the system was 800 MHz and supported MIMO transmission of two layers providing a peak rate of about 7.8 Gbps.

**FIGURE 7.18**

GoB. Beam grid used in the measurements.

Beam tracking was performed by transmitting beam-formed reference signals which were measured upon and reported by the UE. In the beam selection, beams from all four transmission points were evaluated and the best beam was selected regardless from which antenna the beam was transmitted. The transmission points were connected to the same baseband. Drive tests were performed on the race track at speeds up to 170 km/h. Fig. 7.19 shows the downlink throughput and UE speed as a function of time along the drive route. It can be seen that at low speed the throughput exceeded 6 Gbps and at the highest speed of 170 km/h, a throughput of up to 3.6 Gbps was achieved. Furthermore, successful beam switching was reported in [26], even though the car stayed in a beam for only 25 ms at high speeds.

**FIGURE 7.19**

Throughput. Downlink throughput and UE speed vs. time.

7.4.3 SYSTEM SIMULATIONS

Finally, this section provides some results from system simulations using a 3GPP NR scenario and channel model that show how analog and hybrid beam-forming can increase the SINR at millimeter-wave frequencies.

Ideally, the analog beam-forming gain is $10\log_{10}(N_T N_R)$ (dB), where N_T and N_R is the number of Tx and Rx antenna elements, respectively. However, as discussed in Section 7.2.1, this can only be achieved in highly correlated channels when a LoS or specular reflection direction is known both at the transmitter and the receiver so that a beam can be steered in this direction. In practice angular spread will limit the analog beam-forming gain. Furthermore, with analog beam-forming a limited number of hypothesized directions must be tested sequentially in time to find the best beam direction, typically by using a fixed GoB. Since the true direction will not coincide exactly with an angle grid point, a so-called straddling loss will incur. To evaluate the beam-forming gains that are more realistic to achieve, we show results from system simulations based on models that take these effects into account.

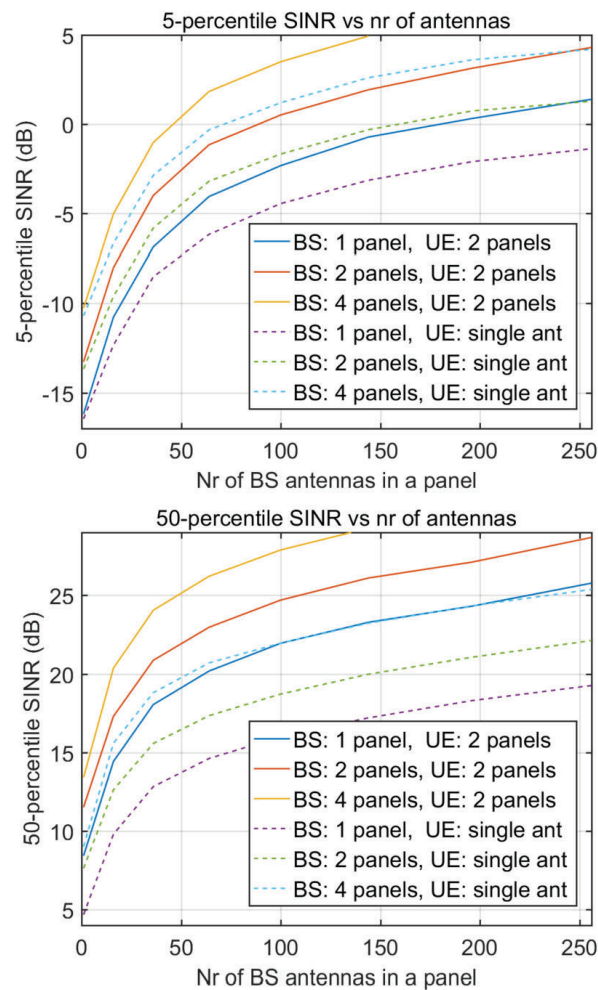
We evaluate the downlink SINR as a function of the number of antenna elements in a BS panel for an urban macro scenario at 30 GHz carrier frequency using the 5G channel model²⁸ in [2]. The inter-site distance is 200 m and other simulation assumptions are as in [37]. We assume that the BS has one, two, or four square panels with $N_T \times N_T$ dual-polarized antenna elements each. Furthermore, we assume that analog beam-forming is performed per polarization within each panel and that the panels and polarizations are combined to a single transmit layer using SVD precoding assuming perfect CSIT. Two different UE antenna configurations are considered: one with and one without beam-forming. For the case without beam-forming the UE is assumed to have one dual-polarized isotropic antenna and for the case with beam-forming the UE is assumed to have two panels mounted back-to-back, where each panel has 2×4 dual-polarized antenna elements. During data reception only the UE panel with highest received power during the beam management procedure is used, i.e., panel combining is not used in the UE. It is assumed that analog beam-forming is performed per polarization within this panel and that the polarizations are combined with interference rejection combining.

An ideal beam management procedure is simulated by finding the best pair of analog Tx beam at the BS and analog Rx beam at the UE. The beams are selected from a GoB constructed from DFT beams without angular oversampling. The zenith angles in the BS GoB are given by the angles $180^\circ \cdot (n - 1/2)/N$, $n = 1, \dots, N$, where N is the number of antenna rows in a panel, that lie in the interval $[90^\circ \ 160^\circ]$. The azimuth angles in the BS GoB are given by the angles $180^\circ \cdot (n - 1/2)/N - 90^\circ$, $n = 1, \dots, N$, where N is the number of antenna columns in a panel that lie in the interval $[-60^\circ \ 60^\circ]$. The same analog beam is used in both polarizations and in all BS panels in the case that it has multiple panels.

Fig. 7.20 shows the 5- and 50-percentile, respectively, in the SINR CDF over all UEs in the network as a function of the number of BS antenna elements, N_T^2 , per panel assuming square panels each having $N_T \times N_T$ elements. The different curves are for different numbers of BS and UE panels. The solid curves are for the cases with one, two, or four BS panels and two UE panels. The dashed curves are the corresponding results for a single dual-polarized antenna at the UE.

The results show that without any BS or UE beam-forming the performance is not acceptable, since the 5-percentile SINR is below -16 dB. With a single 8×8 BS panel and two 2×4 panels in the UE, the 5-percentile SINR is increased to -4 dB, which is a significant improvement. The results also show that doubling the number of BS panels gives roughly a 3 dB increase in SINR. Doubling the number of antenna elements within a panel gives slightly less than 3 dB 5-percentile SINR gain for

²⁸However, recall from Chapter 3 that the highly resolved properties of the channel model have not been experimentally validated.

**FIGURE 7.20**

Simulation results. SINR vs. the number of antenna elements in a BS antenna panel obtained from system simulations using the 3GPP urban macro 5G channel model.

large panels. This may be due to the angular spread in the channel combined with narrow analog beams. Note that the multiple panels in the BS have in these simulations been used for coherent combining to increase coverage. Alternatively, the panels could be used to increase capacity by, e.g., MU-MIMO transmission.

The SINR gain of beam-forming in the UE, i.e., going from a single dual-polarized antenna in the UE to switching between two 2×4 panels, is around 2–3 dB at the 5-percentile and around 5–6 dB at the median. A possible explanation of the lower SINR gain at the 5-percentile is that UEs with low SINR are probably located indoors and/or in NLoS and therefore experience a larger angular spread which will limit the analog beam-forming gain.

REFERENCES

- [1] 3GPP TR 38.802, Study on New Radio Access Technology, Physical Layer Aspects, 2017.
- [2] 3GPP TR 38.900, Study on channel model for frequency spectrum above 6 GHz, 2017.
- [3] 3GPP TR 38.901, Study on channel model for frequencies from 0.5 to 100 GHz, 2017.
- [4] 3GPP TS 36.211, Physical channels and modulation, 2018.
- [5] 3GPP TS 38.101-1, NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone, 2017.
- [6] S.M. Alamouti, A simple transmit diversity technique for wireless communications, *IEEE Journal on Selected Areas in Communications* 16 (8) (1998, Oct.) 1451–1458.
- [7] X. An, C.S. Sum, R.V. Prasad, J. Wang, Z. Lan, J. Wang, R. Hekmat, H. Harada, I. Niemegeers, Beam switching support to resolve link-blockage problem in 60 GHz WPANs, in: 2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, 2009, Sept., pp. 390–394.
- [8] J.B. Andersen, Array gain and capacity for known random channels with multiple element arrays at both ends, *IEEE Journal on Selected Areas in Communications* 18 (11) (2000, Nov.) 2172–2178.
- [9] E. Björnson, M. Bengtsson, B. Ottersten, Optimal multiuser transmit beamforming: a difficult problem with a simple solution structure, *IEEE Signal Processing Magazine* 31 (4) (2014, July) 142–148.
- [10] G. Caire, S. Shamai, On the achievable throughput of a multiantenna Gaussian broadcast channel, *IEEE Transactions on Information Theory* 49 (7) (2003, July) 1691–1706.
- [11] D. Colombi, B. Thors, C. Törnevik, Implications of EMF exposure limits on output power levels for 5G devices above 6 GHz, *IEEE Antennas and Wireless Propagation Letters* 14 (2015) 1247–1249.
- [12] M. Costa, Writing on dirty paper, *IEEE Transactions on Information Theory* 29 (3) (1983, May) 439–441.
- [13] E. Dahlman, S. Parkvall, J. Sköld, 4G LTE-Advanced Pro and The Road to 5G, third ed., Academic Press, 2016.
- [14] R.F.H. Fischer, C. Windpassinger, A. Lampe, J.B. Huber, Space-time transmission using Tomlinson–Harashima precoding, in: Proc. 4th Int. ITG Conf. Source and Channel Coding, 2002, pp. 139–147.
- [15] X. Gu, D. Liu, C. Baks, O. Tageman, B. Sadhu, J. Hallin, L. Rexberg, A. Valdes-Garcia, A multilayer organic package with 64 dual-polarized antennas for 28 GHz 5G communication, in: 2017 IEEE MTT-S International Microwave Symposium (IMS), 2017, June, pp. 1899–1901.
- [16] H2020-ICT-671650-mmMAGIC/D5.2, mmMAGIC Deliverable D5.2, Final multinode and multiantenna transmitter and receiver architectures and schemes, 2017.
- [17] P. Harris, W.B. Hasan, H. Brice, B. Chitambira, M. Beach, E. Mellios, A. Nix, S. Armour, A. Doufexi, An overview of massive MIMO research at the University of Bristol, in: Radio Propagation and Technologies for 5G (2016), 2016, Oct., pp. 1–5.
- [18] B.M. Hochwald, C.B. Peel, A.L. Swindlehurst, A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation, *IEEE Transactions on Communications* 53 (3) (2005, March) 537–544.
- [19] W. Hong, K.H. Baek, Y. Lee, Y. Kim, S.T. Ko, Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices, *IEEE Communications Magazine* 52 (9) (2014, September) 63–69.
- [20] A. Huebner, F. Schuehle, M. Bossert, E. Costa, H. Haas, A simple space-frequency coding scheme with cyclic delay diversity for OFDM, in: 2003 5th European Personal Mobile Communications Conference (Conf. Publ. No. 492), 2003, April, pp. 106–110.
- [21] IEEE Std 145-1993, IEEE Standard Definitions of Terms for Antennas, 1993.
- [22] ITU-R M.2135-1, Guidelines for evaluation of radio interface technologies for IMT-Advanced, 2009.
- [23] S.A. Jafar, A. Goldsmith, Transmitter optimization and optimality of beamforming for multiple antenna systems, *IEEE Transactions on Wireless Communications* 3 (4) (2004, July) 1165–1175.
- [24] M. Joham, W. Utschick, J.A. Nossek, Linear transmit processing in MIMO communications systems, *IEEE Transactions on Signal Processing* 53 (8) (2005, Aug.) 2700–2712.
- [25] J. Jose, A. Ashikhmin, T.L. Marzetta, S. Vishwanath, Pilot contamination and precoding in multicell TDD systems, *IEEE Transactions on Wireless Communications* 10 (8) (2011, August) 2640–2651.
- [26] K. Larsson, B. Halvarsson, D. Singh, R. Chana, J. Manssour, M. Na, C. Choi, S. Jo, High-speed beam tracking demonstrated using a 28 GHz 5G trial system, in: VTC 2017 Fall, 2017.
- [27] X. Li, E. Björnson, E.G. Larsson, S. Zhou, J. Wang, A multicell MMSE precoder for massive MIMO systems and new large system analysis, in: 2015 IEEE Global Communications Conference (GLOBECOM), 2015, Dec., pp. 1–6.
- [28] X. Li, E. Björnson, E.G. Larsson, S. Zhou, J. Wang, Massive MIMO with multicell MMSE processing: exploiting all pilots for interference suppression, *EURASIP Journal on Wireless Communications and Networking* (2017, June), <https://doi.org/10.1186/s13638-017-0879-2>.

- [29] D.J. Love, R.W. Heath, V.K.N. Lau, D. Gesbert, B.D. Rao, M. Andrews, An overview of limited feedback in wireless communication systems, *IEEE Journal on Selected Areas in Communications* 26 (8) (2008, October) 1341–1365.
- [30] R.J. Mailloux, *Phased Array Antenna Handbook*, second ed., Artech House, 2005.
- [31] T. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas, *IEEE Transactions on Wireless Communications* 9 (11) (2010, November) 3590–3600.
- [32] A. Paulraj, R. Nabar, D. Gore, *Introduction to Space–Time Wireless Communications*, Cambridge University Press, ISBN 9780521826150, 2003.
- [33] C.B. Peel, B.M. Hochwald, A.L. Swindlehurst, A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization, *IEEE Transactions on Communications* 53 (1) (2005, Jan.) 195–202.
- [34] 5G collaboration in lead up to 2018 winter sporting events, press release, available at <https://www.ericsson.com/en/news/2017/5/5g-collaboration-in-lead-up-to-2018-winter-games>, 2017, May.
- [35] IBM & Ericsson announce research advance for 5G communications networks, press release, available at <https://www.ericsson.com/en/press-releases/2017/2/ibm--ericsson-announce-research-advance-for-5g-communications-networks>, 2017, Feb.
- [36] Going Massive with MIMO, available at <https://www.ericsson.com/en/news/2018/1/massive-mimo-highlights>, 2018, Jan.
- [37] R1-1703536, Evaluation assumptions for Phase 2 NR MIMO system level calibration, 3GPP TSG RAN WG1 Meeting #88, 2017, February.
- [38] R1-1708688, Codebook design for Type II CSI feedback, 3GPP TSG RAN WG1 Meeting #89, 2017, May.
- [39] M. Sadek, A. Tarighat, A.H. Sayed, A leakage-based precoding scheme for downlink multiuser MIMO channels, *IEEE Transactions on Wireless Communications* 6 (5) (2007, May) 1711–1721.
- [40] B. Sadhu, Y. Tousi, J. Hallin, S. Sahl, S. Reynolds, Ö. Renström, K. Sjögren, O. Haapalahti, N. Mazor, B. Bokinge, G. Weibull, H. Bengtsson, A. Carlinger, E. Westesson, J.E. Thillberg, L. Rexberg, M. Yeck, X. Gu, D. Friedman, A. Valdes-Garcia, A 28 GHz 32-element phased-array transceiver IC with concurrent dual polarized beams and 1.4 degree beam-steering resolution for 5G communication, in: 2017 IEEE International Solid-State Circuits Conference (ISSCC), 2017, Feb., pp. 128–129.
- [41] H. Sampath, Linear precoding and decoding for multiple input multiple output (MIMO) wireless channels, Ph.D. thesis, Stanford University, 2001.
- [42] A. Simonsson, M. Thurfjell, B. Halvarsson, J. Furuskog, S. Wallin, S. Itoh, H. Murai, D. Kurita, K. Tateishi, A. Harada, Y. Kishiyama, Beamforming gain measured on a 5G test-bed, in: 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), 2017, June, pp. 1–5.
- [43] V. Tarokh, H. Jafarkhani, A.R. Calderbank, Space–time block codes from orthogonal designs, *IEEE Transactions on Information Theory* 45 (5) (1999, Jul.) 1456–1467.
- [44] S.H. Tsai, Transmit equal gain precoding in Rayleigh fading channels, *IEEE Transactions on Signal Processing* 57 (9) (2009, Sept.) 3717–3721.
- [45] S.H. Tsai, An equal gain transmission in MIMO wireless communications, in: 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, 2010, Dec., pp. 1–5.
- [46] D. Tse, P. Viswanath, *Fundamentals of Wireless Communications*, 2005.
- [47] A.M. Tulino, A. Lozano, S. Verdu, Capacity-achieving input covariance for single-user multiantenna channels, *IEEE Transactions on Wireless Communications* 5 (3) (2006, March) 662–671.
- [48] D. Tuninetti, On the capacity of the AWGN MIMO channel under per-antenna power constraints, in: 2014 IEEE International Conference on Communications (ICC), 2014, June, pp. 2153–2157.
- [49] M.K. Varanasi, T. Guess, Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel, in: *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers* (Cat. No.97CB36136), vol. 2, 1997, Nov., pp. 1405–1409.
- [50] J. Vieira, F. Rusek, O. Edfors, S. Malkowsky, L. Liu, F. Tufvesson, Reciprocity calibration for massive MIMO: proposal, modeling, and validation, *IEEE Transactions on Wireless Communications* (ISSN 1536-1276) 16 (5) (2017, May) 3042–3056.
- [51] H. Weingarten, Y. Steinberg, S.S. Shamai, The capacity region of the Gaussian multiple-input multiple-output broadcast channel, *IEEE Transactions on Information Theory* 52 (9) (2006, Sept.) 3936–3964.
- [52] B. Yu, K. Yang, C.Y.D. Sim, G. Yang, A novel 28 GHz beam steering array for 5G mobile device with metallic casing application, *IEEE Transactions on Antennas and Propagation* 66 (1) (2018, Jan.) 462–466.
- [53] X. Zheng, Y. Xie, J. Li, P. Stoica, MIMO transmit beamforming under uniform elemental power constraint, in: 2007 IEEE 8th Workshop on Signal Processing Advances in Wireless Communications, 2007, June, pp. 1–5.

CHANNEL CODING

8

Forward error correction (FEC) schemes play a fundamental role in every digital communication system, because they provide robustness against noise and other channel uncertainties (e.g., imperfect channel-state information). The large range of use cases envisaged for 5G NR (see Chapter 1) makes the problem of designing good FEC schemes particularly challenging. Indeed, together with the traditional *enhanced mobile broadband* (eMBB) use case, which involves providing a high data rate to mobile users, 5G NR will also support two new use cases: *massive machine-type communication* (mMTC), which aims at guaranteeing connectivity to a very large number of low-cost and low-energy devices, and *ultrareliable low-latency communications* (URLLC), which deals with providing connectivity with latency and reliability levels that are orders of magnitude more stringent than in eMBB and mMTC [28]. As we shall see, the FEC schemes that are optimal for data transmission in eMBB may be suboptimal in URLLC applications, where the latency constraint imposes severe restrictions on the code blocklength. Furthermore, whereas the problem of designing good FEC schemes for large blocklength values is well understood both from a theoretical and a practical implementation viewpoint, designing good FEC schemes for short blocklength is—as we shall see—still an area of research.

In this chapter, we shall provide an introduction to FEC schemes for 5G NR. Focusing first on a simple binary-input additive white Gaussian noise (AWGN) channel, we shall first characterize the fundamental trade-off between rate, packet error probability, and blocklength using recently developed nonasymptotic tools in information theory [29].

We shall then review the family of low-density parity-check (LDPC) and polar codes that have recently been selected for data and control channel, respectively, in 5G new-radio (NR) eMBB transmission [1], and benchmark their performance against the theoretical bounds given by nonasymptotic information theory. We will also review other FEC schemes that are competitive for short blocklengths and may be relevant for URLLC applications.

Finally, looking beyond what is currently standardized in NR, we shall consider the problem of communicating over multiantenna fading channels, and we shall use nonasymptotic information theory to shed light on the role of frequency and spatial diversity (see Chapter 1), with specific focus on the URLLC use case. We shall see that the use of space-frequency codes is necessary to achieve the required ultrahigh reliability. All the performance results reported in this chapter can be reproduced using the numerical routines described in [10,20], which are available online.

8.1 FUNDAMENTAL LIMITS OF FORWARD ERROR CORRECTION

8.1.1 THE BINARY AWGN CHANNEL

We shall focus for simplicity of exposition on the binary-input AWGN (bi-AWGN) channel, i.e., a memoryless discrete-time AWGN channel

$$y_k = \sqrt{\rho}x_k + w_k, \quad k = 1, \dots, n, \quad (8.1)$$

whose input symbols $\{x_k\}$ belong to the binary alphabet $\{-1, 1\}$. We assume that the additive noise samples $\{w_k\}$ are independent and identically distributed zero mean, unit variance Gaussian random variables. Hence, ρ can be thought of as the signal-to-noise ratio (SNR). Finally, n in (8.1) denotes the number of discrete-time channel uses that can be employed to transmit a packet of information bits.

8.1.2 CODING SCHEMES FOR THE BINARY-AWGN CHANNELS

We shall next review the fundamental limits on the rate at which we can communicate over this channel for a given latency (expressed in terms of number of channel uses) and reliability constraint. To do so, we first introduce the concept of a channel coding scheme for the channel (8.1). We shall focus on coding schemes with codewords whose entries belong to the binary field $\mathbb{F}_2$ and assume that each binary coded symbol c_k is mapped into the binary phase-shift-keying (BPSK) symbol $x_k = 2c_k - 1$ in the Euclidean space.

Definition 1. A (n, M, ϵ) (binary) coding scheme for the channel (8.1) consists of:

- An encoder $f : \{1, 2, \dots, M\} \mapsto \mathbb{F}_2^n$ that maps the information message $j \in \{1, 2, \dots, M\}$, where $M = 2^k$ and k is the number of information bits, to a codeword in the set $\{\mathbf{c}_1, \dots, \mathbf{c}_M\}$, with $\mathbf{c}_m \in \mathbb{F}_2^n$, $m = 1, \dots, M$. The set of M codewords is commonly referred to as the channel code or the codebook.
- A decoder $g : \mathbb{R}^n \mapsto \{1, 2, \dots, M\}$ that maps the received sequence $\mathbf{y} \in \mathbb{R}^n$ into a message $\hat{j} \in \{1, 2, \dots, M\}$, or, possibly, it declares an error. This decoder satisfies the average packet error probability constraint

$$\mathbb{P}\{\hat{j} \neq j\} \leq \epsilon. \quad (8.2)$$

□

The rate R of a (n, M, ϵ) coding scheme is $R = \log_2(M)/n = k/n$. A remark on terminology is at this point appropriate. Following [26], we differentiate between a *code* (i.e., the list of codewords) and a *coding scheme* (i.e., the code together with the encoder and the decoder), because a given code can be decoded using different decoding algorithms (often of drastically different complexity), yielding different coding schemes. We warn the reader that this distinction is frequently omitted in the coding-theory literature.

8.1.3 PERFORMANCE METRICS

We define the maximum coding rate $R^*(n, \epsilon)$ as the largest rate achievable with (n, M, ϵ) coding schemes. Mathematically,

$$R^*(n, \epsilon) = \sup \left\{ \frac{\log_2(M)}{n} : \exists (n, M, \epsilon) \text{ coding scheme} \right\}. \quad (8.3)$$

Determining $R^*(n, \epsilon)$, which describes the fundamental trade-off between rate, blocklength, and packet error probability in the transmission of information, is a fundamental problem in information theory, with a long history. In a seminal and groundbreaking contribution, Shannon characterized the asymptotic behavior of $R^*(n, \epsilon)$ in the limit $n \rightarrow \infty$ for general memoryless channels [35]. Specifically, he showed that, for every $0 < \epsilon < 1$, the maximum coding rate $R^*(n, \epsilon)$ converges to a constant C —the so-called channel capacity—that depends on the characteristics of the channel. The consequence of this result is that, for every transmission rate R less than C , there exists a sequence of coding schemes with rate R and vanishing packet error probability as $n \rightarrow \infty$. Conversely, one can show that if $R > C$, then the packet error probability over most memoryless channels of practical relevance (including the bi-AWGN channel (8.1)) goes to 1. This means that reliable communication is possible only at rates less than C .

For the bi-AWGN channel in (8.1), the channel capacity is given by

$$C = \frac{1}{\sqrt{2\pi}} \int e^{-z^2/2} \left(1 - \log_2(1 + e^{-2\rho + 2z\sqrt{\rho}}) \right) dz. \quad (8.4)$$

The proof of Shannon's coding theorem is based on a random-coding argument, and it does not provide a constructive way to approach the channel capacity. Indeed, it took about 50 years from the publication of Shannon's paper for the coding community to demonstrate near-capacity performance with practical coding schemes [11]. We shall review some of these coding schemes in Section 8.2.

It is worth stressing at this point that the channel capacity C is an asymptotic performance metric describing the behavior of the maximum coding rate $R^*(n, \epsilon)$ in the limit $n \rightarrow \infty$. This means that capacity cannot be used to benchmark the performance of coding schemes in which the blocklength n is short, as it is expected in some 5G use cases, due, for example, to a latency constraint.

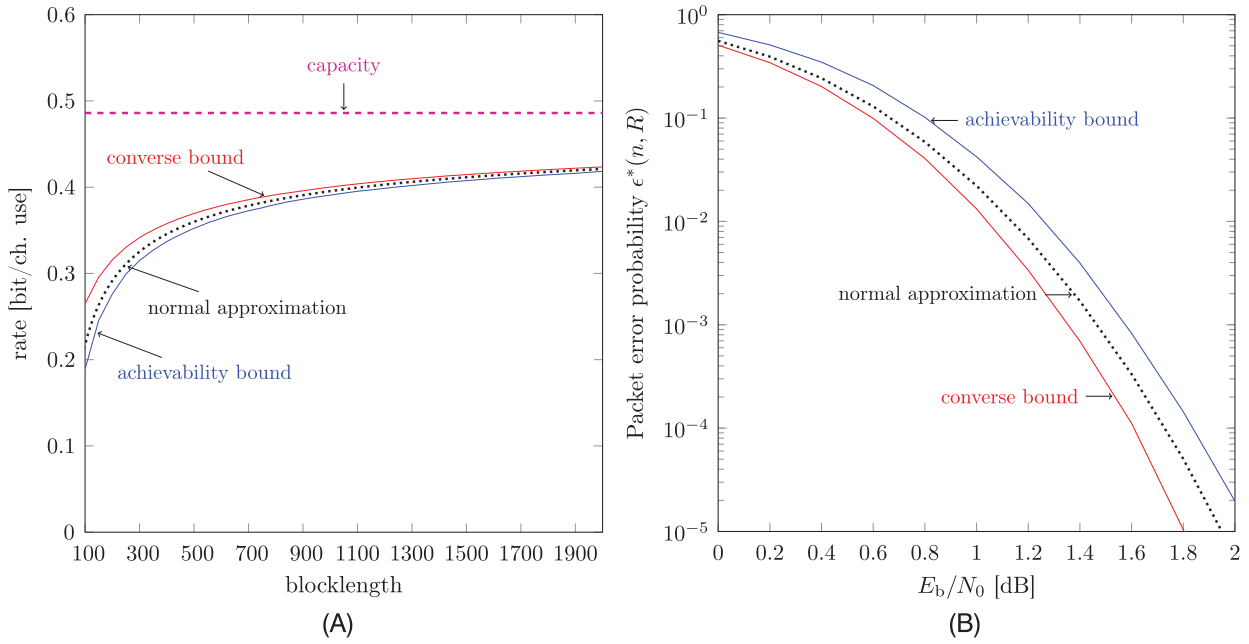
This observation has renewed the interest in nonasymptotic characterizations of the maximum coding rate $R^*(n, \epsilon)$. The exact computation of such a quantity is a formidable task unless the number M of codewords is very small (e.g., $M = 4$; see [39]). However, tight upper (converse) and lower (achievability) bounds on $R^*(n, \epsilon)$ that can be computed efficiently can be obtained for a variety of channels for practical interest for 5G, including the bi-AWGN channel (8.1), using the finite-blocklength information-theoretic tools recently introduced in [29].

Such upper and lower bounds are depicted in Fig. 8.1 for the bi-AWGN channel. Specifically, Fig. 8.1A illustrates the tightest known bounds as a function of the blocklength n , for a target error probability of 10^{-4} and $\rho = 0$ dB. Equivalently, we can use the bounds to study the minimum packet error probability $\epsilon^*(n, R)$ achievable for a fixed blocklength n and rate R :

$$\epsilon^*(n, R) = \min \left\{ \epsilon : \exists (n, \lceil 2^{nR} \rceil, \epsilon) \text{ coding scheme} \right\}. \quad (8.5)$$

This is illustrated in Fig. 8.1B where the bounds on $\epsilon^*(n, R)$ are plotted as a function of the minimum energy per bit E_b normalized with respect to the noise power spectral density N_0 , which for the bi-AWGN channel, is given by

$$\frac{E_b}{N_0} = \frac{\rho}{2R}. \quad (8.6)$$

**FIGURE 8.1**

Bounds on the maximum coding rate $R^*(n, \epsilon)$ and on the minimum probability of error $\epsilon^*(n, R)$ achievable on the bi-AWGN channel (8.1). (A) $R^*(n, \epsilon)$ as a function of n for $\epsilon = 10^{-4}$, $\rho = 0$ dB. (B) $\epsilon^*(n, R)$ as a function of E_b/N_0 for $R = 0.5$, $n = 512$.

In both figures, the converse bound (which is an upper bound on $R^*(n, \epsilon)$ and a lower bound on $\epsilon^*(n, R)$) is based on the minimax converse [29, Thm. 27] (see [14] for details). The achievability bound (which is a lower bound on $R^*(n, \epsilon)$ and an upper bound on $\epsilon^*(n, R)$) is the random-coding union bound with parameter s (RCUs) [23, Thm. 1], a relaxation of the RCU bound [29, Thm. 16] that lends itself to efficient numerical evaluation. The dotted curve in Fig. 8.1A is the so-called normal approximation [29, Eq. (223)] to the maximum coding rate $R^*(n, \epsilon)$, which is given by

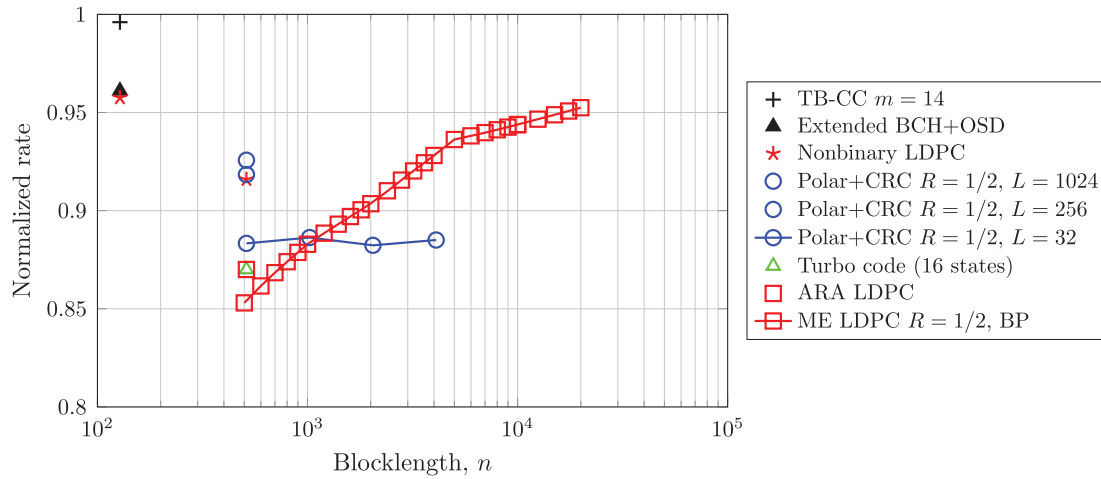
$$R^*(n, \epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + \frac{1}{2n} \log_2 n. \quad (8.7)$$

Here, V is the channel dispersion, which, for the bi-AWGN channel, is

$$V = \frac{1}{\sqrt{2\pi}} \int e^{-z^2/2} \left(1 - \log_2(1 + e^{-2\rho + 2z\sqrt{\rho}}) - C \right)^2 dz, \quad (8.8)$$

and $Q^{-1}(\cdot)$ is the inverse of the Gaussian Q -function. In Fig. 8.1B, the normal approximation is

$$\epsilon^*(n, R) \approx Q\left(\frac{C - R + (2n)^{-1} \log_2 n}{\sqrt{V/n}}\right). \quad (8.9)$$

**FIGURE 8.2**

Normalized rate R_{norm} of some of the coding schemes reviewed in this chapter when used over the bi-AWGN channel (8.1). Here, $\epsilon = 10^{-4}$.

As can be seen from Fig. 8.1, the bounds characterize tightly the nonasymptotic performance metrics of interest, i.e., $R^*(n, \epsilon)$ and $\epsilon^*(n, R)$. Furthermore, the normal approximation, which is simple to evaluate numerically, turns out to be accurate in the parameter range considered in the figure. It is worth highlighting that this approximation is typically inaccurate in scenarios where both the required packet error probability and the SNR are low. In such scenarios, approximations based on saddle-point methods [23] should be used instead.

As we shall see in Section 8.3, the nonasymptotic bounds depicted in Fig. 8.1 can be generalized to include the presence of fading, and the use of pilot-aided transmission, and multiple antennas. Such generalizations provide valuable insights on how to optimally communicate over multiantenna fading channels.

The nonasymptotic bounds in Fig. 8.1 provide a natural way to benchmark the performance of practical coding schemes, which is more informative than the classic error probability vs. E_b/N_0 curves. Specifically, one fixes a code of a given blocklength n and rate R , and determines the minimum SNR $\rho_{\min}(\epsilon)$ needed to achieve a target error probability ϵ . Then one defines the normalized rate

$$R_{\text{norm}} = \frac{R}{R^*(n, \epsilon)} \quad (8.10)$$

where the maximum coding rate $R^*(n, \epsilon)$ can be evaluated, for example, using the normal approximation (8.7) with ρ in (8.4) and (8.8) replaced by $\rho_{\min}(\epsilon)$. Note that R_{norm} is a normalized performance metric that allows one to compare coding schemes of different blocklengths and different rates. The larger R_{norm} , the better the coding scheme.

Fig. 8.2 summarizes the performance over the bi-AWGN channel of some of the coding schemes we will review in the next sections of this chapter. As we shall see, there exist three different regimes in which different coding schemes are preferable: the short-blocklength regime, the moderate-blocklength regime, and the large-blocklength regime.

In the large-blocklength regime ($n \geq 1000$), modern codes that are decoded with belief propagation, such as multiedge-type LDPC codes and turbo codes, are the most competitive solutions. At moderate blocklengths ($400 \leq n \leq 1000$), good performance can be achieved using polar codes decoded with successive-cancellation decoding with a large list size and combined with an outer cyclic-redundancy check (CRC) code. Finally, in the short-blocklength regime ($n \leq 400$), some of the most promising solutions involve the use of short algebraic codes or linear block codes based on tail-biting trellises, decoded using near-maximum-likelihood (ML) decoding algorithms such as ordered-statistic decoding (OSD) [17], or LDPC codes over high-order finite fields. These insights have been taken into account in the 3GPP standardization activities. Indeed, LDPC codes will be used to protect the new-radio (NR) eMBB data channel and polar codes to protect the NR eMBB control channel [1].¹

This confirms that nonasymptotic information-theoretic analyses provide concrete and useful guidelines on the design and the selection of actual coding schemes. We will see further examples of this principle in Section 8.3.

8.2 FEC SCHEMES FOR THE BI-AWGN CHANNEL

8.2.1 INTRODUCTION

Designing codes for large blocklengths is a well-investigated problem and effective solutions are available. Indeed, modern codes (e.g., turbo and LDPC codes) offer excellent performance under suboptimal but low-complexity iterative decoding algorithms such as belief propagation (BP). The design problem is more open for short blocklengths. On the one hand, the BP decoding performance becomes increasingly unsatisfactory when the blocklength decreases; on the other hand, a reduction in the blocklength makes it feasible to use near-maximum-likelihood decoding algorithms, which, when applied to, e.g., classical algebraic codes, yield performances that are sometimes superior to what can be achieved by modern codes with BP decoding.

In this section, we shall review some of the code constructions that are of interest for NR. Our emphasis will be mainly on the short- and moderate-blocklength regimes. For simplicity, we shall focus exclusively on the bi-AWGN channel. Also, in the spirit of the book, we shall highlight the general principles of each coding scheme, without delving too much into the many additional features (e.g., rate flexibility and suitability to hybrid automatic-repetition-request protocols) that are required for a coding scheme to be compatible with the requirements set in NR. Since our focus is on the bi-AWGN channel, we will not discuss coded-modulation techniques. We just highlight that the approach used in LTE to map coded bits into modulation symbols, which relies on bit-interleaved coded modulation (BICM), is suboptimal because it yields a well-known shaping loss for high-order constellations. This is particularly relevant for NR, which will support constellations belonging to sets of cardinality as large as 256. A different approach, which is becoming increasingly popular in fiber-optic applications, is to use the probabilistic shaping method proposed in [9]—an ingenious technique that provides rate adaptation and reduces significantly the shaping loss.

¹The standardization of the coding schemes to be used in the other use cases is still ongoing.

8.2.2 SOME DEFINITIONS

We start our review by collecting here some standard definitions concerning linear block codes that will turn out to be useful in the remainder of this chapter (see, e.g., [31] for more details).

We say that the list of codewords $\mathcal{C}$ of an $(n, 2^k, \epsilon)$ binary coding scheme (see Definition 1) is a (n, k) *linear block code* if the 2^k codewords are a k -dimensional subspace of the vector space of all binary n -tuples. Here, addition and multiplication are the ones of the binary field $\mathbb{F}_2$. It follows by this definition that the codewords of a (n, k) linear block code can be expressed as a linear combination of k linearly independent codewords $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_k$. In other words, the set $\{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_k\}$ forms a basis for $\mathcal{C}$. We can use this basis to perform encoding as follows. Let the so-called $k \times n$ *generator matrix* of the code be defined as²

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_k \end{bmatrix}. \quad (8.11)$$

Then the encoder output $\mathbf{c} = f(j)$ corresponding to the input message $j \in \{1, \dots, 2^k\}$ is

$$\mathbf{c} = \mathbf{b}\mathbf{G} \quad (8.12)$$

where $\mathbf{b}$ is the k -dimensional binary representation of j .

The $(n - k)$ -dimensional dual space $\tilde{\mathcal{C}}$ of $\mathcal{C}$ is the set of all binary n -tuples $\tilde{\mathbf{c}}$ satisfying

$$\tilde{\mathbf{c}}\mathbf{c}^T = 0, \quad \forall \mathbf{c} \in \mathcal{C}. \quad (8.13)$$

Let $\{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{n-k}\}$ be a basis of $\tilde{\mathcal{C}}$. The parity check matrix (PCM) $\mathbf{H}$ of $\mathcal{C}$ is the $(n - k) \times n$ binary matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_{n-k} \end{bmatrix}. \quad (8.14)$$

Note that if $\mathbf{c} \in \mathcal{C}$ then

$$\mathbf{c}\mathbf{H}^T = \mathbf{0}. \quad (8.15)$$

This highlights the important role of the PCM for error detection at the decoder. A linear block code is defined uniquely by the matrices $\mathbf{G}$ and $\mathbf{H}$.

For a given (n, k) linear block code, let $\mathcal{X} = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(2^k)\}$ be the set of BPSK coded sequences in the n -dimensional Euclidean space corresponding to the 2^k codewords. Under the assumption that the information message j is drawn uniformly from $\{1, 2, \dots, 2^k\}$, the decoding rule that minimizes the packet error probability ϵ is the maximum-likelihood (ML) decoding rule

²Following the standard convention in coding theory, all vectors in the remainder of the chapter are row vectors.

$$\hat{j} = \arg \max_{m \in \{1, \dots, 2^k\}} p(\mathbf{y} | \mathbf{x}(m)) \quad (8.16)$$

where $p(\mathbf{y} | \mathbf{x})$ denotes the channel law (8.1),

$$p(\mathbf{y} | \mathbf{x}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\|\mathbf{y} - \sqrt{\rho}\mathbf{x}\|^2}{2}\right). \quad (8.17)$$

It follows from (8.17) that the ML decoding rule (8.16) can be equivalently expressed as

$$\hat{j} = \arg \min_{m \in \{1, \dots, 2^k\}} \|\mathbf{y} - \sqrt{\rho}\mathbf{x}(m)\|^2. \quad (8.18)$$

Stated explicitly, the ML decoder selects the message whose corresponding BPSK coded sequence is closest to the received signal $\mathbf{y}$ in the Euclidean space.

Assume that the coded sequence $\mathbf{x}(1)$, corresponding to $j = 1$, is transmitted. The probability that a different coded sequence $\mathbf{x}(\ell)$ with $\ell \neq 1$ is closer to the received signal $\mathbf{y}$ than $\mathbf{x}(1)$ depends on the Euclidean distance between $\mathbf{x}(\ell)$ and $\mathbf{x}(1)$. Specifically,

$$\mathbb{P}\{\hat{j} = \ell | j = 1\} = Q\left(\frac{\sqrt{\rho}\|\mathbf{x}(\ell) - \mathbf{x}(1)\|}{2}\right). \quad (8.19)$$

Let now K_d be the average number of coded sequences that have a neighbor at distance d . It follows from (8.19) and from an application of the union bound that [16, Eq. (2.32)]

$$\epsilon \leq \sum_d K_d Q\left(\frac{\sqrt{\rho}d}{2}\right). \quad (8.20)$$

The upper bound in (8.20) is typically dominated by the first term

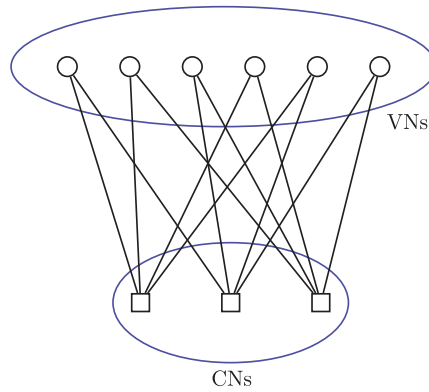
$$K_{d_{\min}} Q\left(\frac{\sqrt{\rho}d_{\min}}{2}\right) \quad (8.21)$$

where $d_{\min}$ is the minimum Euclidean distance between any two coded sequences. This quantity is equal to twice the minimum Hamming weight of the nonzero codewords in $\mathcal{C}$. The bound (8.20) highlights the dependence of the packet error probability ϵ on the distance spectrum of the code. More sophisticated and tighter bounds on the packet error probability of linear block codes under ML decoding are described in [33]. It is worth highlighting that the evaluation of the ML rule (8.18) is in practice unfeasible already for values of k larger than a few tens of bits because of the complexity, unless the code possesses structures that facilitate it.

8.2.3 LDPC CODES

8.2.3.1 Fundamentals of LDPC Codes

LDPC codes are a class of linear block codes characterized by a PCM that is sparse, i.e., it contains only few nonzero entries. Originally proposed by Gallager [18] in the 1960s, and later rediscovered and generalized in the 1990s [22, 5], LDPC codes provide a performance close to capacity for a large set of communication channels. These codes are currently deployed in several standards, including

**FIGURE 8.3**

Tanner graph of the linear code with parity-check matrix given in (8.22).

IEEE802.11n, IEEE802.16e (WiMAX), IEEE 802.11ad (WiGig) and DVB-S2. As we shall briefly review, the sparseness of the PCM enables the use of low-complexity iterative decoding algorithms, which provide often near-ML performance. Throughout this section, we shall focus on binary LDPC codes. Extensions to higher fields are discussed in Section 8.2.5.3.

An (n, k) binary LDPC code is defined in terms of a $m \times n$ sparse PCM. Here, $m \geq n - k$, which implies that the PCM may be rank deficient. A convenient way to represent the PCM is through its Tanner graph representation [37]. A Tanner graph of an LDPC code is a bipartite graph, i.e., a graph in which the nodes are of two different types, and the edges connect only nodes of different types. These two types of nodes are commonly referred to by variable nodes (VNs), which are as many as the codeword length n , and by check node (CN), which are as many as the number of rows m in the PCM, i.e., as many as the parity-check equations. The Tanner graph is constructed from the PCM by drawing an edge between the i th check node and the j th variable node whenever the entry $[\mathbf{H}]_{ij}$ of the PCM contains a 1. As an example, the linear block code with PCM given by

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix} \quad (8.22)$$

is equivalently described by the Tanner graph depicted in Fig. 8.3.

We say that a LDPC code is regular if all VNs have the same degree, i.e., they are connected to the same number of edges, and the same occurs for the CNs. In the example of Fig. 8.3, all VNs have degree 2 and all CNs have degree 4. Allowing for VNs and CNs of different degrees, which results in irregular LDPC codes, turns out to be beneficial from a performance viewpoint [30]. Irregular LDPC codes are typically described in terms of their degree distributions, which provide the fraction of all edges connected to VNs/CNs of each degree.

Coarsely speaking, the decoding of LDPC codes is an iterative process, often referred to as BP, in which log-likelihood ratios (LLRs) about the coded bits are exchanged along the edges of the Tanner graph. Each decoding iteration consists of two phases: a first phase in which, at each VN, the LLRs from the channel and from the upcoming edges are processed and transformed in updated LLRs sent to

the neighboring CNs; and a second phase in which the LLRs arriving at each CN from the neighboring VNs is processed and updated LLRs are sent back. This process is repeated until a codeword is found or the maximum number of iterations is exceeded. The key observation is that the processing at the CNs and VNs depends only on locally available information, which allows for an efficient and parallelizable decoding process. However, since the decoding process is local, the globally optimal ML solution may not be found through this procedure, especially in the presence of short cycles in the Tanner graph. Indeed, such cycles constrain the decoding process to remain local. This observation also shows the importance of the low-density assumption, which facilitates the design of Tanner graphs free of short cycles.

One way to design LDPC codes is to use a pseudorandom algorithm that constructs a PCM with given degree distributions and avoids short cycles. Such an approach, despite yielding LDPC codes with extremely good performance [30], is impractical from an implementation viewpoint, because the absence of further structures in the PCM makes both the encoding and the decoding complexity too high for practically relevant blocklengths and rates.

A more practically appealing approach is to design structured LDPC codes constructed from a smaller protograph. The LDPC codes that one obtains through this construction form a subset of the more general class of MET-LDPC codes, whose performances are illustrated in Fig. 8.2. The PCM of protograph-based LDPC codes can be specified in terms of a small base matrix. The actual PCS is constructed from the base matrix by replacing each entry in the base matrix by a $Q \times Q$ binary matrix whose rows and columns have a weight equal to the corresponding entry in the base matrix. Here, Q is the so-called *lifting factor*. It is particularly convenient to pick as binary matrix a $Q \times Q$ cyclic permutation matrix, whose row and column weights are one. The resulting code is quasi-cyclic, a property that allows for simplified encoding and decoding, with negligible performance loss [21,31]. As an example, consider the following base matrix $\mathbf{H}_b$:

$$\mathbf{H}_b = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad (8.23)$$

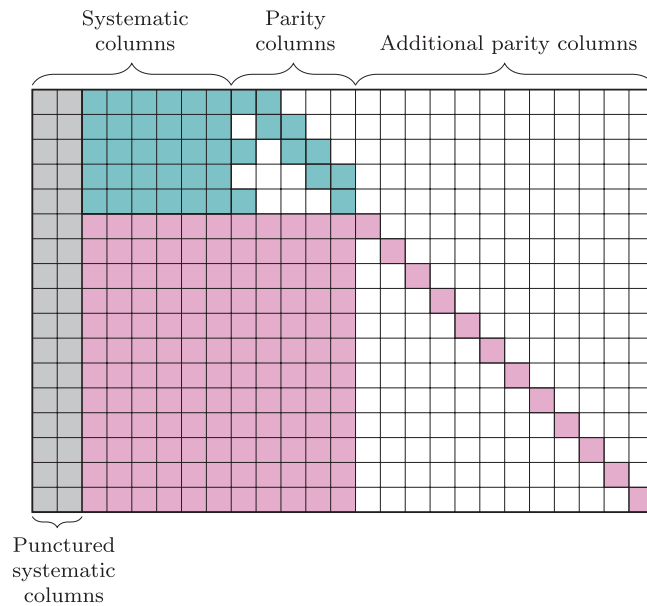
and assume that 3×3 cyclic permutation matrices are used to extend the base matrix to a PCM. Such a PCM may have the following structure:

$$\mathbf{H} = \left[\begin{array}{ccc|ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right]. \quad (8.24)$$

8.2.3.2 The LDPC-Code Solution Chosen for 5G NR

The LDPC codes chosen for the data channel in 5G NR are quasi-cyclic and have a rate-compatible structure that facilitates their use in hybrid automatic-repetition-request (HARQ) protocols.³

³This section is largely based on [32], where further information about the LDPC-code family chosen for 5G NR, including the rate-matching procedure and their use in combination with HARQ, can be found.

**FIGURE 8.4**

The general structure of the base matrix used in the quasi-cyclic LDPC codes selected for the data channel in NR.

To cover the large range of information payloads and rates that need to be supported in 5G NR, two different base matrices are specified. The general structure of these base matrices is provided in Fig. 8.4. In the figure, each white square represents a zero in the base matrix and each nonwhite square represents a one. The first two columns in gray correspond to punctured systematic bits that are actually not transmitted. Their addition is known to improve the threshold of the resulting code, i.e., its minimum SNR operating point [13]. The blue (dark gray in print version) part constitutes the kernel of the base matrix, and it defines a high-rate code. The dual-diagonal structure of the parity subsection of the kernel enables efficient encoding. Transmission at lower code rates is achieved by adding additional parity bits, i.e., by including an appropriately chosen subset of rows and columns in the base matrix containing entries marked in pink (light gray in print version). To enable maximum parallelism, the rows of the base matrix outside the kernel are designed so as to be orthogonal or quasi-orthogonal. The maximum lift factor $Q_{\max}$ is 384. This number is chosen to trade optimally between the parallel processing opportunities enabled by a large Q and the performance loss in terms of threshold due to the resulting higher amount of structure.

The base matrix #1, which is optimized for high rates and long blocklengths, supports LDPC codes of a nominal rate between 1/3 and 8/9. This matrix is of dimension 46×68 and has 22 systematic columns. Together with a lift factor of 384, this yields a maximum information payload of $k = 8448$ bits (including CRC).

The base matrix #2 is optimized for shorter blocklengths and smaller rates. It enables transmissions at a nominal rate between 1/5 and 2/3, it is of dimension 42×52 , and it has 10 systematic columns. This implies that the maximum information payload is $k = 3840$.

The choice of each of the $Q \times Q$ circulant matrices to be substituted into the entries of the base matrix to form the full PCM is specified in [1]. These circulant matrices are selected so as to ensure

efficient encoding and to obtain, at the same time, a good performance in terms of both error floor and threshold.

It is worth pointing out that, as observed in [32], the base matrix #2 tends to yield lower-complexity decoding and should in general be used whenever the information payload k is less than 3840 and the rate is less than $2/3$, whereas base matrix #1 should be used in the rest of the parameter range. Two exceptions are the case $k \leq 308$ for which base matrix #2 should be used for all rates, and the case $R \leq 1/4$, for which the base matrix #2 should be used for all information-payload sizes k .

8.2.4 POLAR CODES

8.2.4.1 Fundamentals of Polar Codes

Polar codes, introduced by Arikan [6], are a class of linear block codes that provably achieve the capacity of memoryless symmetric channels, such as the bi-AWGN, with low encoding and decoding complexity, and a recursive structure that facilitates their hardware implementation.

To introduce the core idea behind polar codes, i.e., the so-called *channel polarization*, we start by observing that, among all bi-AWGN channels (8.1), there are two extreme types, for which the communication problem is trivial⁴:

- The perfect (noiseless) channel $y_k = \sqrt{\rho}x_k$.
- The useless channel $y_k = w_k$.

Uncoded transmission is sufficient to achieve the capacity of the first channel, whereas no information can be transmitted on the second channel. Arikan's polarization technique is a lossless and low-complexity method to convert any binary-input symmetric channel into a mixture of extremal binary-input channels.

Polarization is achieved through the *polar transform*, which operates as follows: given two copies of a binary-input channel $W : \mathbb{F}_2 \rightarrow \mathcal{Y}$, where $\mathcal{Y}$ denotes the output set ($\mathcal{Y} = \mathbb{R}$ for the bi-AWGN channel), the polar transform creates, under the assumption of uniformly distributed inputs, two new *synthetic channels* $W^- : \mathbb{F}_2 \rightarrow \mathcal{Y} \times \mathcal{Y}$ and $W^+ : \mathbb{F}_2 \rightarrow \mathcal{Y} \times \mathcal{Y} \times \mathbb{F}_2$, defined as follows:

$$W^-(y_1, y_2 | u_1) = \frac{1}{2} \sum_{u_2 \in \mathbb{F}_2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2), \quad (8.25)$$

$$W^+(y_1, y_2, u_1 | u_2) = \frac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2). \quad (8.26)$$

Here, $\oplus$ is the addition in $\mathbb{F}_2$. Such a transform is illustrated in Fig. 8.5. Note that the channel W^+ , which has input u_2 and output (y_1, y_2, u_1) , contains the channel W . Hence, its capacity under uniform inputs (which we shall assume throughout this section) is no smaller than the capacity of the channel W . Since the transformation

$$x_1 = u_1 \oplus u_2, \quad (8.27)$$

$$x_2 = u_2, \quad (8.28)$$

⁴See, e.g., [26, Ch. 14] and Telatar's plenary talk at the 2017 International Symposium on Information Theory <https://goo.gl/zQz6nB> for a more comprehensive introduction to polar codes.

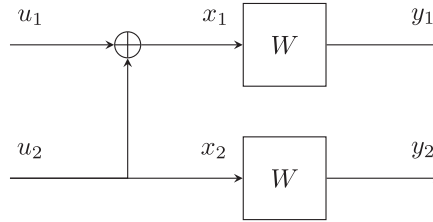


FIGURE 8.5

The polar transform.

is invertible, the sum of the capacities of W^+ and W^- must be equal to twice the capacity of W . This implies that the capacity of W^- must be smaller than that of the original channel W . To summarize, we started from two identical copies of W . Through the application of the polar transform we obtained two new synthetic channels: W^- , whose capacity is smaller than W , and W^+ , whose capacity is larger than W .

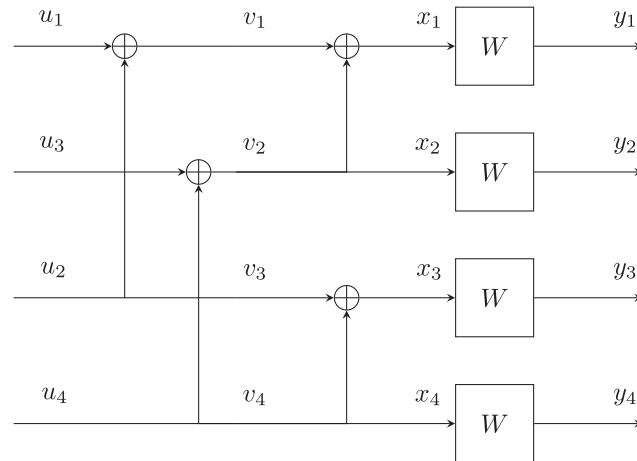
Example. Consider a binary erasure channel (BEC), i.e., a discrete memoryless channel with input–output relation

$$y = \begin{cases} x & \text{with prob. } 1 - p, \\ ? & \text{with prob. } p, \end{cases} \quad (8.29)$$

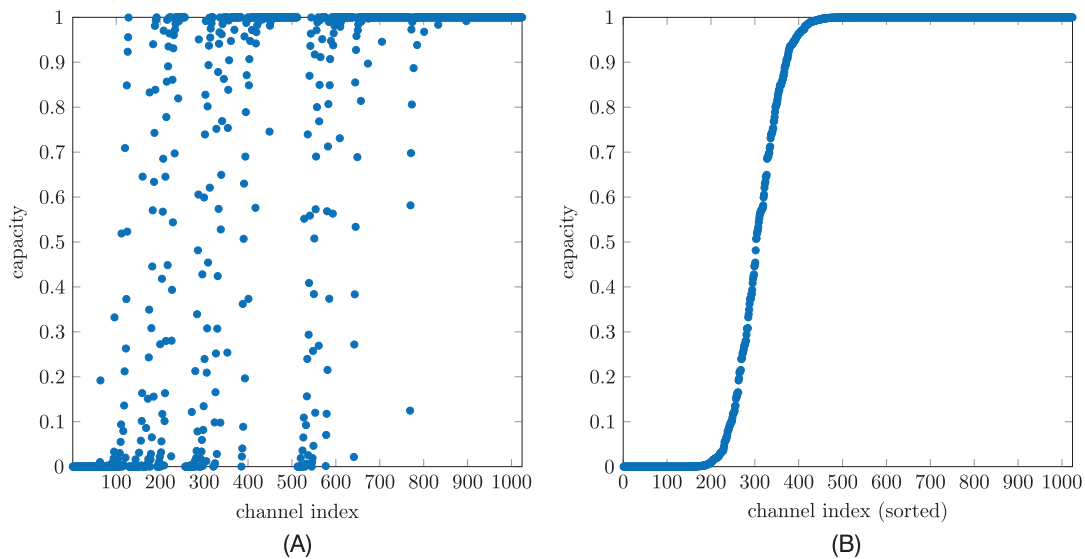
where x denotes the binary input and the symbol “?” denotes an erasure. In words, with probability $(1 - p)$ the input symbol x is received correctly, and with probability p it is erased. One can verify that the synthetic channels W^- and W^+ induced by the polar transform are also BECs, with erasure probability $p^- = p(2 - p) \geq p$ and $p^+ = p^2 \leq p$, respectively.

Some comments on the synthetic channels W^- and W^+ are in order. The channel W^- , which has input u_1 and output (y_1, y_2) , is indeed a genuine channel, because the receiver has access to both y_1 and y_2 . The channel W^+ , however, has u_1 as output, which is not available at the receiver. However, W^+ can be synthesized by imposing a decoding order. Namely, we first decode u_1 using y_1 and y_2 . Then we use y_1, y_2 , and the estimate $\hat{u}_1$ of u_1 to decode u_2 . This corresponds to successive-cancellation decoding. One can show that the block-error probability $\Pr\{(\hat{u}_1, \hat{u}_2) \neq (u_1, u_2)\}$ achievable with successive-cancellation decoding coincides with the one attainable by a genie-aided successive-cancellation decoder that uses u_1 instead of $\hat{u}_1$ in the second step of the decoding procedure. In other words, error propagation is not an issue if we measure performance in terms of block-error probability.

The polarization transform can now be applied again to the inputs of W^- and W^+ . This results in the four channels $W^{--}, W^{-+}, W^{+-},$ and W^{++} illustrated in Fig. 8.6. This process can be applied recursively N times to synthesize 2^N channels out of 2^N copies of W . Applying this process to a BEC with erasure probability $p = 0.3$ for the case $N = 10$ yields 1024 synthetic channels, whose capacities, which are equal to one minus the corresponding erasure probability, are illustrated in Fig. 8.7. This figure exemplifies the polarization phenomenon: roughly 30% of the synthetic channels have capacity close to zero, i.e., they are useless, whereas 70% have capacity close to 1, i.e., they are almost perfect. Observe now that the capacity of the underlying BEC is exactly $1 - p = 0.7$. Consequently, Fig. 8.7

**FIGURE 8.6**

The four synthetic channels obtained by applying twice the polar transform.

**FIGURE 8.7**

Capacity of the 1024 synthetic channels obtained by applying recursively the polar transform $N = 10$ times to a BEC with erasure probability p . (A) Unsorted. (B) Sorted.

suggests that, to achieve capacity, one can use a simple rate 0.7 binary code in which the k information bits are mapped to the almost perfect channels; the codeword entries corresponding to the remaining almost useless channels are frozen, i.e., they contain symbols known to the decoder (i.e., zeros).

More formally, we define a polar code as follows. Fix an integer N and let the blocklength be $n = 2^N$. Denote by $\mathbf{G} = \mathbf{F}^{\otimes n}$ the $n \times n$ polar transform matrix, where $\otimes$ stands for the Kronecker

product, and

$$\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad (8.30)$$

is the polar transform. Let also $\mathcal{U} \subset \{0, 1, \dots, 2^N - 1\}$ be a set of indices of cardinality k , where k is the size of the payload, and let $\mathcal{F}$ be the complementary set of size $n - k$. The k information bits are mapped to the entries belonging to $\mathcal{U}$ of a vector $\mathbf{u}$ of size n . The remaining entries, i.e., the ones with indices belonging to the set $\mathcal{F}$, are frozen to zero. In the remainder of the section, we shall refer to $\mathcal{F}$ as the frozen set. This set is constructed so as to contain the $n - k$ synthetic channels that have the smallest capacity. Finally, the resulting binary codeword $\mathbf{c}$ is obtained by performing the polarization mapping $\mathbf{c} = \mathbf{u}\mathbf{G}$.

The receiver uses successive-cancellation decoding; the decoding order is obtained by applying a bit-reversal permutation to the channel index set: specifically, we number the channels from 0 to $n - 1$, we obtain a new index for each channel by reversing the binary representation of its index, and we decode following the ordering induced by the new indices.

One can formally show [7] that the block-error probability ϵ achievable with the polar-coding scheme just described decays roughly as $2^{-\sqrt{n}}$, where $n = 2^N$ is the blocklength, for every rate below capacity. This proves that polar codes are indeed capacity achieving. Furthermore, they have the additional practical benefit over modern codes decoded with BP of not suffering from an error floor [24].

As far as complexity is concerned, the recursive nature of the polar transform allows one to perform encoding and decoding with a complexity that scales as $n \log_2 n$. Furthermore, these operations are naturally parallelizable.

One disadvantage of polar codes is that, when combined with successive-cancellation decoding, they offer mediocre performance in the short- and moderate-blocklength regimes that are of interest for control-channel applications. The performance of polar codes in this regime can be improved significantly by using instead a successive-cancellation list decoder [36]. This decoder considers simultaneously, at each decoding stage, L alternative paths, where L is the list size. The complexity of such a decoder is of order $Ln \log_2 n$. Performance can be further improved by concatenating the polar code with a high-rate CRC code. This code is used at the end of the list-decoding process to select the final codeword among the ones contained in the list. Concatenating a polar code with a CRC code turns out to be beneficial, because it improves the minimum distance of the polar code. As shown in Fig. 8.2, such an approach yields a state-of-the-art performance at moderate-blocklength values.

8.2.4.2 The Polar-Code Solution Chosen for 5G NR

Polar codes in 5G NR will be used to protect control signaling, with the exception of the transmission of payloads of up to $k = 11$ bits, for which Reed–Muller codes will be used, similar to 4G. Polar codes of different lengths and rates are supported. Such codes are obtained from an underlying parent code of length $2^{10} = 1024$. As recently summarized in [40], the specific polar code adopted in 5G NR relies on three innovations compared to what was described in the previous section, which allow the resulting coding scheme to satisfy the flexibility, processing latency, and complexity desiderata in 5G NR. Such innovations are:

- the offline computation of a deterministic reliability ordering of the synthetic channels;

- the use of parity-check bits to assist successive-cancellation list decoding also during its intermediate stages;
- a low-complexity rate-matching algorithm providing the needed rate and blocklength flexibility.

Deterministic Reliability Ordering

As explained in the previous section, one crucial step in the design of polar codes is the construction of the frozen-bit set $\mathcal{F}$. As shown in Fig. 8.7, the useless channels do not seem to follow a regular pattern. It turns out that, apart from the BEC, no efficient algorithm is known to rank the synthetic channels according to their reliability.⁵ Since the reliability of the synthetic channels and their relative ordering depends on time-varying parameters such as the SNR, adaptive offline and online algorithms, which are able to order the synthetic channels as a function of the current values of channel parameters, appear to be unfeasible because of latency and/or memory requirements. The solution adopted in 5G NR is to assign to each synthetic channel of the parent code a deterministic, i.e., channel-parameter-independent, polarization weight that expresses its reliability. The polarization weights induce an ordering on the subchannels that is prestored so as to avoid online computations. Using techniques such as the β -expansion in number theory [19,40], these polarization weights can be chosen so as to satisfy the natural universal partial ordering existing among the synthetic channels [34].

To save complexity, the polarization weights of the parent code are used also for transmission involving shorter codes of length $n = 2^N$, $N < 10$. Specifically, let $\mathbf{q}^{n_{\max}}$ be the vector containing the ordered sequence of the 1024 synthetic-channel indices of the parent code, ordered according to increasing polarization weights. Then a reliability index list $\mathbf{q}^n$ for the shorter code is obtained by removing from $\mathbf{q}^{n_{\max}}$ all indices larger or equal to n . Although this choice is suboptimal, the significant complexity reduction justifies the resulting moderate performance loss. The polarization weights in 5G NR have been chosen through extensive simulations to guarantee good performance for all blocklengths and code rates.

Parity-Check Coding

The polar-code structure selected in NR relies on the addition of a more general, yet hardware friendly, outer code than just a CRC. Specifically, n_{pc} parity-check bits are appended to the information payload. Some of them are assigned to the $k + n_{\text{pc}}$ unfrozen synthetic channels with lowest polarization weight. This improves the error performance of the code. Some of them are assigned to the unfrozen-bit positions in the vector $\mathbf{u}$ that correspond to the rows in the polar transform matrix $\mathbf{G}$ with smallest Hamming weights. This improves the distance spectrum of the resulting concatenated code.

To aid the intermediate steps of the successive-cancellation list-decoding algorithm, each parity-check bit is designed to depend only on preceding information bits. Specifically, the parity bits are computed through a length 5 shift register that evaluates the $\mathbb{F}_2$ -sum of information bits that are five positions apart; see [40, Algorithm 3] for details. Empirical evidence suggests that enforcing such spacing makes the resulting scheme robust against error propagation. The same shift-register architecture can be used by the successive-cancellation list decoder to prune all paths that result in an erroneous parity-check bit.

⁵See, however, [34,25,19] for recent progress on this problem, based on universal partial ordering, and the β -expansion in number theory.

Rate Adaptation

The polar-code construction reviewed so far allows one to generate codes of blocklength $n = 2^N$ for some integer $N \leq 10$. Additional blocklength values can be obtained by *puncturing* or *shortening* [8].

In puncturing, one transforms the original (n, k) polar code into a $(n - p, k)$ polar code by removing $p < n - k$ bits from each codeword $\mathbf{c}$. At the receiver, the channel LLRs corresponding to the punctured coded bits are set to 0 (the bits are assumed erased), and then the decoder of the original (n, k) polar code is applied. The presence of p zero-valued channel LLRs induces, through the decoding process, zero-valued LLRs for p of the entries of the n -dimensional input vector $\mathbf{u}$. To avoid poor performance, these entries must be set to frozen bits. This can be achieved by puncturing the bits of $\mathbf{c}$ with indices corresponding to the reverse-bit permutations of $\{1, 2, \dots, p\}$ and by freezing the corresponding bits of $\mathbf{u}$. The remaining frozen bits are chosen among the ones with lowest polarization weight, as usual.

We next discuss shortening. Let us assume that the original (n, k) code is systematic. Shortening allows one to obtain a $(n - p, k - p)$ code, $p < k$, by setting p systematic bits to zero and by not transmitting them. At the decoder side, the channel LLRs corresponding to the punctured systematic bits are set to infinity, and the decoder of the original code is applied. For nonsystematic polar codes, a natural way to select the coded bits to shorten is by requiring them to be linear combinations of frozen bits. This can be achieved by shortening the bits of $\mathbf{c}$ with indices corresponding to the reverse-bit permutations of $\{n - p + 1, \dots, n\}$ and by freezing the corresponding bits of $\mathbf{u}$. As before, the remaining frozen bits are chosen among the ones with lowest polarization weight.

8.2.5 OTHER CODING SCHEMES FOR THE SHORT-BLOCKLENGTH REGIME

To conclude our overview, we shall next present a selection of additional coding schemes that exhibit a favorable performance/complexity trade-off in the short-blocklength regime ($n < 400$). Even though these schemes are not standardized in 5G, some of them may enter future releases due to their suitability for URLLC.

8.2.5.1 Short Algebraic Linear Block Codes With Ordered-Statistics Decoding

As already mentioned, when the blocklength is short, classic algebraic codes such as BCH and extended BCH codes can be decoded using near ML decoding algorithms. OSD, which we shall review next, is one such algorithm.

Recall that the evaluation of the ML decoding rule (8.16), which reduces to (8.18) in the bi-AWGN case, has in general a prohibitive complexity, already for small values of the information-payload size k . The idea behind OSD is to replace (8.18) with

$$\hat{j} = \arg \min_{m \in \mathcal{L}} \|\mathbf{y} - \sqrt{\rho} \mathbf{x}(m)\|^2, \quad (8.31)$$

where the optimization is performed over a list $\mathcal{L}$ of much smaller cardinality than 2^k . OSD constructs such a list through the following steps.

Let us consider the problem of decoding a (n, k) linear block code $\mathcal{C}$ that is used over the bi-AWGN channel (8.1). Let $\mathbf{G}$ be the generator matrix of the code, $\mathbf{u}$ the k -dimensional information vector, $\mathbf{c}$ the corresponding codeword, and $\mathbf{x}$ its vector representation in the Euclidean space after BPSK modulation.

For a given received vector $\mathbf{y}$, we let $\mathbf{r} = [|y_1|, \dots, |y_n|]$; furthermore, we construct the additional vector $\mathbf{r}'$ by ordering the entries of $\mathbf{r}$ in decreasing order. Note that the scalars $\{y_\ell\}_{\ell=1}^n$ are proportional

to the channel LLRs. Hence, the vector $\mathbf{r}'$ contains a scaled version of the channel LLRs ordered according to their reliability. Let π_1 be the permutation that maps $\mathbf{r}$ into $\mathbf{r}'$. If we apply this permutation to the columns of the generator matrix $\mathbf{G}$, we obtain a new generator matrix:

$$\mathbf{G}' = \pi_1(\mathbf{G}). \quad (8.32)$$

It turns out to be convenient to associate to each column of $\mathbf{G}'$ a reliability value, which is given by the corresponding entry of $\mathbf{r}'$. Next we rearrange the columns of $\mathbf{G}'$ so that the first k columns of this matrix are the k linear independent columns with the highest reliability, ordered in decreasing order of reliability, and the remaining $n - k$ columns are also ordered in decreasing order of reliability. We denote the resulting matrix by $\mathbf{G}''$ and the corresponding column permutation by π_2 . Finally, we put $\mathbf{G}''$ in the systematic form $\mathbf{G}''' = [\mathbf{I}_k \mathbf{P}]$ by performing standard row operations. Here, $\mathbf{P}$ is of size $k \times n - k$. Note that the codes generated by $\mathbf{G}$ and by $\mathbf{G}'''$ are equivalent.

To construct the list, we now apply the second permutation π_2 to $\mathbf{r}'$ and obtain $\mathbf{r}'' = \pi_2(\mathbf{r}')$. Next, we perform a hard decision on the first k entries of $\mathbf{r}''$, to obtain the k -dimensional vector $\hat{\mathbf{u}}$. Specifically, we set $\hat{u}_i = 0$ if $z_i > 0$ and $\hat{u}_i = 1$ otherwise. It is worth remarking that the permutations ensure that the hard decision is performed on the most reliable linear-independent set of channel outputs.

For a given integer t , the list $\mathcal{L}$ is finally constructed by considering all codewords,

$$\hat{\mathbf{c}} = \pi_1^{-1}(\pi_2^{-1}(\tilde{\mathbf{u}}\mathbf{G}''')) \quad (8.33)$$

where $\tilde{\mathbf{u}}$ spans all k -dimensional vectors whose Hamming distance from $\hat{\mathbf{u}}$ is smaller or equal to t . The final decision is taken by computing the Euclidean distance between the BPSK vectors corresponding to each codeword in $\mathcal{L}$ and $\mathbf{y}$, and by selecting the codeword closest to $\mathbf{y}$.

The complexity of this procedure grows with the size of the list

$$|\mathcal{L}| = \sum_{m=0}^t \binom{k}{m}. \quad (8.34)$$

Clearly, the larger t , the larger the list and the better the performance of OSD; but also the greater the decoding complexity. Indeed, in the extreme case $t = k$, OSD coincides with ML decoding. In general, the value of t needed to approach ML decoding performance grows with the blocklength n . The decoding complexity can be reduced if the minimum distance $d_{\min}$ of the code is known. In such a case, one can stop the list construction procedure as soon as one finds a codeword whose corresponding BPSK vector has a Euclidean distance from $\mathbf{y}$ less than $\sqrt{\rho d_{\min}}$, since there cannot be a codeword that is closer to $\mathbf{y}$ than this.

8.2.5.2 Linear Block Codes With Tail-Biting Trellises

Short linear block codes can sometimes be represented efficiently by finite-length trellises with multiple initial and final states, in which each codeword corresponds to a tail-biting (TB) path with the same initial and final state. One example of codes that have this property and have also good distance spectra are linear block codes obtained through a tail-biting termination of suitably chosen convolutional codes.

The Viterbi algorithm can be used to decode such codes. However, since the initial state is not known, ML decoding requires running the Viterbi algorithm as many times as the number of initial states.

A suboptimal but lower-complexity decoding method is to use the so-called wrap-around Viterbi algorithm (WAVA). One assumes that all initial states are equiprobable and then runs the Viterbi algorithm one time. If the most likely path returned by the algorithm is a tail-biting path, decoding stops and this path is given as output. If the returned path is not tail-biting, the final state of the trellis is copied to the initial state and the Viterbi algorithm is run again.

The process is repeated until a tail-biting path is returned, or a maximum number of iterations is exceeded, in which case the decoder declares an error.

8.2.5.3 Nonbinary LDPC Codes

The performance under iterative decoding of LDPC codes in the short-blocklength regime can be significantly improved by constructing such codes on higher-order fields than $\mathbb{F}_2$ [12]. Low-complexity implementations of the required decoding algorithm are, however, still an active area of research.

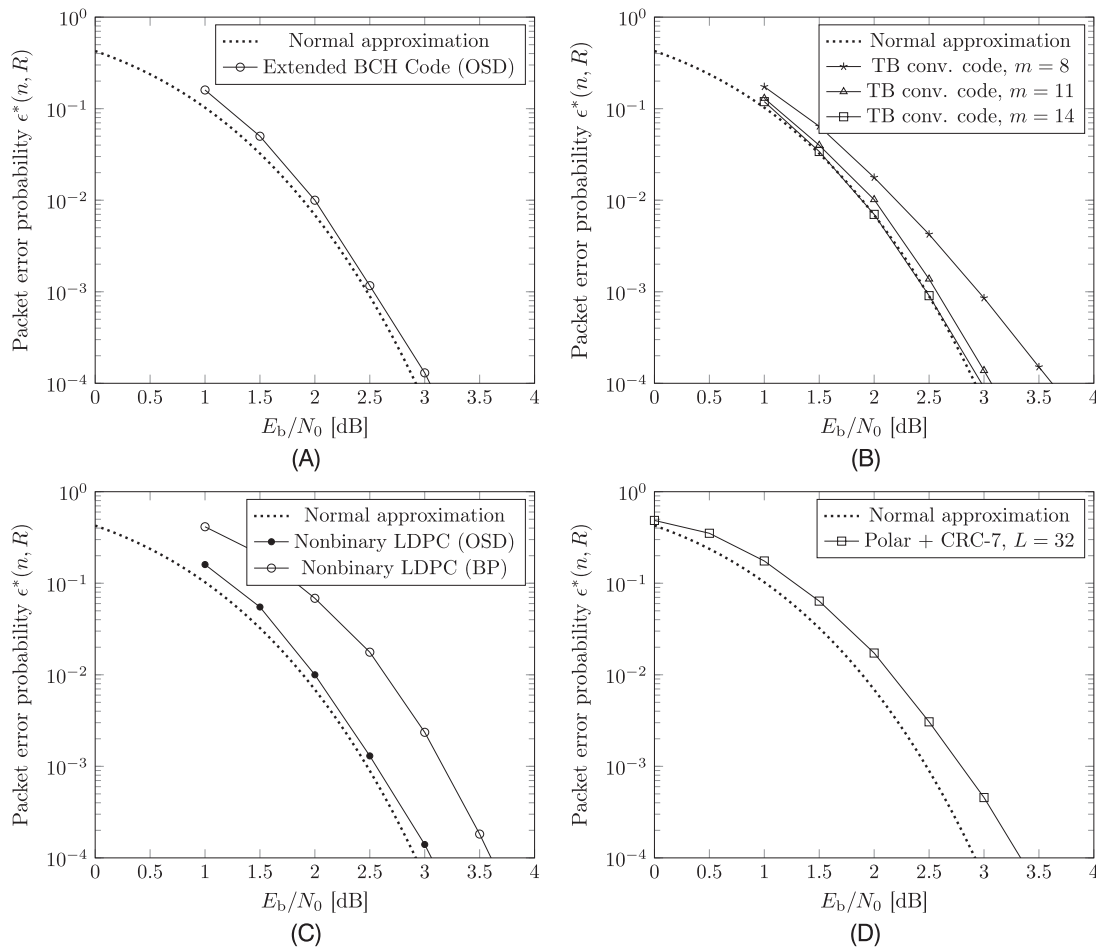
8.2.5.4 Performance

We illustrate next the performance of the coding schemes described in Sections 8.2.5.1–8.2.5.3. We consider a short code of length 128 and rate $R = 1/2$. Hence, $k = 64$.

Fig. 8.8A, illustrates the performance of a (128, 64) extended BCH code with minimum distance 22 [17]. The t parameter in the OSD decoder is set to 4, which results in a list of size 679121. In Fig. 8.8B, we consider three tail-biting convolutional codes with different memory m . Specifically, we analyze a memory 8 convolutional code with generator polynomial (given in octal form) [515, 677], a memory 11 convolutional code with generator polynomial [5537, 6131] and a memory 14 convolutional code with generator polynomial [75063, 56711]. We see from the figure that both the eBCH and the tail-biting convolutional codes with memory 11 and 14 operate remarkably close to the finite-blocklength normal approximation limit. The memory-8 convolutional code exhibits a loss of about 0.5 dB at 10^{-4} packet error probability. In Fig. 8.8C, we consider the performance of a nonbinary LDPC code constructed over a finite field of order 256. The PCM of the code has a constant row weight of 4 and a constant column weight of 2. We consider both iterative decoding with 200 iterations, and OSD with $t = 4$. One sees that the performances with OSD are close to the normal approximation, whereas the performance gap with iterative decoding is about 0.6 dB at 10^{-4} . Finally, in Fig. 8.8D we present the performance of a (128, 71) polar code, combined with a (71, 64) shorted cyclic code that serves as CRC. The list size is limited to 32. The performance gap to the normal approximation is about 0.4 dB at 10^{-4} .

8.3 CODING SCHEMES FOR FADING CHANNELS

So far, we have focused on the problem of transmitting information over the bi-AWGN channel (8.1). In this final section, we shall instead consider the more practically relevant scenario of communications over a multiantenna fading channel. The purpose is to illustrate the additional design challenges brought about by fading. Our focus will be on the short-packet regime and on the URLLC use case. We shall first discuss the single-input single-output (SISO) case and then move to multiple-input multiple-output (MIMO) transmissions.

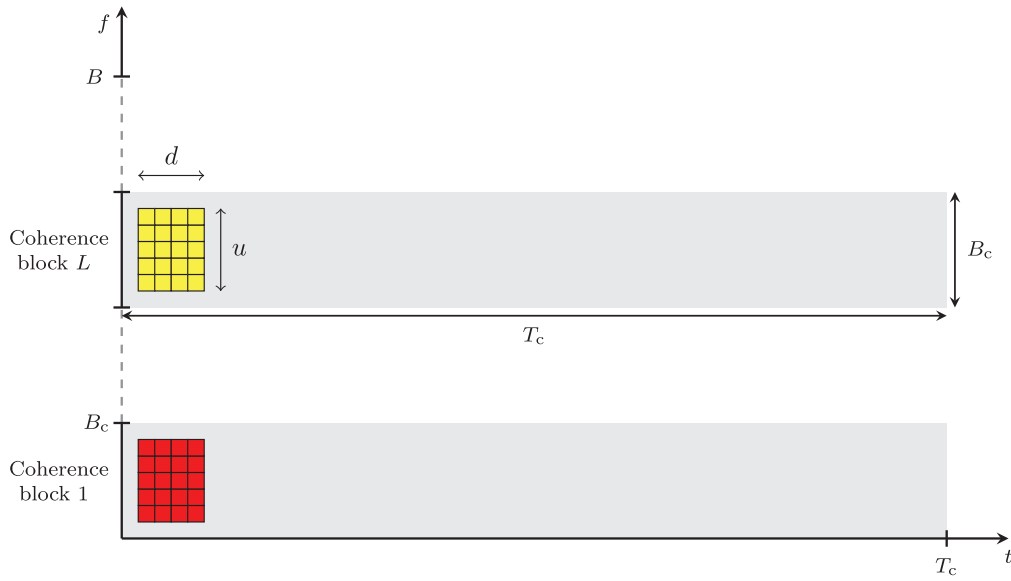
**FIGURE 8.8**

Performance of the coding schemes described in this section. Here, $R = 1/2$ and $n = 128$.

8.3.1 THE SISO CASE

We assume orthogonal frequency-division multiplexing (OFDM) operations and assume that each codeword spans multiple resource blocks (RBs), which are transmitted in the same time slot but at different frequencies. We assume that each RB contains d OFDM symbols consisting of u subcarriers. Hence, a RB conveys $n_c = d \times u$ complex-valued symbols. Low latency is achieved by selecting a sufficiently small value for d . For example, in downlink control channels, d may be chosen from the set $\{1, 2, 3\}$.

We assume that the coherence time T_c of the channel is larger than the transmission duration, and we let $L_{\max} = \lfloor B/B_c \rfloor$ be the ratio between the transmission bandwidth B and the coherence bandwidth of the channel B_c . Hence, $L_{\max}$ corresponds to the maximum number of frequency diversity branches offered by the channel. As we shall see, exploiting diversity is fundamental to achieve the reliability constraints set in URLLC.

**FIGURE 8.9**

Signaling strategy in the time-frequency plane.

For simplicity, we model channel variations in frequency using the block-memoryless fading assumption. According to this assumption, the channel stays constant over each coherence interval and changes independently across diversity branches. This means that the fading channel is fully characterized by $L_{\max}$ complex channel coefficients.

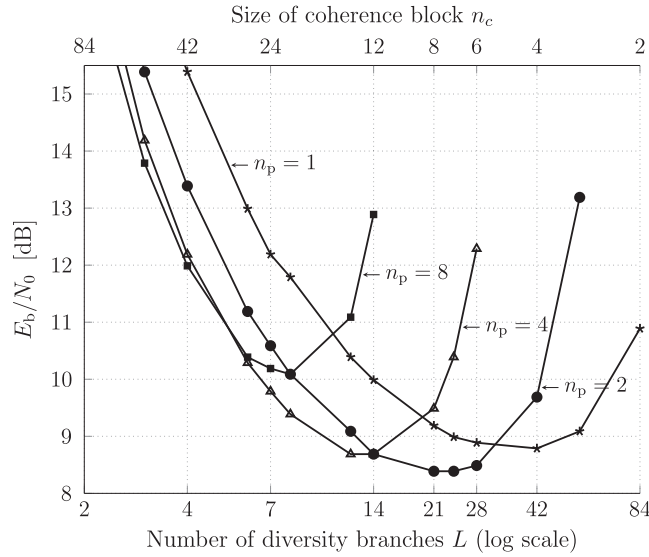
As illustrated in Fig. 8.9, we assume that each RB fits within a coherence interval, i.e., all the complex symbols contained in the RB experience the same fading gain. We also assume that different RBs are allocated at different frequency branches. Finally, we assume that each codeword consists of $L \leq L_{\max}$ RBs.

An example is in order. Assuming $B = 20$ MHz, as in LTE, we obtain $L_{\max} = 4$ for the extended pedestrian type-A (EPA) 5 Hz channel model [2], whose coherence bandwidth is 4.4 MHz, whereas $L_{\max} = 30$ for the tapped-delay-line type-C (TDL-C) 300 ms–3 km/h channel model [3], whose bandwidth is 0.66 MHz. In both cases, the coherence interval is about 85 ms, which exceeds by far the duration of a RB in all practically relevant scenarios. Indeed, recall that, with 15 kHz subspacing, the duration of an OFDM symbol is just 66.7 μ s. We shall denote by $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L] \in \mathbb{C}^{Ln_c}$ the vector of the transmitted complex symbols, where $n = L_c$ is the blocklength. The corresponding received vector is $[\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_L]$ where $\mathbf{y}_\ell$ is the received signal corresponding to the ℓ th RB; we have

$$\mathbf{y}_\ell = \sqrt{\rho} h_\ell \mathbf{x}_\ell + \mathbf{w}_\ell, \quad \ell = 1, \dots, L. \quad (8.35)$$

Here, $\{h_\ell\}$ are i.i.d. fading coefficients, in agreement with the memoryless block-fading assumption. We shall assume that the variance of the $\{h_\ell\}$ is normalized to one. Furthermore, $\mathbf{w}_\ell$ is the additive white complex-Gaussian noise whose entries are i.i.d. and have zero mean and unit variance.

We shall focus on pilot-assisted transmission [38] according to which n_p out of the n_c entries of a RB are allocated to pilot symbols known to the receiver, and the remaining entries are reserved for

**FIGURE 8.10**

Minimum energy per bit E_b/N_0 to achieve $\epsilon = 10^{-3}$ for the Rayleigh-fading case, as a function of the number of diversity branches L . Here, $k = 81$ and $n = 168$.

coded data symbols. Specifically, we assume that $\mathbf{x}_\ell = [\mathbf{x}_\ell^{(p)} \mathbf{x}_\ell^{(d)}]$ where $\mathbf{x}_\ell^{(p)}$ is the n_p -dimensional pilot-symbol vector and $\mathbf{x}_\ell^{(d)}$ is the $n_c - n_p$ data vector. Throughout, we will focus on the scenario in which both pilot and data symbols are transmitted at the same power. More precisely, we assume that the entries of $\mathbf{x}_\ell^{(p)}$ and $\mathbf{x}_\ell^{(d)}$ are QPSK symbols with unit energy. QPSK modulation is indeed suitable for the low-rate low-power scenarios that are relevant for URLLC.

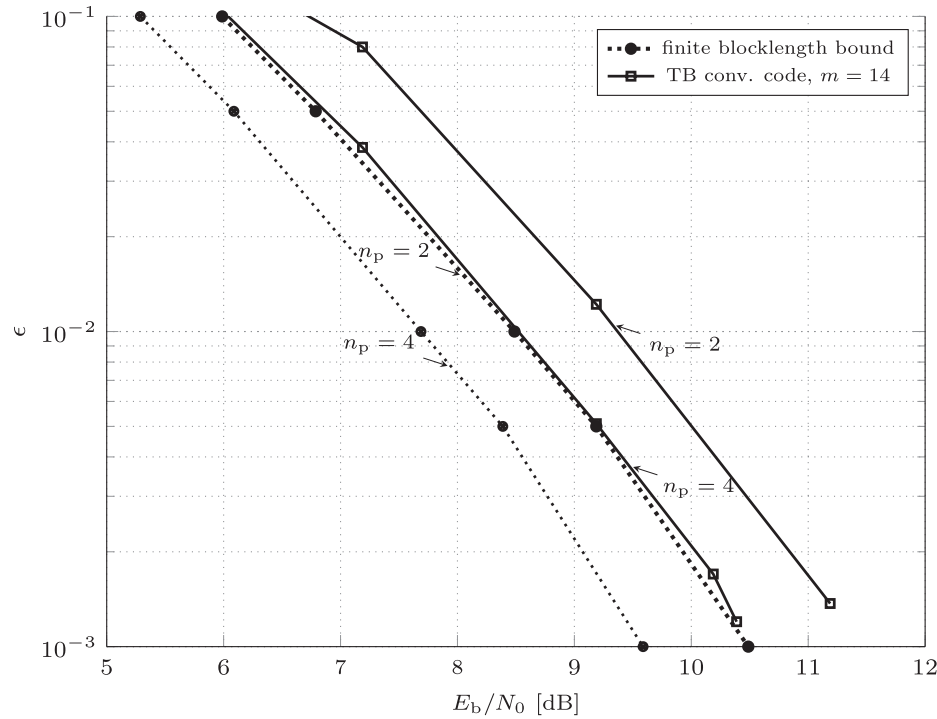
At the receiver side, we assume a practically relevant mismatch-decoding structure, in which the n_p -dimensional received vector $\mathbf{y}_\ell^{(p)}$ that corresponds to the pilot symbols is used to obtain a ML estimated $\hat{h}_\ell$ of the fading channel h_ℓ according to

$$\hat{h}_\ell = \mathbf{y}_\ell^{(p)} \frac{(\mathbf{x}_\ell^{(p)})^H}{n_p \sqrt{\rho}}. \quad (8.36)$$

Then the channel estimate is fed to a scaled minimum-distance decoder that treats it as perfect and produces a message estimate as follows:

$$\hat{j} = \arg \min_{m \in \{1, 2, \dots, 2^k\}} \sum_{\ell=1}^L \|\mathbf{y}_\ell^{(d)} - \hat{h}_\ell \mathbf{x}_\ell^{(d)}(m)\|^2. \quad (8.37)$$

The performance of this transceiver architecture has been recently studied in [27, 15] using finite-blocklength information-theoretic methods similar to the ones described in Section 8.1. This kind of theoretical analyses provides useful insights on the optimal number of diversity branches L one should code over for a given fixed blocklength n , and on the optimal number of pilot symbols n_p that should be allocated in each resource block. This is illustrated in Fig. 8.10 where we have depicted

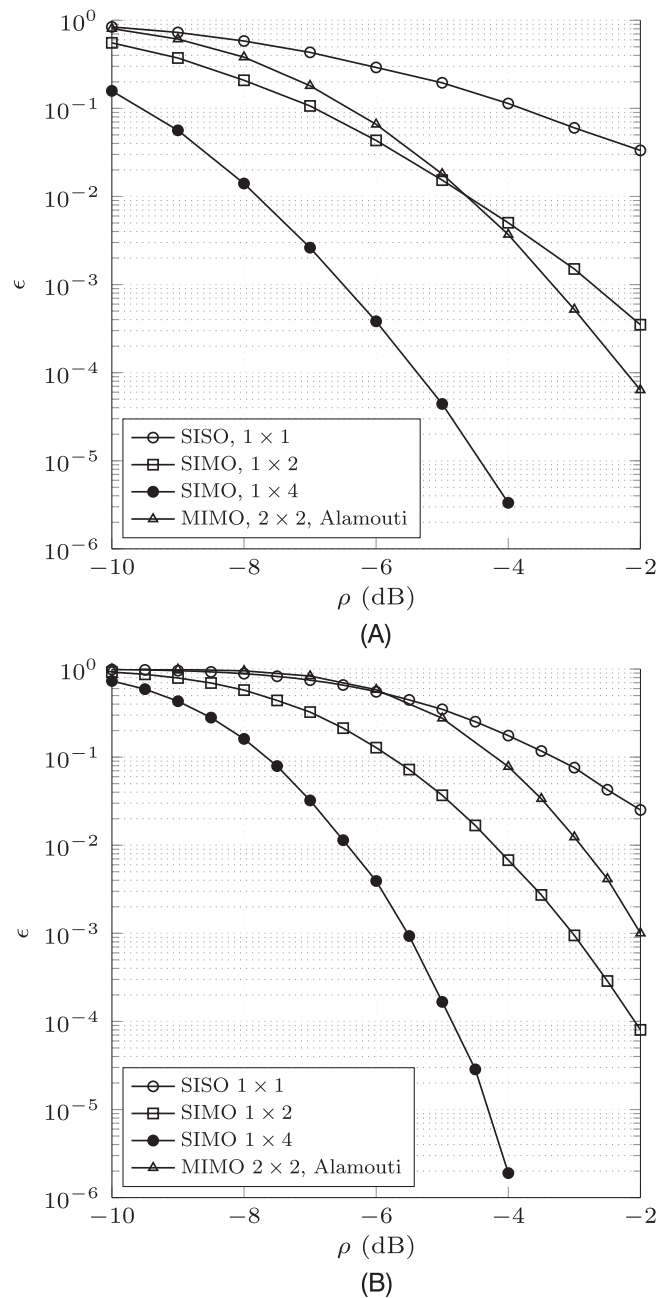
**FIGURE 8.11**

Packet error probability versus energy per bit for the case of Rayleigh fading, $k = 81$, $n = 186$, $L = 7$. Information-theoretic bounds [27, Th. 3] and performance of an actual coding scheme based on tail-biting convolutional codes and OSD.

the information-theoretic upper bound [27, Th. 3] on the minimum energy per bit E_b/N_0 required to achieve $\epsilon = 10^{-3}$ when transmitting $k = 81$ information bits. The blocklength $n = Ln_c$ is 168 and a different number of diversity branches L is considered (the larger L , the smaller n_c).

We see that there exists an optimum number of diversity branches $L = 21$ that minimizes the energy per bit. When $L < 21$, the system is penalized by the insufficient amount of frequency diversity available, which makes reliable transmission costly from an energy viewpoint. When $L > 21$, and, hence $n_c = n/\ell < 8$, the system suffers from the large pilot-symbol overhead, which is required to track the fast channel variations. We also observe that the number n_p of pilot symbols per resource block needs to be chosen carefully as a function of the number L of frequency diversity branches. Indeed, setting n_p to a suboptimal value yields a significant performance loss.

We next consider the performance of an actual coding scheme and benchmark it against the information-theoretic bounds. Specifically, we choose a $(324, 81)$ binary quasi-cyclic code obtained by tail-biting termination of a rate $1/4$ convolutional code with memory $m = 14$. The output of the encoder is passed through a pseudorandom interleaver and then some of the coded symbols are punctured to accommodate the desired number of pilot symbols per resource block (RB) after QPSK modulation. Decoding is performed via OSD with $t = 3$. The performance of this coding scheme when $L = 7$ is illustrated in Fig. 8.11 for the case $n_p = 2$ and $n_p = 4$. In both cases, the gap to the information-theoretic bound is about 1 dB.

**FIGURE 8.12**

Packet error probability versus E_b/N_0 ; $k = 30$, $n = 288$, spatially white Rayleigh fading. (A) $L = 4$. (B) $L = 12$.

8.3.2 THE MIMO CASE

Exploiting the additional spatial diversity provided by MIMO transmission and reception is crucial to achieve the reliability level targeted by URLLC. The information-theoretic bounds depicted in Fig. 8.10

can be extended to MIMO communications, which allows one to explore the benefit of multiple antennas. As discussed in [15], the information-theoretic bounds can be extended to cover the case in which a space-frequency code is used at the transmitter to let the available antennas provide spatial diversity, in the absence of channel-state information at the transmitter.

In Fig. 8.12, we report the performance of different MIMO configurations for the case $k = 30$, which is relevant for downlink control-information transmission, and $n = 288$. We compare the performance of single-input single-output (SISO), 1×2 and 1×4 SIMO, and 2×2 MIMO with Alamouti encoding [4]. In Fig. 8.12A, we consider the case $n_c = 72$ and $L = 4$ (EPA 5 Hz channel model), whereas in Fig. 8.12B we set $n_c = 24$ and $L = 12$ (TDL-C channel model). In both cases, the number of pilot symbols is optimized to minimize the error probability. Furthermore, QPSK transmission is assumed.

We observe that, within the range of SNR values considered in the figure, only the 1×4 SIMO configuration is able to achieve an error probability below 10^{-5} , a common requirement in URLLC. Although the 2×2 MIMO Alamouti configuration offers the same amount of spatial diversity as 1×4 single-input multiple-output (SIMO), it is more sensitive to channel-estimation errors. This is particularly evident in Fig. 8.12B, where the small value of n_c results in a noisy channel estimate. As a consequence, the 2×2 MIMO Alamouti configuration performs worse than 1×2 SIMO over the range of SNR values considered in the figure.

REFERENCES

- [1] 3GPP, TS 38.212 V15.0.0: multiplexing, and channel coding, 2017, Dec.
- [2] 3GPP, TS 36.104: Technical specification group radio access network, 3GPP, 2012.
- [3] 3GPP, TR 38.901: Study on channel model for frequencies from 0.5 to 100 GHz, 3GPP, 2017.
- [4] S. Alamouti, A simple transmit diversity technique for wireless communications, *IEEE Journal on Selected Areas in Communications* (ISSN 0733-8716) 16 (8) (1998, Oct.) 1451–1458, <https://doi.org/10.1109/49.730453>.
- [5] N. Alon, M. Luby, A linear time erasure-resilient code with nearly optimal recovery, *IEEE Transactions on Information Theory* 42 (11) (1996, Nov.) 1732–1736.
- [6] E. Arıkan, Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, *IEEE Transactions on Information Theory* 55 (7) (2009, Jul.) 3051–3073.
- [7] E.E. Arıkan, I. Telatar, On the rate of channel polarization, in: *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, 2009, Jul., pp. 1493–1495, <https://arxiv.org/abs/0807.3806>.
- [8] V. Bioglio, F. Gabry, I. Land, Low-complexity puncturing and shortening of polar codes, in: *Proc. IEEE Wireless Commun. Netw. Conf.*, San Francisco, CA, U.S.A., 2017, Mar.
- [9] G. Böcherer, F. Steiner, P. Schulte, Bandwidth efficient and rate-matched low-density parity-check coded modulation, *IEEE Transactions on Communications* 63 (12) (2015, Dec.) 4651–4665.
- [10] A. Collins, G. Durisi, T. Erseghe, V. Kostina, J. Östman, Y. Polyanskiy, I. Tal, W. Yang, SPECTRE: short-packet communication toolbox, v2.0, <https://github.com/yp-mit/spectre>, 2016, Sep.
- [11] D.J. Costello Jr., G.D. Forney Jr., Channel coding: the road to channel capacity, *Proceedings of the IEEE* 95 (6) (2007, Jun.) 1150–1177.
- [12] M.C. Davey, D. MacKay, Low density parity-check codes over GF(q), *IEEE Communications Letters* 2 (6) (1998, Jun.) 165–167.
- [13] D. Divsalar, S. Dolinar, C.R. Jones, K. Andrews, Capacity approaching protograph codes, *IEEE Journal on Selected Areas in Communications* 27 (6) (2009, Aug.) 876–888.
- [14] T. Erseghe, Coding in the finite-blocklength regime: bounds based on Laplace integrals and their asymptotic approximations, *IEEE Transactions on Information Theory* 62 (12) (2016, Dec.) 6854–6883.
- [15] G.C. Ferrante, J. Östman, G. Durisi, K. Kittichokechai, Pilot-assisted short-packet transmission over multiantenna fading channels: a 5G case study, in: *Conf. Inf. Sci. Sys. (CISS)*, Princeton, NJ, 2018, Mar.

- [16] G.D. Forney Jr., G. Ungerboeck, Modulation and coding for the linear Gaussian channels, *IEEE Transactions on Information Theory* 44 (6) (1998, Oct.) 2384–2415.
- [17] M. Fossorier, S. Lin, Soft-decision decoding of linear block codes based on ordered statistics, *IEEE Transactions on Information Theory* (ISSN 0018-9448) 41 (5) (1995, Sep.) 1379–1396, <https://doi.org/10.1109/18.412683>.
- [18] R. Gallager, Low-density parity-check codes, *IRE Transactions on Information Theory* 8 (1) (1962, Jan.) 21–28.
- [19] G. He, J.C. Belfiore, X. Liu, Y. Ge, R. Zhang, I. Land, Y. Chen, R. Li, J. Wang, G. Yang, W. Tong, β -expansion: a theoretical framework for fast and recursive construction of polar codes, in: *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Singapore, 2017, Dec., <https://arxiv.org/abs/1704.05709>.
- [20] G. Liva, F. Steiner, pretty-good-codes.org: online library of good channel codes, <http://pretty-good-codes.org/>, 2017.
- [21] G. Liva, W.E. Ryan, M. Chiani, Quasi-cyclic generalized LDPC codes with low error floors, *IEEE Transactions on Communications* 56 (1) (2008, Jan.) 49–57.
- [22] D. MacKay, R. Neal, Good codes based on very sparse matrices, in: C. Boyd (Ed.), *IMA Conf. Cryptography and Coding*, Springer-Verlag, 1995, Oct.
- [23] A. Martinez, A. Guillén i Fàbregas, Saddlepoint approximation of random-coding bounds, in: *Proc. Inf. Theory Applicat. Workshop (ITA)*, San Diego, CA, U.S.A., 2011, Feb.
- [24] M. Mondelli, S.H. Hassani, R. Urbanke, Unified scaling of polar codes: error exponent, scaling exponent, moderate deviations, and error floors, *IEEE Transactions on Information Theory* 62 (12) (2016, Dec.) 6698–6712.
- [25] M. Mondelli, S.H. Hassani, R. Urbanke, Construction of polar codes with sublinear complexity, <https://arxiv.org/abs/1612.05295>, 2017, Jul.
- [26] S.M. Moser, *Information Theory (Lecture Notes)*, fifth ed., ETH Zurich/National Chiao Tung University, Switzerland/Taiwan, 2017, Mar.
- [27] J. Östman, G. Durisi, E.G. Ström, M.C. Coşkun, G. Liva, Short packets over block-memoryless fading channels: pilot-assisted or noncoherent transmission?, <https://arxiv.org/abs/1712.06387>, 2017, Dec.
- [28] S. Parkvall, E. Dahlman, A. Furuskär, M. Frenne, NR: the new 5G radio access technology, *IEEE Communications Standards Magazine* (2017, Dec.) 24–30.
- [29] Y. Polyanskiy, H.V. Poor, S. Verdú, Channel coding rate in the finite blocklength regime, *IEEE Transactions on Information Theory* 56 (5) (2010, May) 2307–2359.
- [30] T.J. Richardson, R. Urbanke, *Modern Coding Theory*, Cambridge Univ. Press, Cambridge, U.K., 2008.
- [31] W.E. Ryan, S. Lin, *Channel Codes: Classical and Modern*, Cambridge Univ. Press, 2009.
- [32] S. Sandberg, M. Andersson, A. Shirazinia, Y. Blankenship, LDPC Codes for 5G New Radio, 2018.
- [33] I. Sason, S. Shamai (Shitz), Performance analysis of linear codes under maximum-likelihood decoding: a tutorial, *Foundations and Trends in Communications and Information Theory* (2006), <https://doi.org/10.1561/01000000009>.
- [34] C. Schürch, A partial order for the synthesized channels of a polar code, in: *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, 2016, Jul., pp. 220–224.
- [35] C.E. Shannon, A mathematical theory of communication, *The Bell System Technical Journal* 27 (1948, July and October) 379–423, 623–656.
- [36] I. Tal, A. Vardy, List decoding of polar codes, *IEEE Transactions on Information Theory* (ISSN 0018-9448) 61 (5) (2015, May) 2213–2226, <https://doi.org/10.1109/TIT.2015.2410251>.
- [37] R.M. Tanner, A recursive approach to low complexity codes, *IEEE Transactions on Information Theory* 27 (9) (1981, Sep.) 533–547.
- [38] L. Tong, B.M. Sadler, M. Dong, Pilot-assisted wireless transmissions, *IEEE Signal Processing Magazine* 21 (6) (2004, Nov.) 12–25, <https://doi.org/10.1109/MSP.2004.1359139>.
- [39] G. Vazquez-Vilar, A.T. Campo, A.G. i Fàbregas, A. Martinez, Bayesian M-ary hypothesis testing: the meta-converse and Verdú-Han bounds are tight, *IEEE Transactions on Information Theory* 62 (5) (2016, May) 2324–2333.
- [40] H. Zhang, R. Li, J. Wang, S. Dai, G. Zhang, Y. Chen, H. Luo, J. Wang, Parity-check polar coding for 5G and beyond, in: *IEEE Int. Conf. Commun. (ICC)*, 2018, 01, <https://arxiv.org/abs/1801.03616>.

SIMULATOR

9

For further research and development of the NR physical layer, we complement this book with a MATLAB-based link level simulator. We shall present an open-source simulator, including several waveforms, i.e., cyclic prefix based orthogonal frequency division multiplexing (CP-OFDM), windowed-OFDM (W-OFDM), universal filtered OFDM (UF-OFDM), discrete Fourier transform spread OFDM (DFTS-OFDM), offset QAM filter-bank multicarrier (FBMC-OQAM), and QAM filter-bank multicarrier (FBMC-QAM), various hardware impairments, i.e., power amplifier (PA) nonlinearity, oscillator phase noise (PN), and carrier frequency offset (CFO), and a geometry-based stochastic channel model supporting millimeter-wave channel emulation up to 80 GHz.

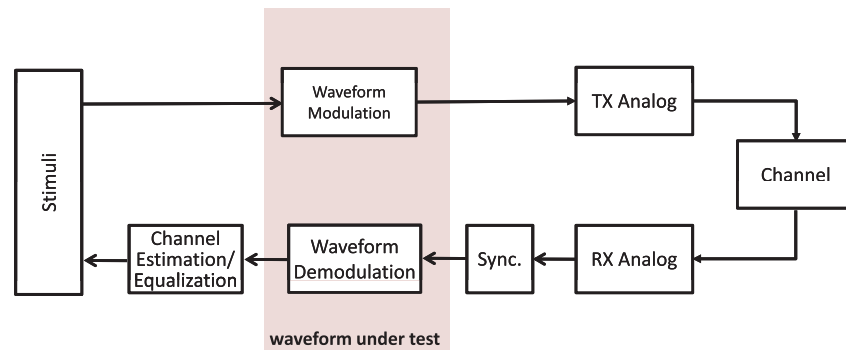
The simulator is developed by Qamcom Research & Technology AB. We release it with this book as an open-source simulator (with permission from Qamcom Research & Technology AB). The simulator can be accessed at www.qamcom.se/research/5gsim.

The simulator was originally developed in the EU funded mmMAGIC project [15] to assess performance of various waveforms in a common environment. As discussed in Chapter 5, several waveforms were proposed for 5G NR. Due to lack of a common waveform simulator, sometimes contradictory results were observed. Therefore, the mmMAGIC project realized importance of a transparent simulator for fair comparisons. Furthermore, the simulator was used to develop receiver algorithms for NR. At present (August 2018), the simulator has the following key components:

- A channel model up to 80 GHz (discussed in Chapter 3);
- A phase-noise model for a free-running oscillator and a phase-noise model for a phased-lock loop based oscillator (discussed in Chapter 4);
- A power-amplifier model based on a generalized memory polynomial (discussed in Chapter 4);
- Modulation/demodulation modules for various waveforms (discussed in Chapter 5);
- Algorithms for channel estimation & equalization, synchronization, and phase-noise compensation (discussed in Chapter 6).

It is likely that the simulator will evolve in the future with contributions from the authors and the open-source community.

This chapter is organized as follows. Section 9.1 provides an overview of the simulator, followed by details of different functional modules and waveforms in Section 9.2 and Section 9.3, respectively. Our ambition is not to discuss fundamentals (which are already covered in previous chapters); rather the focus is on implementation specific details. Finally, Section 9.4 shows a number of simulation exercises with various waveforms subject to the above-mentioned impairments and receiver algorithms. With these exercises, important observations are made.

**FIGURE 9.1**

Block diagram of the simulator.

9.1 SIMULATOR OVERVIEW

A functional diagram of the simulator is shown in Fig. 9.1. The stimuli module sends (receives) QAM symbols to be modulated (demodulated) by the waveform modulator (demodulator). The modulated signal goes through analog modules at the transmitter (Tx) and receiver (Rx), and the radio channel. The Tx analog module adds hardware impairments, such as power amplifier (PA) nonlinearity, oscillator phase noise (PN), and carrier frequency offset (CFO) to the transmitted signal. The transmitted signal then experiences delay spread, Doppler spread, and interference, incurred by the channel model. After the radio channel, the received signal is further impaired by additive white Gaussian noise (AWGN) and PN that occur in the Rx analog module (before demodulation). At the receiver, timing and frequency synchronization are performed first. The synchronized signal is transformed to the frequency (subcarrier) domain by the waveform demodulator. Channel estimation and equalization are performed afterwards. (Note that it is possible to estimate the channel in the time domain before waveform demodulation. Nevertheless, the conventional frequency-domain channel estimation is chosen in this simulator.) At the moment, included waveforms are cyclic prefix (CP) based orthogonal frequency division multiplexing (OFDM) [5], windowed-OFDM (W-OFDM) [20], universal filtered OFDM (UF-OFDM) [22], discrete Fourier transform spread OFDM (DFT-S-OFDM) [4], offset QAM filter-bank multicarrier (FBMC-OQAM) [21], and QAM filter-bank multicarrier (FBMC-QAM) [17]. Other waveforms can be integrated into the open-source simulator as well.

We describe each module of the simulator separately in the following sections. Exemplary simulations are conducted to compare different waveforms and to illustrate the functionality of the simulator.

9.2 FUNCTIONAL MODULES

We present all the functional modules (except for waveforms) of the simulator separately in this section. The presentation of different waveforms is deferred to the next section.

9.2.1 CHANNEL MODEL

Both geometry-based channel models and 5G stochastic channel models [11–16] have been discussed in Chapter 2. Among the 5G stochastic channel model, the mmMAGIC (also known as QuaDRiGa) channel model [16] is the only one that has been made open-source. Therefore, we integrate the mmMAGIC channel model into the simulator. For the sake of completeness, we briefly present the mmMAGIC channel model here. The mmMAGIC channel model is an extension of the well-known WINNER channel model [1]. While the WINNER model is valid up to 6 GHz, the current version of the mmMAGIC channel model supports millimeter-wave channel up to 80 GHz. In addition, the mmMAGIC channel model can well predict the time variance of the channel (as compared to the WINNER model). The data flow of the mmMAGIC channel model is given as follows:

- define network layout, terminal trajectories, propagation scenarios along the trajectories, and antenna patterns of the terminals and base stations;
- calculate the maps of large-scale parameters (i.e., RMS delay spread, Rician K -factor, shadow fading, azimuth spread of departure/arrival, elevation spread of departure/arrival);
- calculate the initial delays, the path powers, and the angles of departure/arrival;
- calculate the polarized channel coefficients including the transmission loss, the shadow fading, and the K -factor.

A detailed description of the channel model can be found in [14]. The channel model can be switched off, in which case the channel boils down to a simple AWGN channel. This can be useful if one is only interested in making manifest the effect of a certain hardware impairment.

9.2.2 POWER AMPLIFIER MODEL

As discussed in Chapter 4, the PA effects can be accurately modeled by the Volterra series. Nevertheless, the Volterra series model can be quite complicated. By ignoring all the memory and crossterms, the Volterra series model boils down to the polynomial model. The polynomial model is very simple and allows analytical study. However, it cannot capture the dynamic or memory effect of the PA. As a trade-off, the memory polynomial and the generalized memory polynomial are introduced in modeling the PA effects. The former omits all the crossterms and, therefore, captures only the first degree of dynamic effects, whereas the latter only ignores part of the crossterms and is very close to the Volterra series model. The generalized memory polynomial (GMP) model is a good compromise between accuracy and complexity. As the nonlinear order increases, the variances of the estimated parameters also increases. To overcome this problem, local basis functions are used. Specifically, a vector-switched GMP (VS-GMP)-based PA model [2] is included in the simulator. The interested reader may refer to [2] for a detailed description of the PA model.

Note that the PA nonlinearity causes spectral regrowth, i.e., power leakage to adjacent channels or increased out-of-band (OOB) emissions. In order to correctly examine the PA effects, interpolation and decimation ought to be performed before and after the PA model. The implemented PA model [2] requires an interpolation with an oversampling factor of 5 to 7. Otherwise, the OOB emissions caused by the PA will incur a (nonphysical) inband distortion due to the spectral periodicity of time-discrete signals. The Parks–McClellan FIR filter [13] is used as anti-aliasing filter in the interpolation and decimation. Even though the Parks–McClellan FIR filter is close to an ideal low-pass filter, there will inevitably be some distortion in the transition around the cut-off frequency. As a result, guard bands

(i.e., zero subcarriers in the beginning and end of each multicarrier symbol) [10] are needed. The Parks–McClellan FIR filter has another effect: delaying the signal by half of the filter length. Since it is used twice (i.e., before and after the PA model), the filter delay equals its filter length. The filter delay and the delay caused by the random multipath channel will be estimated in the synchronization module.

The PA model can be switched off, in which case the Parks–McClellan FIR filter will be switched off automatically.

9.2.3 PHASE-NOISE MODEL

The PN of an oscillator is a multiplicative noise. It causes a random phase rotation of the received signal. For a multicarrier waveform, the PN causes common phase error (CPE) and inter-carrier interference (ICI). While the former represents a common phase rotation to all the subcarrier, the latter refers to the interference for each subcarrier caused by all the other ones. The CPE can be readily corrected using scattered pilots, whereas the ICI may be difficult to eliminate. There are ICI correction algorithms, e.g., [19], [7], which can effectively mitigate the ICI effect. However, these algorithms are computationally heavy.

As discussed in Chapter 3, the PN of a free-running oscillator can be modeled by a random walk (or Wiener) process, whose discrete-time expression is given by $\varphi(n+1) = \varphi[n] + w[n]$, where φ is the PN and w is a zero-mean Gaussian random variable whose variance is $4\pi\beta/f_s$ with f_s denoting the sampling frequency and β representing the 3-dB bandwidth of the phase noise. The Wiener PN model is perhaps the most popular PN model in the literature due to its simplicity and well-known statistical properties. However, its variance increases linearly with time. Hence, in practice, a phase-locked loop (PLL) is usually added to regularize the PN. A PN model of the PLL-based oscillator has been proposed in mmMAGIC [12]. This model was developed specifically for millimeter-wave communications. It is a more realistic (yet complicated) model as compared with the simple Wiener PN model. The PLL PN model is briefly mentioned in Chapter 3.

Both PN models are included in the simulator. The parameter β is used to specify the quality of the free-running oscillator. The PLL PN model involves many parameters. For simplicity, “low” and “high” PN modes were implemented for good and bad PLL-based oscillators. The detailed modeling parameters of the PN of the PLL-based oscillator are listed in Table 9.1.

Either PN model can be chosen at the transmitter and the receiver. The PN model at either side can be switched on and off independently.

9.2.4 SYNCHRONIZATION

As mentioned before, the multipath channel and the filters will cause delay (timing offset) of the received signal. The timing offset has to be estimated before further processing of the received signal. The CFO represents the frequency difference between the carrier frequencies at the transmitter and at the receiver. If left uncompensated, it will cause a phase rotation of the received signal and the phase increases linearly with time. The synchronization module performs timing and frequency synchronization. By assuming common preamble, the timing offset, and CFO are estimated using the methods in [24] and [23], respectively. One can also choose perfect timing and frequency synchronization. In the former case, the true value of the timing offset is used for timing synchronization. In the latter case, the CFO is set to zero and the frequency synchronization module is disabled (since perfect frequency synchronization results in a CFO-free signal).

Table 9.1 Parameter of “low” and “high” phase-noise settings

Component	Parameter	Value (f_c is the carrier frequency in GHz)	
		Low	High
Reference	S_0	$-60 + 20\log_{10}(f_c)$ (dBc/Hz)	$-60 + 20\log_{10}(f_c)$ (dBc/Hz)
	f_p	1 Hz	1 Hz
	f_x	10 kHz	3980 Hz
PLL	S_0	$-140.35 + 20\log_{10}(f_c)$ (dBc/Hz)	$-122.35 + 20\log_{10}(f_c)$ (dBc/Hz)
	f_p	∞	∞
	f_p	∞	∞
	f_{fc} (flicker corner)	2 MHz	0.8 MHz
	k (flicker exponent)	1	1
	f_{fp} (flicker pole)	200 Hz	80 Hz
VCO	S_0	$-24.34 + 20\log_{10}(f_c)$ (dBc/Hz)	$-3.35 + 20\log_{10}(f_c)$ (dBc/Hz)
	f_p	1 Hz	1 Hz
	f_x	20 MHz	80 MHz
Transfer function from reference and loop noise to output	$\frac{K_D K_{VCO} Z(s)}{sN_D + K_D K_{VCO} Z(s)}$	$6.366 \times 10^{-7}s + 1$ (Numerator)	$1.592 \times 10^{-6}s + 1$ (Numerator)
		$7.076 \times 10^{-29}s^4 + 5.780 \times 10^{-21}s^3$ $+ 1.013 \times 10^{-13}s^2 + 6.366 \times 10^{-7}s + 1$ (Denominator)	$6.214 \times 10^{-20}s^3 + 6.333 \times 10^{-13}s^2$ $+ 1.592 \times 10^{-6}s + 1$ (Denominator)
Transfer function from reference and loop noise to output	$\frac{sN_D}{sN_D + K_D K_{VCO} Z(s)}$	$7.076 \times 10^{-29}s^4 + 5.780 \times 10^{-21}s^3$ $+ 1.013 \times 10^{-13}s^2$ (Numerator)	$6.214 \times 10^{-20}s^3 + 6.333 \times 10^{-13}s^2$ $+ 1.013 \times 10^{-13}s^2$ (Numerator)
		$7.076 \times 10^{-29}s^4 + 5.780 \times 10^{-21}s^3$ $+ 1.013 \times 10^{-13}s^2 + 6.366 \times 10^{-7}s + 1$ (Numerator)	$6.214 \times 10^{-20}s^3 + 6.333 \times 10^{-13}s^2$ $+ 1.592 \times 10^{-6}s + 1$ (Denominator)

9.2.5 CHANNEL ESTIMATION AND EQUALIZATION

The frequency-domain channel estimation (equalization) is used in the simulator. The exact channel estimation methods may differ slightly for different waveforms. Yet, in essence, the channel transfer functions (CTFs) are estimated at the pilot subcarriers using the least-square (LS) estimator and the CTFs at other subcarriers are obtained by filtering [6].

Like the synchronization, the simulator allows for perfect channel estimation, in which case the true value of the channel transfer function is used for channel equalization.

9.3 WAVEFORMS

Several waveforms have been presented in Chapter 5. In the following subsections, we will briefly introduce these waveforms from an implementation point of view in the simulator. For all waveforms, the received (equivalent baseband) signal subject to phase noise, CFO, and multipath channel with

additive noise is modeled as

$$y[n] = \exp(j(\phi[n] + 2\pi\epsilon n)) \sum_{l=0}^{L-1} h_l x[n-l-v] + w[n] \quad (9.1)$$

where ϵ denotes the CFO (normalized by the sampling frequency), ϕ denotes the PN at the receiver, x represents the transmitted time-domain signal, h_l is the l th tap of the channel impulse response (CIR), L denotes the channel length, and w denotes the AWGN. Note that, for notational simplicity and without loss of generality, the effects of PA and PN at the transmitter have been omitted in the expression.

9.3.1 CP-OFDM

The CP-based OFDM (CP-OFDM) enjoys the simplest (de)modulation implementation among all the multicarrier waveforms. Its modulation can be expressed as

$$\mathbf{x} = [\mathbf{F}_{\text{CP}} \mathbf{F}]^H \mathbf{s}, \quad (9.2)$$

where $\mathbf{s}$ denotes the $N \times 1$ column vector consisting of the subcarrier symbols, $\mathbf{F}$ denotes the $N \times N$ discrete Fourier transform (DFT) matrix whose element is given by $\exp(-j2\pi kn/N)/\sqrt{N}$ ($k, n = 0, \dots, N-1$), and $\mathbf{F}_{\text{CP}}$ consists of the last N_{CP} columns of $\mathbf{F}$ (where N_{CP} denoting the CP length). Note that $\mathbf{F}_{\text{CP}} \mathbf{s}$ forms the CP of the CP-OFDM symbol. The CP-OFDM demodulation matrix is given by $[\mathbf{0} \mathbf{F}]$, where $\mathbf{0}$ is an $N \times N_{\text{CP}}$ zero matrix for CP removal.

9.3.2 W-OFDM

One can improve the OOB emission of the CP-OFDM by applying a windowing function to the transmitted time-domain signal, i.e., W-OFDM. Because of the windowing, the W-OFDM is also called weighted CP-OFDM (WCP-OFDM) [20]. By tapering the edges of the time-domain signal, the OOB emission can be reduced significantly. The tapering at the left edge causes no distortion of the useful signal thanks to the CP. To correct for the distortion due to the tapering at the right edge, the left edge of the received signal is added to the right edge, as explained in Chapter 5. This operation is called weighted overlap and add (WOLA) [9].

The raised-cosine filter is chosen as the windowing function in the simulator. The WOLA operation is implemented in the timing synchronization module at the W-OFDM receiver.

9.3.3 UF-OFDM

The UF-OFDM is also known as the universal filtered multicarrier (UFMC) [22]. It is achieved by first grouping the active subcarriers into B blocks of consecutive subcarriers, and then performing inverse discrete Fourier transform (IDFT) and filtering to each of the blocks. For the sake of convenience in the implementation, each block contains N_0 subcarriers. The filter bank is obtained by exponential modulation of a prototype filter of length N_v . The Dolph–Chebyshev filter is used as the prototype

filter in the simulator. The UF-OFDM modulation can be expressed as

$$\mathbf{x} = \sum_{i=1}^B \mathbf{V}_i \mathbf{D}_i \mathbf{s}_i \quad (9.3)$$

where the subscript i is the block index, $\mathbf{s}_i = [s_i(0), \dots, s_i(N_0 - 1)]^T$ consists of symbols and scattered pilots to be modulated on the N_0 subcarriers in the i th block, $\mathbf{D}_i$ consists of N_0 columns corresponding the subcarriers in the i th block submatrix of the IDFT matrix $\mathbf{F}^H$, and $\mathbf{V}_i$ is a Toeplitz matrix whose first column vector is given by $[v_i(0), \dots, v_i(N_v), 0, \dots, 0]^T$ where $v_i[n] = v[n] \exp(-j2\pi n k_i / N_v)$ with k_i denoting the center frequency of the i th block and $v[n]$ denoting the n th tap of the prototype filter. The UF-OFDM demodulation can be achieved by performing $2N_f$ -point DFT of the zero-padded time-domain UF-OFDM symbol and then down sampling the frequency-domain signal by a factor of 2.

Note that a time-domain preprocessing has to be performed at the transmitter after the UF-OFDM modulation in order to remove the deterministic distortion introduced by the filter bank at the transmitter [22].

By performing filtering on groups of subcarriers, the OOB emission can be reduced significantly at the cost of increased complexity. There is also modest increase of symbol length due to the filters. Note that the performance of the UF-OFDM can be improved by introducing a CP in the presence of large delay spread [9] at the cost of increased overhead.

9.3.4 FBMC-OQAM

The FBMC-OQAM is also known as FBMC [3]. It applies a filter to each of the subcarriers. The filter length is K times of the number of subcarriers. K is the overlapping factor (which is a positive integer number). The filter bank can be implemented in the frequency domain using the frequency spreading method [3], which increases the DFT size from N to KN . Since the IDFT output and DFT input are overlapped in the time domain, a significant amount of redundancy is present in the computations of the frequency spreading method. In order to reduce this redundancy, the simulator implements the FBMC-OQAM in the time domain using the polyphase network (PPN) method [3].

Unlike the CP-OFDM, the neighboring subcarriers of the FBMC-OQAM are overlapped. The imaginary part of the impulse response of the subcarrier interference filter crosses the time axis at the integer multiples of the symbol period and the real part crosses the time axis at the odd multiples of half the symbol period [3]. The OQAM can be used in combination of the FBMC to avoid subcarrier interference without sacrificing the spectral efficiency. Specifically, this is achieved by doubling the symbol rate and using alternatively the real and the imaginary part for each subchannel. The FBMC-OQAM has good frequency localization (e.g., very low OOB emission) and high spectral efficiency (i.e., no CP overhead). Nevertheless, the FBMC-OQAM suffers from high complexity and long delay, i.e., $K - 1/2$ symbol durations. Another drawback is that it is not directly compatible with MIMO systems due to the staggered OQAM.

In the simulator, we adopt the PHYDYAS prototype filter with an (integer) overlapping factor K ranging from 2 to 4 [3]. A larger K results in better frequency localization, yet higher complexity and longer delay.

9.3.5 FBMC-QAM

To avoid the inconvenience caused by the OQAM, the QAM-based FBMC (FBMC-QAM) with dual filter banks was proposed [18]. A prototype filter in one filter bank is obtained by block interleaving the prototype filter in the other filter bank. The FBMC-QAM modulator performs filtering for even and odd subcarriers separately. The proposed filter banks in [18] ensure orthogonal subcarriers yet has higher OOB emission than that of the CP-OFDM. As a tradeoff between low OOB emission and high signal-to-interference ratio (SIR), two types of filters were proposed in [25]. Generally speaking, filters with lower OOB emission will result in lower SIR. The filter type I has a spectrum similar to that of the FBMC-OQAM with PHYDYAS filter (i.e., good frequency localization) yet poor SIR (10.6 dB). The filter type II has a wider spectrum, yet 9-dB higher SIR (19.6 dB) [25].

Like the FBMC-OQAM, the FBMC-QAM enjoys a high spectral efficiency (i.e., there is no CP overhead), yet suffers from high complexity and long delay. Unlike the FBMC-OQAM, the FBMC-QAM is compatible with MIMO systems at the cost of degradations in SIR and/or OOB emission.

In the simulator, the two types of filters in [25] and the filters in [18] are implemented for a fixed overlapping factor of $K = 4$. They are referred to as filter type I, II, and III, respectively.

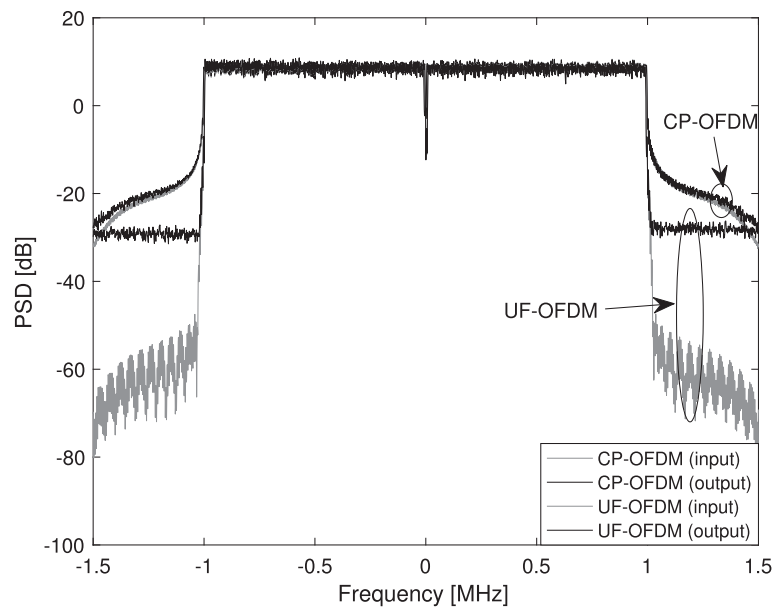
9.4 SIMULATION EXERCISES

In this section, we compare the performances of different waveforms using the simulator. The simulator is written using object-oriented programming in MATLAB. To run the simulator, one needs to first direct to the subfolder `[.../waveformSimulator/ptplink]` in MATLAB. System parameters (waveform, symbol rate, subcarrier number, etc.) can be changed in a file named `sSysParDefault.m`. Different analog impairments (PN, CFO, AWGN, etc.) can be switched on and off in a file named `sAnaParDefault.m`. After all the parameters have been set, one needs to run `setpath.m` to add the paths of all the necessary folders. Finally, running `runSim.m`, it will perform simulation of the chosen waveform. Some simulation examples can be found in the subfolder `[.../ptplink/examples]`.

9.4.1 SPECTRAL REGROWTH

As mentioned in the previous section, certain waveforms (e.g., UF-OFDM and FBMC-OQAM) outperform the CP-OFDM in terms of OOB emissions (at the cost of increased complexity). Nevertheless, there can be spectral regrowth when waveforms are subject to PA nonlinearities. Here we show the spectral regrowths of CP-OFDM, UF-OFDM, W-OFDM, and FBMC-OQAM using the simulator.

The measurement setup is given below. We assume a DFT size of 512, out of which 158 are reserved as a guard band (i.e., 79 subcarriers in the beginning and end of each multicarrier symbol are set to zero), and QPSK modulation. In order to correctly include the OOB emission due to the PA nonlinearity, an interpolation with an oversampling factor of 5 and the Parks–McClellan optimal finite impulse response (FIR) filter as the anti-aliasing filter are used. For UF-OFDM modulation, we additionally assume Dolph–Chebyshev filter (with 64 filter taps and 40-dB stop band suppression) as the prototype filter and 16 subcarriers per filter. For W-OFDM modulation, we additionally assume a time-domain raised-cosine window (with 32 tapering length at the left and right soft edges). For FBMC-OQAM modulation, we assume an overlapping factor of 4 ($K = 4$) unless otherwise specified.

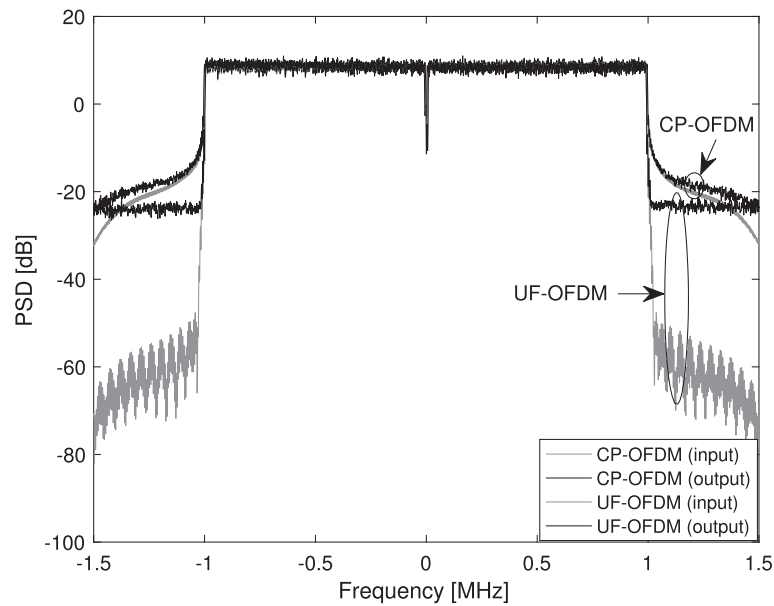
**FIGURE 9.2**

Spectral regrowths of CP- and UF-OFDM under PA nonlinearity: low input power (with 10-dB additional backoff).

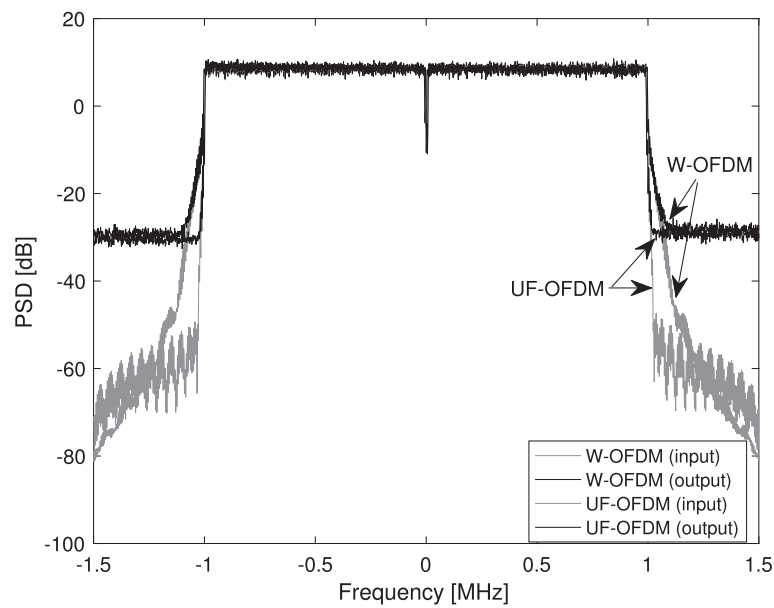
Figs. 9.2–9.6 show the spectral regrowths of CP-, UF-, and W-OFDM under PA nonlinearities. As can be seen, the UF-OFDM has lower OOB emission as compared with the CP-OFDM, especially at low input power. Nevertheless, as the input power increases, the OOB emission of the UF-OFDM increases more than that of the CP-OFDM. That is, the OOB advantage of the UF-OFDM over the CP-OFDM diminishes at high input power. It is noted that the OOB emission of the CP-OFDM can be greatly improved by introducing a time-domain window to it, i.e., the W-OFDM. The OOB emissions of W- and UF-OFDM become comparable under PA nonlinearities. It is apparent that the FBMC-OQAM with an overlapping factor of 4 has much narrower OOB emission than the UF-OFDM does. Nevertheless, their spectra overlap under PA nonlinearities. It should also be noted that, in the absence of PA nonlinearities, the OOB emission of the FBMC-OQAM with an overlapping factor of 2 ($K = 2$) becomes larger than that with an overlapping factor of four (yet still smaller than that of the CP-OFDM). For the conciseness of the paper, the corresponding results are omitted here. All in all, it is shown that OOB emissions can be improved by filter banks; however, the improvements become insignificant when the PA nonlinearities are taken into account.

9.4.2 IMPAIRMENT OF CFO

The carrier frequencies generated by oscillators at the transmitter and the receiver are usually not identical. The difference between the two carrier frequency is referred to as the carrier frequency offset (CFO). The CFO destroys the orthogonality of multicarrier waveforms, resulting in performance degradations for different waveforms. In this exercise, we study the CFO impairment to different waveforms by simulations.

**FIGURE 9.3**

Spectral regrowths of CP- and UF-OFDM under PA nonlinearity: high input power.

**FIGURE 9.4**

Spectral regrowths of W- and UF-OFDM with low input power.

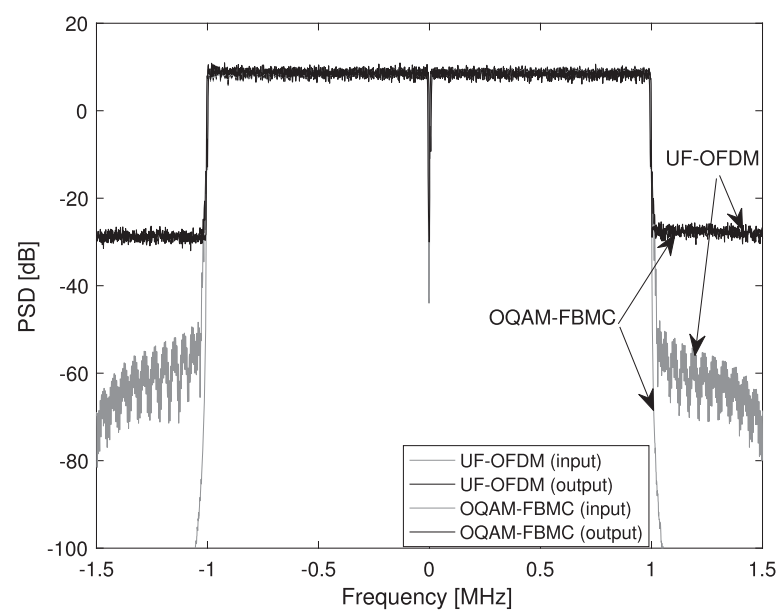


FIGURE 9.5
Spectral regrowths of UF-OFDM and FBMC-OQAM with low input power.

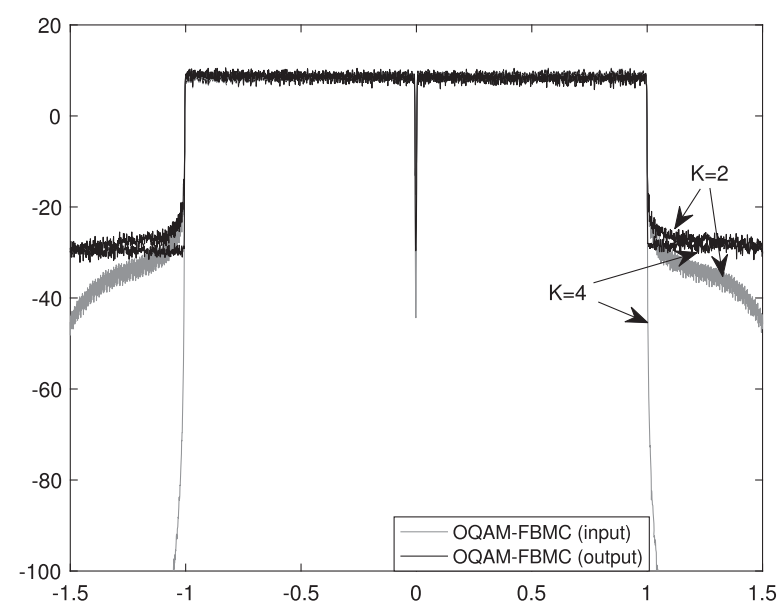
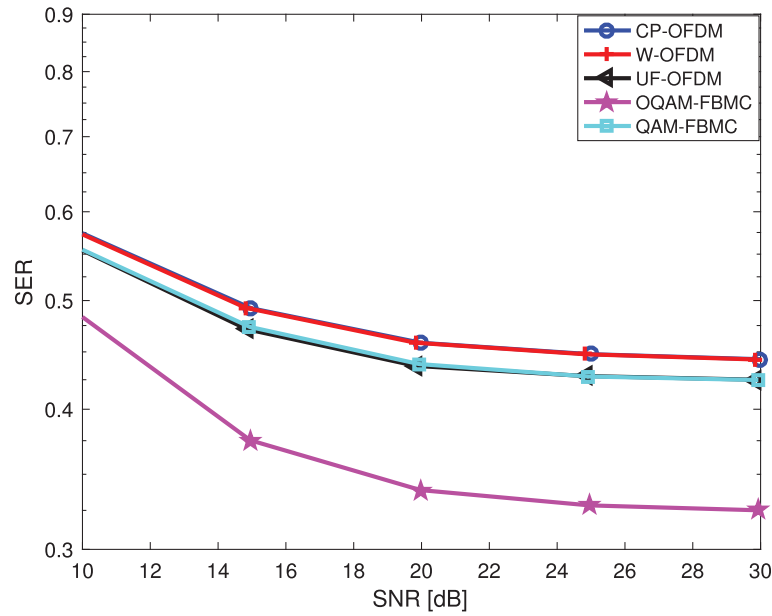


FIGURE 9.6
Spectral regrowths of FBMC-OQAM with different overlapping factors and low input power.

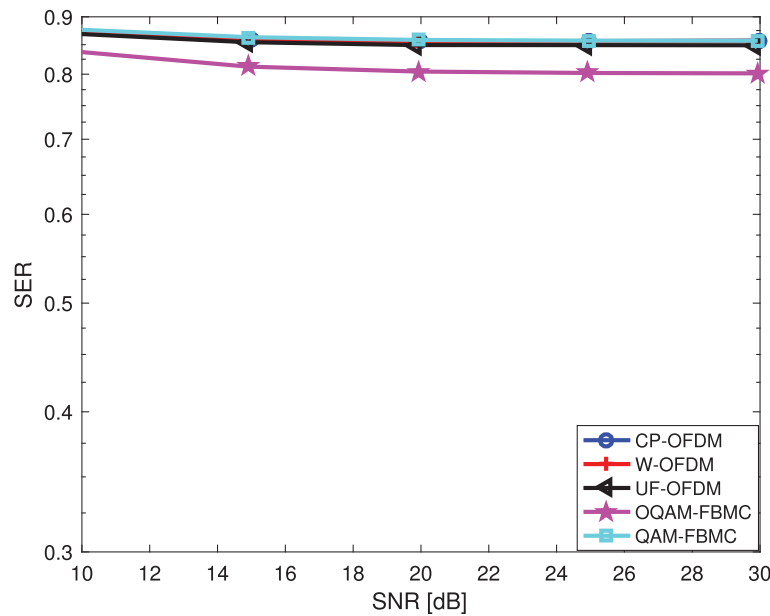
**FIGURE 9.7**

SER performances of different waveforms with 100 Hz CFO.

For fair comparisons of different waveforms, we disable the frequency synchronization module and assume 100 and 1000 Hz CFOs. In addition, we assume 1024 subcarriers, 100 MHz bandwidth, 16-QAM, and 100 multicarrier symbols. For the UF-OFDM, we assume the Dolph–Chebyshev filter (with 64 filter taps and 40-dB stop band suppression) as the prototype filter and 16 subcarriers per filter. For the W-OFDM, we assume a time-domain raised-cosine window with a tapering length of 32 at the left and right soft edges. For the FBMC-OQAM, we assume a PHYDYAS prototype filter with an overlapping factor of 4. For the FBMC-QAM, we assume the filter type II with an overlapping factor of 4 (cf. Section 9.3), since it is a good tradeoff between OOB emission and SIR [25]. In order to focus on the CFO effect, we ignore all the other hardware effects and assume an AWGN channel.

Fig. 9.7 and Fig. 9.8 show the symbol error rate (SER) performances of different waveforms under 100 and 1000 Hz CFOs, respectively. As can be seen, the FBMC-OQAM has the best SER performance and the UF-OFDM has the second best SER performance due to their filter bank. The SER performance of the FBMC-QAM depends on the value of the CFO and is no better than that of the UF-OFDM. The latter feature is because the filter type II in the FBMC-QAM is a compromise between OOB emission and SIR and because it does not guarantee orthogonality. Nevertheless, it should be noted that the SER improvements of the filter-bank-based waveforms (e.g., FBMC-OQAM and UF-OFDM) over that of the CP-OFDM tend to diminish (or become less significant) as the CFO increases.

As can be seen, without frequency synchronization, the system will not work. In practice, a coarse frequency synchronization (e.g., [23]) is performed prior to demodulation at the receiver. The residual CFO from the imperfect frequency synchronization can be further mitigated (together with the PN) using PN mitigation algorithms [8]. Since different waveforms apply different filters (which affect the signal powers of the waveforms), the exact SNR differs from the conventional E_S/N_0 (symbol energy

**FIGURE 9.8**

SER performances of different waveforms with 1000 Hz CFO.

over noise power density). As a result, we estimate the signal powers of different waveforms from their transmitted (time-domain) signals, respectively.

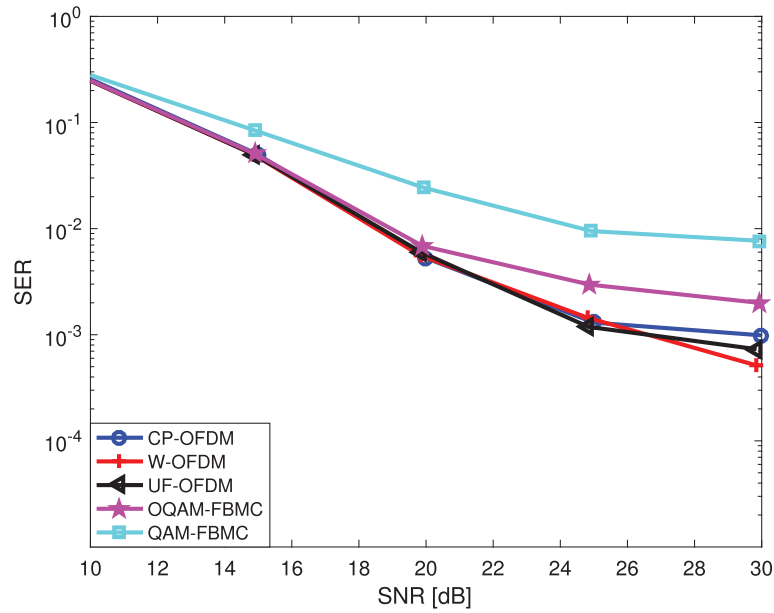
9.4.3 IMPAIRMENT OF PN

On top of the carrier frequency, the oscillator also produces random jitter frequency, which results in random PN. The PN causes random phase rotation of the signal and destroys the orthogonality of the multicarrier waveforms. In this exercise, we study the PN impairment to different waveforms by simulations.

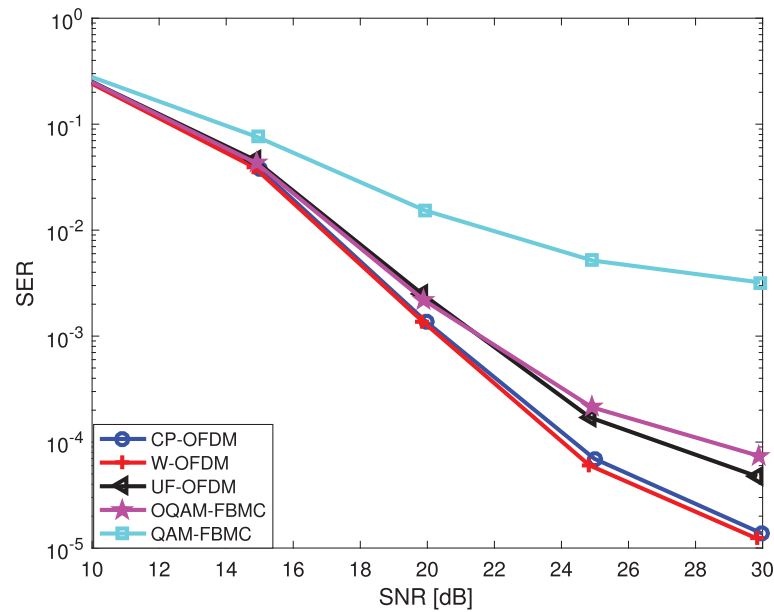
As mentioned in Section 9.2.3, the PN effects in the frequency domain can be categorized as CPE and ICI. While the CPE can easily be estimated, the ICI is different to track. The advanced ICI correction algorithms (see, e.g., [19], [7]) are usually computationally expensive and waveform-specific, and may not always be feasible.

We assume perfect timing and frequency synchronization, 1024 subcarriers, 100 MHz bandwidth, 16-QAM, and 100 multicarrier symbols. For the UF-OFDM, we assume the Dolph–Chebyshev filter (with 64 filter taps and 40-dB stop band suppression) as the prototype filter and 16 subcarriers per filter. For the W-OFDM, we assume a time-domain raised-cosine window with 32 tapering length at the left and right soft edges. For the FBMC-OQAM, we use the PHYDYAS prototype filter with an overlapping factor of 4. For the FBMC-QAM, we assume the filter type II with an overlapping factor of four (cf. Section 9.3) since it is a good tradeoff between OOB emission and SIR [25]. In order to focus on the PN effect, we ignore all the other hardware effects and assume an AWGN channel.

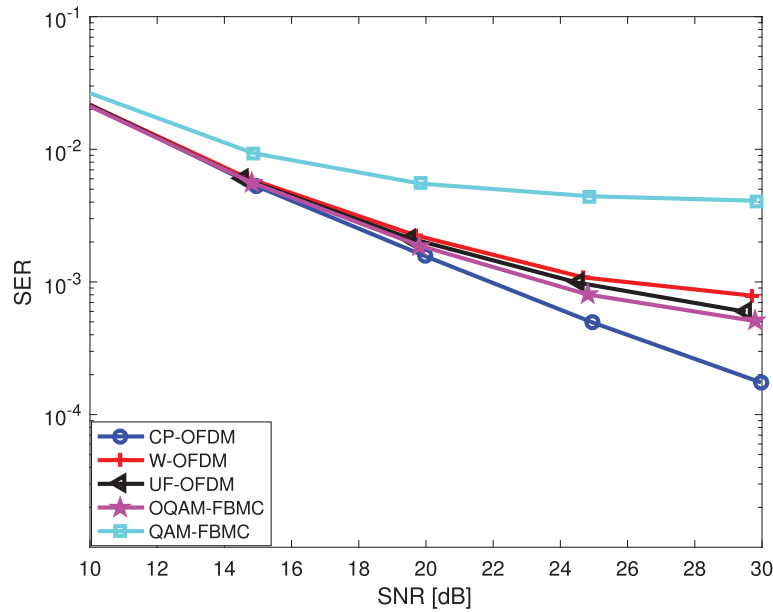
Fig. 9.9 and Fig. 9.10 shows the SER performance of different waveforms in the presence of PN without and with CPE correction, respectively. As can be seen, the CPE corrections improve the SER

**FIGURE 9.9**

SER performances of different waveforms subject to PN without any receiver compensation.

**FIGURE 9.10**

SER performances of different waveforms subject to PN with CPE correction.

**FIGURE 9.11**

SER performances of different waveforms in time-varying fading channel with a user speed of 60 km/h at 6 GHz.

performances of the considered waveforms. Unlike the (constant) CFO, the PN arises from random jitter frequency. In this case, the W- and CP-OFDM have better SER performances than the other filter-bank-based waveforms, especially after the CPE removal. Due to the intrinsic SIR, the FBMC-QAM (with filter type II) has the worst SER performance under the PN.

9.4.4 IMPAIRMENT OF FADING CHANNEL

The multicarrier waveforms were introduced to cope with the multipath effect of the propagation channel in broadband communications. In this exercise, we compare SER performances of different waveforms in the multipath fading channel. We assume perfect timing and frequency synchronization, 512 subcarriers, 120 MHz bandwidth, and QAM. Like before, we assume the Dolph–Chebyshev filter (with 64 filter taps and 40-dB stop band suppression) as the prototype filter and 16 subcarriers per filter for UF-OFDM, a time-domain raised-cosine window with 32 tapering length at the left and right soft edges for W-OFDM, the PHYDYAS prototype filter with an overlapping factor of 4 for FBMC-OQAM, and the filter type II with an overlapping factor of 4 for FBMC-QAM. In order to focus on the multipath channel effect, we assume perfect channel estimation, ignore all the hardware effects, and resort to the QuaDRiGa channel model [16]. Unlike the WINNER channel model [1], the QuaDRiGa channel model can well emulate the time variation of the fading channel [14]. We assume a user speed of 60 km/h at 6 GHz. For each channel drop (realization), 50 multicarrier symbols were transmitted. In total, 60 channel drops were simulated. The SER performances of different waveforms are shown in Fig. 9.11. As can be seen, the CP-OFDM has the best SER performance in the time-varying fading channel; the FBMC-OQAM, UF-OFDM, and W-OFDM have similar SER performances; and the

FBMC-QAM has the worst SER performance. Note that the performance of the FBMC-QAM can be improved by using the type III filter at the expense of a higher OOB emission than the CP-OFDM [18].

REFERENCES

- [1] P. Kyosti, J. Meinila, L. Hentila, et al., IST-4-027756 WINNER II D1.1.2 V1.2 WINNER II Channel Models: Part I Channel Models, 2007.
- [2] S. Afsardoost, T. Eriksson, C. Fager, Digital predistortion using a vector-switched model, *IEEE Transactions on Microwave Theory and Techniques* (ISSN 0018-9480) 60 (4) (2012, April) 1166–1174, <https://doi.org/10.1109/TMTT.2012.2184295>.
- [3] M. Bellanger, et al., FBMC physical layer: a primer, <http://www.ict-phydyas.org>, 2010, 06.
- [4] G. Berardinelli, Generalized DFT-S-OFDM waveforms without cyclic prefix, *IEEE Access* 6 (2018) 4677–4689, <https://doi.org/10.1109/ACCESS.2017.2781122>.
- [5] J.A.C. Bingham, Multicarrier modulation for data transmission: an idea whose time has come, *IEEE Communications Magazine* (ISSN 0163-6804) 28 (5) (1990, May) 5–14, <https://doi.org/10.1109/35.54342>.
- [6] X. Chen, A. Wolfgang, Phase noise mitigation in OFDM-based backhaul in the presence of channel estimation and synchronization errors, in: 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), 2016, May, pp. 1–5.
- [7] X. Chen, S. Zhang, Antenna mutual coupling effect on MIMO-OFDM system in the presence of phase noise, in: 2017 11th European Conference on Antennas and Propagation (EUCAP), 2017, March, pp. 653–657.
- [8] X. Chen, A. Wolfgang, A. Zaidi, MIMO-OFDM for small cell backhaul in the presence of synchronization errors and phase noise, in: 2017 IEEE International Conference on Communications Workshops (ICC Workshops), 2017, May, pp. 1221–1226.
- [9] A. Daher, E.H. Baghious, G. Burel, E. Radoi, Overlap-save and overlap-add filters: optimal design and comparison, *IEEE Transactions on Signal Processing* (ISSN 1053-587X) 58 (6) (2010, June) 3066–3075, <https://doi.org/10.1109/TSP.2010.2044260>.
- [10] M. Faulkner, The effect of filtering on the performance of OFDM systems, *IEEE Transactions on Vehicular Technology* (ISSN 0018-9545) 49 (5) (2000, Sep.) 1877–1884, <https://doi.org/10.1109/25.892590>.
- [11] K. Guan, G. Li, T. Kurner, A.F. Molisch, B. Peng, R. He, B. Hui, J. Kim, Z. Zhong, On millimeter wave and THz mobile radio channel for smart rail mobility, *IEEE Transactions on Vehicular Technology* (ISSN 0018-9545) 66 (7) (2017, July) 5658–5674, <https://doi.org/10.1109/TVT.2016.2624504>.
- [12] H2020-ICT-671650-mmMAGIC/D5.1 2016 mmMAGIC Deliverable D5.1. Initial multi-node and antenna transmitter and receiver architectures and schemes.
- [13] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1996.
- [14] F.H.H. Institute, Quasi deterministic radio channel generator user manual and documentation, document revision: v1.4.8-571, 2016, Sep.
- [15] J. Luo, A.A. Zaidi, J. Vihriälä, D. Giustiniano, et al., Preliminary radio interface concepts for mm-wave mobile communications. Deliverable D4.1, Millimetre-Wave Based Mobile Radio Access Network for Fifth Generation Integrated Communications (mmMAGIC), <https://5g-mmagic.eu/results>, 2016.
- [16] S. Jaeckel, L. Raschkowski, K. Borner, L. Thiele, Quadriga: a 3-d multi-cell channel model with time evolution for enabling virtual field trials, *IEEE Transactions on Antennas and Propagation* (ISSN 0018-926X) 62 (6) (2014, June) 3242–3256, <https://doi.org/10.1109/TAP.2014.2310220>.
- [17] C. Kim, K. Kim, Y.H. Yun, Z. Ho, B. Lee, J.Y. Seol, QAM-FBMC: a new multi-carrier system for post-ofdm wireless communications, in: 2015 IEEE Global Communications Conference (GLOBECOM), 2015, Dec., pp. 1–6.
- [18] H. Nam, M. Choi, C. Kim, D. Hong, S. Choi, A new filter-bank multicarrier system for QAM signal transmission and reception, in: 2014 IEEE International Conference on Communications (ICC), 2014, June, pp. 5227–5232.
- [19] D. Petrovic, W. Rave, G. Fettweis, Effects of phase noise on OFDM systems with and without pll: characterization and compensation, *IEEE Transactions on Communications* (ISSN 0090-6778) 55 (8) (2007, Aug.) 1607–1616, <https://doi.org/10.1109/TCOMM.2007.902593>.
- [20] D. Roque, C. Siclet, Performances of weighted cyclic prefix OFDM with low-complexity equalization, *IEEE Communications Letters* (ISSN 1089-7798) 17 (3) (2013, March) 439–442, <https://doi.org/10.1109/LCOMM.2013.011513.121997>.
- [21] P. Siohan, C. Siclet, N. Lacaille, Analysis and design of OFDM/OQAM systems based on filterbank theory, *IEEE Transactions on Signal Processing* (ISSN 1053-587X) 50 (5) (2002, May) 1170–1183, <https://doi.org/10.1109/78.995073>.

- [22] V. Vakilian, T. Wild, F. Schaich, S. ten Brink, J.F. Frigon, Universal-filtered multi-carrier technique for wireless systems beyond LTE, in: 2013 IEEE Globecom Workshops (GC Wkshps), 2013, Dec., pp. 223–228.
- [23] J.J. van de Beek, M. Sandell, P.O. Borjesson, ML estimation of time and frequency offset in OFDM systems, IEEE Transactions on Signal Processing (ISSN 1053-587X) 45 (7) (1997, Jul.) 1800–1805, <https://doi.org/10.1109/78.599949>.
- [24] G. Yi, L. Gang, G. Jianhua, A novel time and frequency synchronization scheme for OFDM systems, IEEE Transactions on Consumer Electronics (ISSN 0098-3063) 54 (2) (2008, May) 321–325, <https://doi.org/10.1109/TCE.2008.4560093>.
- [25] Y.H. Yun, C. Kim, K. Kim, Z. Ho, B. Lee, J.Y. Seol, A new waveform enabling enhanced QAM-FBMC systems, in: 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2015, June, pp. 116–120.

Index

A

ACEPR (adjacent channel error power ratio), 95
Active subcarriers, 24, 120, 123, 125, 128, 134, 136, 150, 195, 284
Additive white Gaussian noise, *see* AWGN
Adjacent channel error power ratio, *see* ACEPR
Adjacent channel leakage ratio (ACLR), 88, 99
Amplifier, 88, 90–92, 100, 146
Analog array, 217, 222
 architectures, 217, 218
Antenna array, 201–203, 215, 216, 219, 222, 230
 architectures, 215
Antenna effects, 37, 38
Antenna elements, 40, 144, 160, 200, 201, 203–207, 209, 215–219, 227, 229, 230, 235, 236, 245, 247, 249, 250
 dual-polarized, 219, 249
Antenna gains, 8, 38, 201, 239
Antenna patterns, 32, 37, 38, 41, 54, 75
Antenna ports, 100, 223, 224, 227, 229, 231–233, 236, 242
Architectures, hybrid array, 216, 218, 219
Array gain, 199, 202–208, 215, 218
 composite link, 203, 206
Asymmetric digital subscriber line (ADSL), 119, 159
AWGN (additive white Gaussian noise), 189, 190, 193, 194, 253, 280, 284, 286

B

Baseband complexity, 18, 144, 146, 161
Basis functions, 90–92, 97, 100
 local, 89, 93, 96, 97, 281
Beam
 correspondence, 239, 244, 245
 management, 29, 32, 217, 224, 230, 232, 235, 238–244
 wide, 239, 240, 242, 243
Beam-forming, 25, 33, 143, 144, 160, 201–204, 217, 219, 222, 225, 228–230, 235, 238, 239, 243, 245–247, 249, 250

Beam-forming gain, 245–247, 249, 250
 estimated, 245–247
Beam-forming weights, 203, 217, 223, 239
BEC (binary erasure channel), 265, 266, 268
Behavioral models, 34, 87, 109, 111, 115
Binary erasure channel, *see* BEC
Blocklength, 253, 255, 257, 258, 266, 267, 269, 270, 273, 275

C

Carrier-frequency offset, *see* CFO
CDD (cyclic delay diversity), 208
CFO (carrier-frequency offset), 159, 186–188, 191, 279, 280, 282–284, 286, 287, 290, 293
Challenges, 3, 7, 16, 21, 32
Channel bandwidths, 21, 161, 163
Channel capacity, 212, 255
Channel coding, 23
Channel correlation, 204–206
Channel delay, 139, 162, 170, 173
Channel estimation, 28, 29, 149, 223, 225, 226, 229, 231, 279, 280
Channel impulse response, *see* CIR
Channel knowledge, 207
 accurate, 211
 detailed, 204, 224, 226, 229
Channel matrix, 208–210, 212, 227
Channel measured, 50–52, 73
Channel model, 17, 59, 61, 74, 75, 82, 171, 172, 175, 248–250, 273, 277, 279–281
 geometry-based stochastic, 18, 279
 stochastic, 74, 281
Channel modeling, 17, 32, 35
Channel prediction, 193, 195, 197
Channel quality, 224, 229
Channel quality indicator, *see* CQI
Channel responses, 42–44, 50
Channel state information, *see* CSI
Channel time selectivity, 18, 160
 robustness to, 144, 160
Channel transfer functions (CTFs), 189, 193, 283
Channel vectors, 204, 205, 212

CIR (channel impulse response), 189, 190, 284
 Codebooks, 209, 225, 227–229, 232, 235–237
 Codes, convolutional, 270, 271, 275
 Codewords, 182, 254, 255, 259, 262, 269, 270, 272, 273
 Coding schemes, 18, 25, 29, 224, 254, 255, 257, 258, 271, 272, 275
 Commercial deployments, 1, 6, 14, 16
 Common phase error, *see* CPE
 Control-plane, 14, 22, 23
 CP (cyclic prefix), 24, 123–127, 129, 132, 141, 161–163, 166, 170, 185, 189, 193, 279, 280, 284, 285
 CP-OFDM, 24, 123, 126–128, 138, 139, 141, 142, 146–149, 151, 152, 235, 279, 284–287, 290, 293, 294
 waveform, 159, 162
 CPE (common phase error), 28, 153, 167–169, 186–188, 282, 291
 CQI (channel quality indicator), 29, 224, 226, 235
 CSI acquisition, 29, 224, 225, 227, 229–232, 235, 242
 CSI (channel state information), 25, 28, 200, 216, 224, 225, 227, 229–232, 234–236
 CSI-RS, 28, 29, 224, 231, 232, 235, 238, 241–245
 CTFs, *see* Channel transfer functions
 Cyclic prefix, *see* CP

D
 DAB (digital audio broadcast), 119, 159
 Data transmission, 23, 25, 27, 32, 201, 227, 231, 233–235, 237, 238, 240, 241, 253
 Demodulation, 28–30, 149, 153, 154, 156, 167, 186, 227, 230, 231, 280, 290
 Deployments, 8, 14, 15, 18, 24, 119, 143, 144, 159, 160, 165–167, 170, 172
 large cell, 142–144, 160
 DFT Spread OFDM (DFTS-OFDM), 24, 119, 141, 142, 149, 160, 179, 180, 184, 235, 279
 Digital array, 216, 217, 229
 Digital audio broadcast (DAB), 119, 159
 Distortion, 33, 87, 88, 95, 97, 98, 109, 112, 140, 180, 182, 195, 281, 284
 Dolph–Chebyshev filter, 284, 286, 290, 291, 293
 Dominant channel directions, 232
 Doppler effect, 163, 172, 186
 Doppler frequencies, 45

Doppler time domains, 48
 Downlink, 14, 23, 24, 26–28, 31, 142, 159, 162, 179, 180, 202, 224–226, 238, 243
 Downlink beam management, 238–241, 244
 Downlink channel, 225, 226, 235
 Downlink MIMO transmission, 232, 235
 Downlink transmissions, 24, 27, 119, 225, 226
 Dynamic TDD, 30, 145, 161

E
 Elevation, 41, 46, 50, 51, 59, 70–74, 78, 220, 245, 247, 281
 EMBB (Enhanced Mobile Broadband), 8–10, 15, 30, 31, 142, 143, 145, 172, 253, 258
 Equalization, 54, 68, 151, 279, 280, 283

F
 Fading channels, 18, 197, 205, 253, 271, 273, 274, 293
 time-varying, 149, 293
 Fast Fourier transform, *see* FFT
 FBMC symbol, 133, 135, 140
 FBMC waveforms, 122, 130, 133, 137, 138, 140, 151, 158
 FBMC-OQAM, 133–135, 137–139, 146–149, 152, 157, 279, 280, 285–287, 289–291, 293
 FBMC-QAM, 135, 137–139, 146, 149, 157, 158, 279, 280, 286, 290, 291, 293, 294
 FDD (frequency division duplex), 2, 31, 202, 209, 225–227, 229, 232, 235
 FEC (forward error correction), 253
 FFT (fast Fourier transform), 125, 127, 130, 185, 220
 Filter bank, 284–287, 290
 Filtered OFDM, 123, 127
 Frame, 6, 26
 structure, 21, 27, 30, 31, 166
 Frequencies
 corner, 107
 dependence, 38, 56, 58, 63, 75, 77, 201
 domain, 28, 29, 42, 43, 45, 119, 120, 122, 130, 132, 141, 143, 151, 161, 172, 175, 185, 186, 207, 208
 higher, 6, 7, 30, 33, 54, 58–60, 66, 68, 82, 87, 144, 159
 higher carrier, 18, 28, 119, 142, 143, 167, 170, 172
 increasing, 38, 54, 58, 75, 221
 lower carrier, 143, 144, 159
 random jitter, 291, 293

Frequency division duplex, *see* FDD
 Function, nonlinear, 92, 97, 109, 110

G

Generalized memory polynomial, *see* GMP
 GMP (generalized memory polynomial), 92, 93, 95–97, 279, 281
 GoB (grid-of-beams), 228, 245–249
 Grid-of-beams, *see* GoB
 Guard interval, 129, 151

H

Half-power beam width (HPBW), 220, 245, 247
 Hann filter, 43, 44, 51
 Hardware impairments, 16, 18, 21, 32, 87, 109, 115, 117, 145, 159, 167, 279–281
 High frequencies, 25, 33, 38, 66, 144, 145, 160, 161, 166, 200, 201, 221, 230, 242

I

IMT (international mobile telecommunications), 5, 8, 78
 Information bits, 254, 266–268, 275
 Information theory, 253, 255, 264
 Input power, low, 287–289
 Inter-symbol interference, *see* ISI
 Intercarrier interference, 133, 153, 167, 168, 172, 186
 Interference, 122, 123, 127, 133, 135, 175, 176, 200, 206, 209, 212–215, 217, 226, 227, 230, 234, 235, 280, 282
 intercell, 213, 214, 231, 234
 measurements, 29, 232, 235
 presubtraction, 212, 213
 suppression, 80, 199, 200, 210–212, 215, 224, 233
 International mobile telecommunications, *see* IMT
 International telecommunications union, *see* ITU
 Internet of things, *see* IoT
 Internumerology interference, 173, 175, 176
 IoT (internet of things), 1, 9, 15
 ISI (inter-symbol interference), 123, 124, 133, 140, 153, 155, 186
 ITU (international telecommunications union), 4, 6, 7, 10, 205

K

Kalman filter, 193–195, 197
 Key performance indicators, *see* KPIs
 Key technology components, 23, 229

KPIs (key performance indicators), 18, 144, 146, 161

L

Large bandwidths, 8, 81, 144, 146, 147, 161, 200, 216
 Layers, 22, 23, 26, 28, 71, 208–212, 214, 215, 224, 228, 231–233, 235–238, 247
 higher, 23, 26, 28, 244
 transmitted, 209, 211, 214
 LDPC code, 25, 261–263, 271
 Linear block codes, 258–261, 264, 269, 270
 Link adaptation, 27, 29, 171, 175, 225, 226
 Link types, 142, 159–162
 Listen-before-talk (LBT), 31, 164, 165
 Low latency, 1, 12, 30, 31, 142, 143, 145, 161, 165, 171, 172, 272
 LTE (long term evolution), 2–5, 14–16, 21, 23–26, 30, 140, 159–161, 163, 165, 179, 199, 200, 208, 209, 223, 224, 226–231, 233–235
 LTE numerology, 165, 166

M

Maximum ratio transmission, *see* MRT
 Measurement data, 55, 56, 73, 74, 78, 80, 84, 93, 95, 96
 Medium access control (MAC) layer, 22, 23
 Millimeter-wave, 3, 6, 7, 25, 30–33, 59, 60, 87, 88, 142, 143, 164, 165, 200, 201, 216, 217, 219–224, 229, 230, 244, 245, 247, 248, 281, 282
 MIMO (multiple-input multiple-output), 25, 33, 48, 137, 142, 143, 199, 277
 Mini-slots, 26, 30, 31, 164, 165, 172
 MMTC (massive Machine Type Communications), 8–10, 23, 142, 143, 145, 172, 253
 Mobile communications, 1–3, 5, 6, 12, 24, 35, 66, 74, 82, 84
 Model parameters, 52, 62, 64, 95
 Modeling error, 93, 189, 190
 Models
 baseline, 55, 79–81
 phase-noise, 105, 169, 279
 system, 101, 102
 MRC (maximum ratio combining), 204, 206, 207, 215
 MRT (maximum ratio transmission), 113, 205–207, 230
 Multiantenna techniques, 8, 18, 24, 25, 87, 199, 200, 202, 203, 215, 216, 223, 224, 230

Multicarrier waveforms, 18, 24, 119–121, 140, 141, 146, 148, 149, 160, 167, 186, 282, 284, 287, 291, 293
 Multipath channel, 119, 123, 125, 185, 212, 282, 283
 Multipath components, 44, 51, 53, 78–80
 Multiple antennas, 199, 200, 206, 208, 215, 257
 Multiple-input multiple-output, *see* MIMO
 Mutual coupling, 88, 89, 100–102

N

Narrowband interference, 120
 Narrowband subcarriers, 176–178
 New radio, 3
 NLoS measurements, 66, 67, 69
 NLoS (non-line-of-sight), 33, 48, 53, 58, 59, 61, 67–69, 79, 201, 205, 206, 245, 250
 Noise, 48, 52, 60, 73, 104–106, 203, 204, 214, 215, 230, 253
 additive, 168, 189, 190, 284
 Non-line-of-sight, *see* NLoS
 Non-stand-alone (NSA), 6, 14
 Number theory, 268
 Numerologies, 24, 26, 126, 163–166, 170–178
 aggressor, 175, 176
 subcarriers of, 176–178

O

OFDM based waveforms, 123, 130, 131, 137
 OFDM numerologies, 24, 30, 163, 166, 167, 172
 OFDM (orthogonal frequency division multiplexing), 2, 24, 119–121, 131, 132, 134, 135, 137, 140, 141, 146, 147, 149, 151, 157–163, 167, 168, 175, 176, 179, 279, 280
 OFDM subcarrier spacing, 29, 163, 167
 OFDM symbols, 26, 28–32, 124, 126, 127, 150, 153, 163, 164, 166, 168, 182, 183, 186, 191, 193–195, 197, 272, 273
 core, 126, 127
 OFDM systems, 113, 160, 167, 168, 170, 172, 179, 183–188, 191, 208, 216, 224
 OFDM waveforms, 122, 124, 125, 127, 140, 141, 179
 OOB, *see* Out-of-band
 Orthogonal frequency division multiplexing, *see* OFDM
 Orthogonality, 110, 112, 287, 291
 Oscillator, 29, 104, 105, 107, 282, 287, 291
 Oscillator frequencies, 154, 170

Out-of-band (OOB), 24, 127, 145, 146, 179, 180, 195, 281

Overlapping factor, 133–137, 147, 151, 285–287, 289–291, 293

P

PA, *see* Power amplifier
 Packet data convergence protocol (PDCP) layer, 22
 Panoramic photographs, 67, 68, 73
 PAPR (peak-to-average-power-ratio), 119, 141, 144, 148, 149, 160, 181, 182
 PAPR reduction, 160, 182
 schemes, 183, 184
 Partial transmit sequence (PTS), 182, 183
 Peak-to-average-power-ratio, *see* PAPR
 Penetration losses, 33, 57, 58, 77, 143
 Phase locked loop, *see* PLL
 Phase-noise
 oscillator, 18, 28, 88, 279, 280
 power spectral density, 104–106
 Phase shifters, 217–219
 PHY layer, 23, 28
 PHYDYAS prototype filter, 285, 290, 291, 293
 Physical broadcast channel (PBCH), 27, 232
 Physical channels, 26
 Physical downlink control channel (PDCCH), 27, 238, 241, 243, 244
 Physical downlink shared channel (PDSCH), 26, 27, 238, 243, 244
 Physical random access channel (PRACH), 27, 239
 Physical signals, 26, 28
 Physical uplink control channel (PUCCH), 27, 242, 245
 Physical uplink shared channel (PUSCH), 27, 235, 236, 242
 Pilots, scattered, 190, 191, 195, 282, 285
 PLL (phase-locked loop), 34, 104, 105, 107, 111, 152, 167, 282, 283
 PMI (precoding matrix indicator), 29, 224, 227, 228, 235, 242
 Polar codes, 25, 253, 258, 264, 266–269, 271
 Polarizations, 38, 41, 203, 207, 216, 219, 220, 228, 232, 233, 247, 249, 264
 Power amplifier, 17, 18, 33, 87–91, 93, 97, 100–103, 110, 111, 114, 116, 140, 141, 144, 147, 148, 179, 205, 279–281
 modeling, 89, 93, 97, 100

Power delay profiles, 53, 54, 59, 62, 67–69, 73, 78, 114

Power spectral density, *see* PSDs

PRB (same as Resource Block (RB)), 26, 29, 175, 176, 179

Precoder matrix, 208, 209, 227, 232

Precoders, 29, 205, 210, 212, 213, 227, 228, 231, 233, 236, 237

Precoding, 202, 204–206, 208, 210, 211, 213, 214, 225, 227, 235

Precoding matrix indicator, *see* PMI

Propagation, 35–37, 40, 84, 215, 226
 distance, 59, 69–71, 73, 79, 80
 free-space, 36, 37, 39, 41, 70, 201, 203
 radio wave, 16, 21, 32

Prototype filters, 120–123, 127, 130–134, 136, 147, 284–286, 290, 291, 293

PSDs (power spectral density), 95, 96, 104, 106, 108, 115, 134, 173, 184, 189

PTS (partial transmit sequence), 182, 183

Q

QoS, *see* Quality of service

QuaDRiGa channel model, 149, 191, 197, 293

Quality of service (QoS), 3, 18, 22, 159, 162

R

Radiation patterns, 98, 99, 205

Radio interface technologies, 5

Radio link control (RLC) layer, 22, 23

Radio link failure (RLF), 23, 244

Radiowave propagation, 35, 36, 48

Rank indicator, *see* RI

RB, *see* Resource block

Received signal vector, 207, 208, 214

Receiver side, 126, 132, 274

Reciprocity, 202, 225–227, 229, 231, 232, 235, 237, 239

Reference signal received power (RSRP), 29, 224

Reference signals, 25, 28, 30, 31, 178, 223–225, 227, 229–232, 239, 240, 243–245
 received, 225, 238, 239, 241

Resource block (RB), 26, 128, 175–179, 230, 272–275

Retransmission, 27, 244

RF power amplifiers, 33

RFICs (radio frequency integrated circuits), 219, 220

RI (rank indicator), 29, 224, 235, 242

Robustness, 18, 144, 160, 161, 199, 238, 240–242, 253
 phase-noise, 144, 146, 160, 161

S

Sampling frequency offset (SFO), 187, 188, 191, 193, 195–197

Scattering channel, 199, 204, 206, 207

Self-interferences, 123, 135, 136

Service data adaptation protocol (SDAP) layer, 22

Sidelink, 144, 145, 159–162, 179

Signal
 bandwidth, 128, 147, 160, 163
 processing, 16, 199

Signal-to-interference ratio, 136, 153, 154, 157, 158, 168, 169, 187, 199, 286, 290, 291
 achievable, 153, 154, 157, 158, 168, 169, 187

Simulation, 113, 146, 147, 171, 190, 233, 234, 250, 286, 287, 291
 exercises, 18, 279

Simulator, 279–286
 open-source, 18, 279, 280

Single-carrier waveforms, 119, 140, 141, 146, 159, 186

SIR, *see* Signal-to-interference ratio

Slot, 26–28, 30–32, 140, 163, 164, 166, 170, 172, 231, 272

SNR (signal-to-noise ratio), 49, 120, 190, 199, 203–205, 207–210, 212, 254, 257, 268

Sounding reference signal, *see* SRS

SRS (sounding reference signal), 28, 29, 231, 238, 244, 245

SSB (synchronization signal block), 232, 238–245

Stochastic model, 110, 111, 113, 115

Subarrays, 218, 219, 245, 246

Subcarrier bandwidth, 172, 173, 186

Subcarrier spacing, 24, 26, 137, 160–164, 166–168, 170, 172–174, 176, 190, 191
 function of, 154, 157
 wider, 164, 170, 171, 173

Subcarriers, 26, 28, 29, 119–123, 127–129, 133–138, 149–151, 153–156, 167, 168, 172, 173, 175–179, 182, 186–189, 216, 282–286, 290, 291
 demodulated, 153, 155, 157, 168, 185–187

Subchannels, 208, 209, 268, 285

Subframe, 26, 140, 165, 230

Symbols, pilot, 273–275, 277

Synchronization signal block, *see* SSB

Synthetic channels, 265–268

System

practical, 191, 193, 209, 210, 213

simulations, 245, 248, 249

T

Tail-biting (TB), 258, 270, 271, 275

TB, *see* Tail-biting

TDD (time division duplex), 2, 30, 31, 143, 145, 202, 226, 227, 229, 231, 232, 235

Time division duplex, *see* TDD

Time domain, 26, 28, 29, 45, 47, 123, 126, 130, 143, 145, 147, 151, 153, 161, 164, 285

Time-frequency resources, 21, 26–28

Timing offset, 185, 186, 282

TM, *see* Transmission mode

TPMI (transmit precoder matrix indicator), 236

Tracking reference signal (TRS), 29, 244

Transmission loss, 7, 25, 32, 37–39, 48, 49, 55, 56, 59, 75, 83, 84, 119, 201, 211, 281

free-space basic, 37–39

Transmission mode (TM), 208, 227, 230, 231

Transmission points, 247, 248

Transmission ranks, 209, 224, 233, 235, 236

Transmission schemes, 180, 227, 232, 235

Transmitter, 57, 58, 61, 151, 175, 181, 193, 199, 202–212, 214, 238, 239, 277, 280, 282, 284, 285, 287

U

UF-OFDM, 123, 128, 146, 149, 151, 279, 280, 284–291, 293

waveform, 128, 129

ULA (uniform linear array), 99, 205, 206, 227, 228

Ultra-reliable low-latency communications, *see*

URLLC

Uniform linear array, *see* ULA

Uniform planar array (UPA), 219, 228

Unlicensed spectrum, 30, 31, 142, 164, 165

Uplink, 14, 23, 24, 28–31, 119, 142–144, 159, 161, 162, 179, 202, 209, 225, 226, 235, 239, 245

Uplink and downlink, 23, 24, 26, 180, 225, 226

Uplink and sidelink, 145, 160, 161

Uplink beam management, 29, 238, 244

Uplink measurements, 202, 229, 230

Uplink transmissions, codebook-based, 235, 236

URLLC (ultra-reliable low-latency communications), 8–10, 30, 31, 142, 145, 161, 172, 253, 269, 272, 274, 276, 277

User-plane, 14, 22, 23

V

Vector network analyzer (VNA), 48

Volterra series, 88–90, 92, 281

W

W-OFDM, 123, 125, 127, 130, 146, 149, 151, 175, 176, 279, 280, 284, 286, 287, 290, 291, 293

Waveform design requirements, 119, 161, 162

Wide range of frequencies, 21, 24, 142, 161, 166, 170

Wireless communications, 35, 203

Z

ZF receiver, 214, 215