CHAPTER 5

New Radio Access RF and Transceiver Design Considerations

The NR RF transceiver characteristics are related to the frequency bands in which the 5G systems will be deployed. Due to wide range of the target frequency bands, spectrum flexibility was required for the new radio in order to operate in diverse spectrum allocations. While spectrum flexibility has been used in the previous generations of radio access technologies, it has become more important for NR development and deployment. Such spectrum flexibility is manifested as feasibility of deployment and resource allocations in frequency blocks of different sizes over an extremely wide range of contiguous or non-contiguous spectrum, both in the form of paired and unpaired frequency bands along with aggregation of different spectrum blocks within and across different bands. The NR has the capability to operate with mixed OFDM numerologies over the same or different RF carrier(s) and has relatively more flexibility compared to LTE in terms of frequency-domain scheduling and multiplexing of devices over the serving base station (BS) RF carrier(s). The use of OFDM waveform in NR provides the desired flexibility in terms of the size of the spectrum allocation and the instantaneous transmission bandwidth adaptation. The application of active antenna system (AAS) concept and multiple antennas in the base stations and the devices, which emerged during LTE development, has taken a giant leap in NR with the support of massive MIMO and control/data channel beamforming both in the existing LTE bands and in the new mmWave bands. Aside from physical layer design implications, the advent of the latter features significantly impact the analog/digital RF hardware system design/implementation including filters, amplifiers, data converters, antennas, etc.

In this chapter, we will discuss the new radio spectrum, RF characteristics, implementation considerations of the NR base stations and the devices as well as the hardware technologies that are used to implement various features of the new radio. In the course of this chapter, we will discuss various types of NR base stations and their external interfaces over which the RF requirements are defined. We further explore the conducted and over-the-air (OTA) RF requirements specified by 3GPP for testing and evaluating the performance of the NR base stations and devices.

5G NR. DOI: https://doi.org/10.1016/B978-0-08-102267-2.00005-1

5.1 NR Radio Parameters and Spectrum

The requirement for spectrum flexibility was a key driving factor for the adoption of OFDM-based technologies in 3GPP LTE which continues to be a major driver for the NR frequency planning and deployments. The need for diverse spectrum allocations in terms of spectral bands, operation bandwidths, duplex schemes (paired and unpaired spectrum) and multiple-access schemes emerged during the 3G and 4G deployments and has led to one of the most distinctive characteristics of 5G NR, which supports a large and diverse spectrum from 450 MHz to 52.6 GHz (and up to 100 GHz in future releases). The maximum frequency currently under study in ITU-R is 86 GHz (see Fig. 5.1). The NR supports operating bandwidths of 5 MHz to 3.2 GHz for both paired and unpaired spectrum as well as supplementary downlink (SDL) or supplementary uplink (SUL) carriers. The NR defines various frequency bands within the 5G spectrum. Although the boundaries of the NR frequency bands can vary in different countries and regions, it must be possible to efficiently allocate RF carriers at positions where the spectrum blocks are used with minimal spectrum wastage. This requires carefully defined channel raster for carrier allocation. There are also a number of other spectrum blocks being considered by regulatory bodies that were not initially considered by 3GPP specifications, which include 5925−7150 MHz in the United States and 5925−6425 MHz by CEPT[1] for unlicensed use as well as frequency range 64−86 GHz for extremely wideband applications [12]. Note that various parts of the 64−86 GHz range are allocated differently in different regions of the world. For instance, 64−71 GHz is set aside for unlicensed use in North America and is under consideration in CEPT (66−71 GHz); 66−76 and 81−86 GHz are under study in ITU-R as possible bands for IMT-2020 to be ratified during WRC-19 [12].

In general, the NR UEs do not receive or transmit using the full channel bandwidth of the gNBs, rather they can be assigned to what is referred to as bandwidth parts. While the concept does not have any direct RF implications, it is important to note that the gNB and the UE channel bandwidths are defined independently, and the device bandwidth capability does not have to match the gNB channel bandwidth. A unified frame structure is defined in NR that supports TDD, FDD, and half-duplex FDD operations. The duplex scheme is specifically defined for each operating band. Some bands are also designated as SDL or SUL bands to be used in FDD operation. Some of the frequency bands that have been identified for NR deployment are the existing ITU-R IMT[2] bands which may have already been accommodating 2G, 3G, and/or 4G [incumbent] deployments. In some regions, a number of

[1] The European Conference of Postal and Telecommunications Administrations (CEPT) was established in 1959 and is an organization where policy makers and regulators from 48 countries across Europe collaborate to harmonize telecommunication, radio spectrum, and postal regulations across Europe. The CEPT conducts its work through three autonomous business committees (ECC, Com-ITU and CERP).

[2] The term International Mobile Telecommunications (IMT) is the generic term used by the ITU community to designate broadband mobile systems. It encompasses IMT-2000, IMT-Advanced, and IMT-2020 collectively.

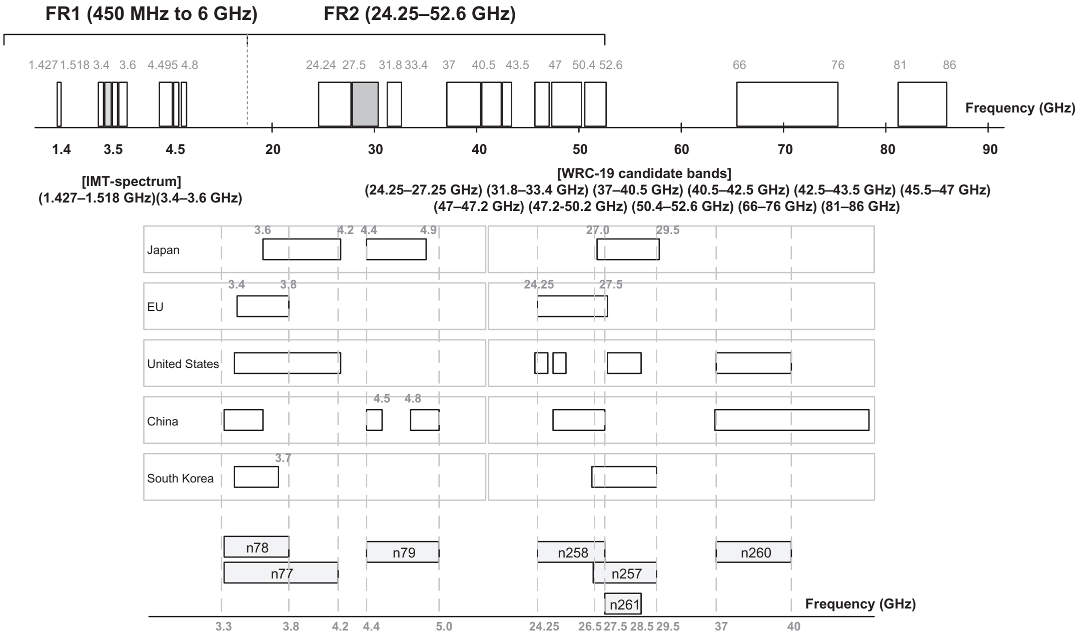

Figure 5.1
NR bands and 5G frequency bands under study in ITU-R WRC-19 [14,29].

frequency bands are designated and regulated as technology-neutral, which means that the coexistence between different technologies is a requirement for deployment. The capability to operate in this wide range of frequency bands for any cellular system, including NR, has direct implications on the RF requirements. The operators in the same band may deploy NR or other IMT technologies such as LTE [11]. Such coexistence requirements are typically specified within 3GPP (and/or ITU-R), but there may also be regional requirements defined by regulatory bodies that must be satisfied in order to be able to deploy the technology. Mobile operators have diverse spectrum holdings and allocations, which in many cases consist of a spectrum block that cannot accommodate one wideband carrier, or the allocation may be non-contiguous, consisting of multiple frequency blocks across multiple bands. In these cases, the NR specifications support carrier aggregation, where multiple carriers within a band, or in multiple bands, can be combined to create effectively wider transmission bandwidths. As shown in Fig. 5.1, the 5G spectrum in various countries is different which makes it difficult to harmonize spectrum and achieve global roaming.

The LTE and NR coexistence in the same spectrum, which is required for non-standalone and early deployments of 5G systems, makes it possible to deploy NR in the existing LTE frequency allocations. Since the co-channel NR and LTE carriers need to be aligned at subcarrier level, some restrictions are imposed on the NR channel raster in order to align the position of the NR and LTE carriers. The NR further supports multiple numerologies with subcarrier spacing (SCS) ranging from 15 to 120 kHz, with direct implications on the time and frequency structures. The subcarrier spacing has certain implications on the RF front-end in terms of the roll-off of the transmitted signal, which impacts the guard bands that are allocated between the transmitted resource blocks and the edge of the frequency band. The NR also supports mixed numerologies on the same carrier, which has further RF implications since the guard bands may need to be different at the two edges of the band.

In some cases, there are certain limitations concerning the site where the gNB equipment is deployed. The candidate sites are often shared between operators or an operator deploys multiple technologies at one site, which creates additional requirements for the gNB transmitters and receivers to operate in close proximity of other BS transceivers (e.g., blocking effects). The coexistence between operators of TDD systems in the same band is in general provided by inter-operator synchronization in order to avoid interference between downlink and uplink transmissions of different operators. This means that all operators need to have the same uplink/downlink configurations and frame synchronization, which is not by itself an RF requirement, but it is implicitly assumed in the 3GPP specifications. The RF requirements for unsynchronized systems are inevitably much stricter. The frequency bands are regionally defined, and new bands are added continuously for each generation of mobile systems, which means each new release of 3GPP specifications will incorporate additional bands. Using a release-independent principle, it is possible to design devices based on an early release of 3GPP specifications that support a frequency band added in a later release. The first set of NR bands is defined in 3GPP Rel-15, and additional bands will be added in a release-independent manner.

The NR can operate over a wide range of frequencies which include sub-6 GHz and above 6 GHz spectrum. The support of this extremely wide range of frequencies implies that the radio characteristics and operating parameters can significantly vary depending on the frequency band and the operating bandwidth; thus it is desirable to be able to configure the radio-related parameters such as subcarrier spacing, OFDM symbol length, cyclic prefix length, channel bandwidth according to the allocated frequency band. The NR specifies two frequency ranges, namely, FR1 and FR2 where each support certain OFDM numerologies. In FR1, which covers 450 MHz to 6 GHz, the subcarrier spacings that can be used for data transmission are 15, 30, and 60 kHz. The FR2 supports subcarrier spacings of 60 and 120 kHz. The UEs are required to support all subcarrier spacings except 60 kHz in FR1. The ratio of cyclic prefix length over OFDM symbol length is the same for all supported subcarrier spacings. Therefore, as the subcarrier spacing increases, the cyclic prefix length decreases, resulting in more susceptibility of NR signals to multipath delay distortion and coverage reduction. The use of larger subcarrier spacings further increases the NR signal tolerance to Doppler shift. It must be noted that the effect of oscillator phase noise is more pronounced in FR2 bands, resulting in inter-carrier interference. The use of larger subcarrier spacing helps mitigate the effects of phase noise. The subcarrier spacing is configured by the network using higher layer signaling. The most appropriate subcarrier spacing can be configured according to the deployment scenario. The FR2 signal propagation is characterized with less diffraction, higher penetration loss, and in general higher path loss. This can be compensated by incorporating more antenna elements at the transmit and receive sides, narrower beams with higher antenna gains, leading to massive MIMO systems.

LTE initially supported channel bandwidths up to 20 MHz, and later through the use of carrier aggregation, it was able to operate in up to 100 MHz bandwidths. The number of component carriers was later increased to 32 in 3GPP Rel-13, resulting in the maximum operating bandwidth of up to 640 MHz. The Rel-15 NR supports inter-band and intra-band contiguous and non-contiguous carrier aggregation as well as dual connectivity which allows simultaneous communication to NR and LTE base stations. Due to much wider channel bandwidths targeted in NR, the maximum channel bandwidth per component carrier has been increased to 100 MHz in FR1 (using SCSs of 30 and 60 kHz) and 400 MHz in FR2 (using 120 kHz SCS). Support of 400 MHz bandwidth in FR2 is optional; however, the NR devices are required to support up to 200 MHz in FR2. The NR new bands and the corresponding channel bandwidths and SCSs are shown in Table 5.1 [29].

The LTE spectrum utilization (defined as the ratio of the transmission bandwidth over the channel bandwidth) was 90% and guard bands were provisioned on both sides of the transmission bandwidth to protect communication in the adjacent channels by limiting the adjacent channel interference. In NR, the spectrum utilization has been increased to 98% (depending on transmission bandwidth), as a result of time and frequency-domain preprocessing of the OFDM signal (i.e., windowing and spectral shaping). While the size of the guard bands have been reduced in NR due to increased spectrum utilization, the out-of-band (OOB) emission and the permissible leakage power requirements have remained the same as LTE.

Table 5.1: NR new bands and channel bandwidths ("x" configurable and "*" optional) [5,29].

Frequency Range	Frequency Band		Subcarrier Spacing (kHz)	Component Carrier Bandwidth (MHz)														
	Band Number	Frequency Band (GHz)		5	10	15	20	25	30	40	50	60	70	80	90*	100	200	400*
FR1	n77/ n78	3.7	15		x	x	x			x	x						–	–
			30		x	x	x			x	x	x		x	x	x	–	–
			60		x	x	x			x	x	x		x	x	x	–	–
	n79	4.5	15							x	x						–	–
			30							x	x			x		x	–	–
			60							x	x			x		x	–	–
FR2	n257/ n258	24–40	60	–	–	–	–	–	–	–	x	–	–	–	–	x	x	
	n260/ n261		120	–	–	–	–	–	–	–	x	–	–	–	–	x	x	x

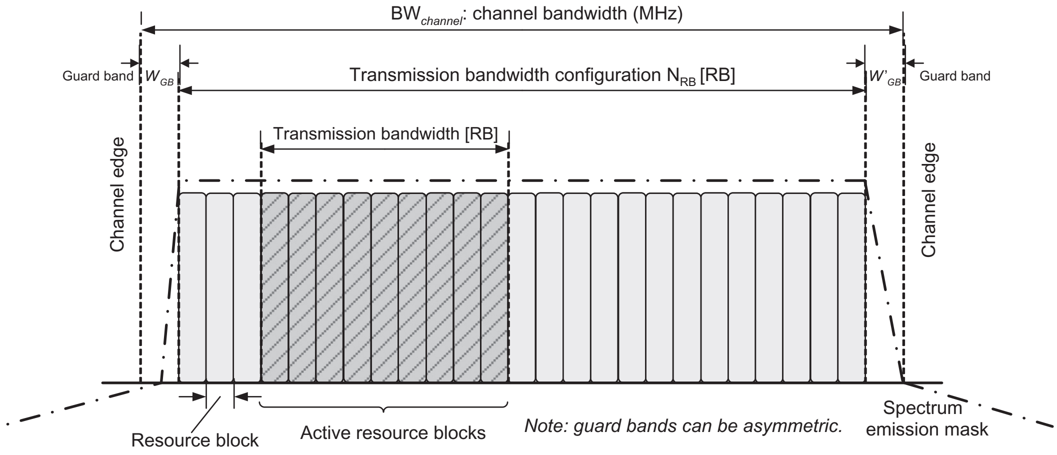

Figure 5.2
Illustration of channel bandwidth, transmission bandwidth, and guard bands [5].

The channel bandwidth of a gNB (BW$_{channel}$) supports transmission on a single (NR) uplink or downlink RF carrier. Different UE channel bandwidths may be supported within the same gNB channel bandwidth for bidirectional communication with the UE. The relative location of the UE channel bandwidth is flexible and within the gNB channel bandwidth. The gNB is able to transmit/receive to/from one or more UE bandwidth parts that are smaller than or equal to the total number of carrier resource blocks on the RF carrier. The relationship between the channel bandwidth, guard band, and the transmission bandwidth is shown in Fig. 5.2. The transmission bandwidth configuration in terms of the number of physical resource blocks (N_{RB}) for each *BS channel bandwidth* and SCS was given in Table 3.9, for FR1 and FR2. The minimum guard band for each *BS channel bandwidth* and SCS is shown in Table 5.2 for FR1 and FR2. The minimum guard band shown for SCS of 240 kHz is only applicable when an SS/PBCH block (SSB) with SCS of 240 kHz is placed adjacent to the edge of the *BS channel bandwidth* within which the SSB is located; otherwise, 240 kHz subcarrier spacing is not used for any other configurations. The number of RBs configured within any *BS channel bandwidth* will guarantee that the minimum guard band shown in Table 5.2 is satisfied.

If multiple numerologies are multiplexed over the same OFDM symbol, the minimum guard band on each side of the carrier is the guard band applied at the configured *BS channel bandwidth* for the numerology that is transmitted/received adjacent to the guard band. In FR1, if multiple numerologies are multiplexed over the same OFDM symbol and the *BS channel bandwidth* is greater than 50 MHz, then the guard band that is inserted adjacent to 15 kHz SCS is the same as the guard band defined for 30 kHz SCS for the same *BS channel bandwidth*. In FR2, if multiple numerologies are multiplexed over the same OFDM symbol

Table 5.2: Minimum guard band (kHz) for FR1 and FR2 [5].

Subcarrier Spacing (kHz)	Bandwidth (MHz)																
	FR1													FR2			
	5	10	15	20	25	30	40	50	60	70	80	90	100	50	100	200	400
15	242.5	312.5	382.5	452.5	522.5	592.5	552.5	692.5	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
30	505	665	645	805	785	945	905	1045	825	965	925	885	845	N/A	N/A	N/A	N/A
60	N/A	1010	990	1330	1310	1290	1610	1570	1530	1490	1450	1410	1370	1210	2450	4930	N/A
120	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1900	2420	4900	9860
240	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3800	7720	215,560

and the *BS channel bandwidth* is larger than 200 MHz, then the guard band inserted adjacent to 60 kHz SCS is the same as the guard band defined for 120 kHz SCS for the same *BS channel bandwidth* [5].

For each *BS channel bandwidth* and each numerology, the *BS transmission bandwidth configuration* is required to satisfy the minimum guard band requirement. The common resource blocks (see Section 3.7.1) are specified for each numerology and the starting point of the associated transmission bandwidth configuration on the common resource block grid for a given channel bandwidth is indicated by an offset to reference point A in the unit of the numerology. For each numerology, all *UE transmission bandwidth configurations* that are indicated to the UEs by the serving gNB through higher layer parameter *carrierBandwidth* must be located within the *BS transmission bandwidth configuration* [5].

In carrier aggregation scenarios, the transmission bandwidth configuration is defined per component carrier. An *aggregated BS channel bandwidth* and *guard bands* are defined for intra-band contiguous carrier aggregation. The *aggregated BS channel bandwidth* ($BW_{channel\_CA}$) is defined as $BW_{channel\_CA} = f_{edge\text{-}high} - f_{edge\text{-}low}$ (MHz). As shown in Fig. 5.3, the lower bandwidth edge $f_{edge\text{-}low}$ and the upper bandwidth edge $f_{edge\text{-}high}$ of the aggregated BS channel bandwidth are used as frequency reference points for transmitter and receiver requirements, which are defined as $f_{edge\text{-}low} = f_{C\text{-}low} - f_{offset-low}$ and

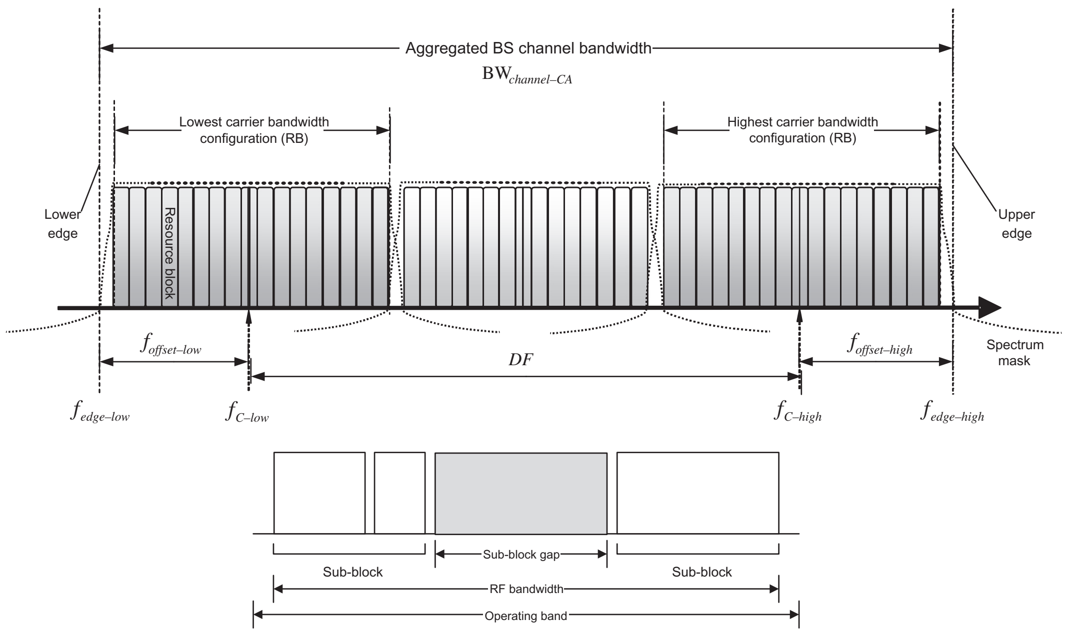

Figure 5.3
Illustration of aggregated BS channel bandwidth for intraband carrier aggregation [5].

$f_{edge\text{-}high} = f_{C\text{-}high} + f_{offset\text{-}high}$. The lower and upper frequency offsets depend on the transmission bandwidth configurations of the lowest and highest assigned edge component carrier and are defined as $f_{offset-low} = (12N_{RB-low} + 1)SCS_{low}/2 + W_{GB}$ (MHz) and $f_{offset\text{-}high} = (12N_{RB\text{-}high} - 1)\mathrm{SCS}_{high}/2 + W_{GB}$(MHz), where W_{GB} denotes the lower or upper minimum guard band defined for the lowest and highest assigned component carrier. In the latter equations, the parameters $N_{RB\text{-}low}$ and $N_{RB\text{-}high}$ are the transmission bandwidth configurations for the lowest and highest assigned component carrier, SCS_{low} and SCS_{high} denote the subcarrier spacing for the lowest and highest assigned component carrier, respectively [5]. It can be shown that the minimum guard band can be defined as $W_{GB} = \mathrm{BW}_{channel}/2 - \mathrm{SCS}(12N_{RB} + 1)/2$, where N_{RB} is the maximum number of resource blocks that can fit into the *BS channel bandwidth* and SCS denotes the subcarrier spacing. Note that an extra SCS/2 guard band is inserted on each side of the carrier due to the relation to the RF channel raster, which has a subcarrier-level granularity and is defined independent of the actual spectrum blocks. Therefore, it may not be possible to place an RF carrier exactly in the center of a spectrum block and an extra guard band would be required to ensure that the RF requirements can be met [10].

Some spectrum allocations may consist of fragmented blocks of spectrum. In the intra-band non-contiguous case, the BS transmits and receives over an RF bandwidth that is split in two (or more) separate subblocks with a subblock gap in between (see Fig. 5.3). A subblock is defined as one contiguous allocated block of spectrum for transmission and reception by the same BS. The lower subblock edge of the subblock bandwidth ($\mathrm{BW}_{channel\text{-}block}$) is defined as $f_{edge\text{-}block\text{-}low} = f_{C\text{-}block\text{-}low} - f_{offset\text{-}low}$. The upper subblock edge of the subblock bandwidth is defined as $f_{edge\text{-}block\text{-}high} = f_{C\text{-}block\text{-}high} + f_{offset\text{-}high}$. The subblock bandwidth $\mathrm{BW}_{channel\text{-}block}$ is defined as $\mathrm{BW}_{channel\text{-}block} = f_{edge\text{-}block\text{-}high} - f_{edge\text{-}block\text{-}low}$ (MHz). The lower and upper frequency offsets $f_{offset\text{-}block\text{-}low}$ and $f_{offset\text{-}block\text{-}high}$ depend on the transmission bandwidth configurations of the lowest and highest assigned edge component carriers within a subblock which are defined as $f_{offset\text{-}block\text{-}low} = (12N_{RB\text{-}low} + 1)\mathrm{SCS}_{low}/2 + W_{GB\text{-}low}$ (MHz) and $f_{offset\text{-}block\text{-}high} = (12N_{RB\text{-}high} - 1)\mathrm{SCS}_{high}/2 + W_{GB\text{-}high}$ (MHz), where $N_{RB\text{-}low}$ and $N_{RB\text{-}high}$ are the transmission bandwidth configurations for the lowest and highest assigned component carrier within a subblock, respectively. In the latter equations, SCS_{low} and SCS_{high} denote the subcarrier spacing for the lowest and highest assigned component carrier within a subblock, respectively; $W_{GB\text{-}low}$ and $W_{GB\text{-}high}$ are the minimum guard bands for the lowest and highest assigned component carriers, respectively. The subblock gap size between two consecutive subblocks W_{gap} is defined as $W_{gap} = f_{edge-block-low_{(n+1)}} - f_{edge-blockn-high_{(n)}}$ (MHz). Because the subblock gap starts from the inner edge of the channel bandwidth and not the center of the channel bandwidth, the subblock gap width is independent of the component carriers' channel bandwidth [5].

The frequency spacing between RF carriers depends on the deployment scenario, the size of the frequency block available and the *BS channel bandwidths*. The nominal channel spacing DF between two adjacent NR carriers is defined based on the frequency range and channel raster. For NR FR1 operating bands with 100 kHz channel raster, $DF = \left(\mathrm{BW}_{channel(1)} + \mathrm{BW}_{channel(2)}\right)/2$. For NR FR1 operating bands with 15 kHz channel raster, $DF = \left(\mathrm{BW}_{channel(1)} + \mathrm{BW}_{channel(2)}\right)/2 + \{-5, 0, 5\}$ kHz. For NR FR2 operating bands with 60 kHz channel raster, $DF = \left(\mathrm{BW}_{channel(1)} + \mathrm{BW}_{channel(2)}\right)/2 + \{-20, 0, 20\}$ kHz where $\mathrm{BW}_{channel(1)}$ and $\mathrm{BW}_{channel(2)}$ are the *BS channel bandwidths* of the two NR RF carriers. The channel spacing can be adjusted depending on the channel raster in order to optimize the system performance in a particular deployment scenario [5].

The channel spacing between adjacent component carriers for intra-band contiguously aggregated carriers is a multiple of least common multiple of channel raster and subcarrier spacing. The nominal channel spacing DF between two adjacent aggregated NR carriers is defined as follows [5]:

- For NR operating bands with 100 kHz channel raster

$$DF = \left\lfloor \frac{\left(\mathrm{BW}_{channel(1)} + \mathrm{BW}_{channel(2)} - 2\left|W_{GB_{channel(1)}} - W_{GB_{channel(2)}}\right|\right)}{0.6} \right\rfloor 0.3 \text{ (MHz)}$$

- For NR operating bands with 15 kHz channel raster

$$DF = \left\lfloor \frac{\mathrm{BW}_{channel(1)} + \mathrm{BW}_{channel(2)} - 2\left|W_{GB_{channel(1)}} - W_{GB_{channel(2)}}\right|}{0.015 \times 2^{n+1}} \right\rfloor 0.015 \times 2^{n} \text{ (MHz)}$$

  and $n = \max(m_1, m_2)$

- For NR operating bands with 60 kHz channel raster

$$DF = \left\lfloor \frac{\mathrm{BW}_{channel(1)} + \mathrm{BW}_{channel(2)} - 2\left|W_{GB_{channel(1)}} - W_{GB_{channel(2)}}\right|}{0.06 \times 2^{n+1}} \right\rfloor 0.06 \times 2^{n} \text{ (MHz)}$$

  and $n = \max\left(\mu_1, \mu_2\right) - 2$

In the above expressions, $\mathrm{BW}_{channel(1)}$ and $\mathrm{BW}_{channel(2)}$ represent the *BS channel bandwidths* of the two NR component carriers with values in MHz; $W_{GB_{channel(i)}}$ is the minimum guard band of the ith channel; μ_1 and μ_2 denote the subcarrier spacing configurations of the component carriers. The channel spacing for intra-band contiguous carrier aggregation can be adjusted to any multiple of least common multiple of the channel raster and SCS less than the nominal channel spacing to optimize performance in a particular deployment scenario. For intra-band non-contiguous carrier aggregation, the channel spacing between two NR component carriers in different subblocks must be larger than the nominal channel spacing [5].

5.2 Base Station Transceiver RF Characteristics and Requirements

5.2.1 General Base Station RF Requirements

The RF requirements in general are defined at the BS antenna connector. Those requirements are referred to as conducted requirements and are typically specified as absolute or relative power levels measured at the antenna connector. The OOB emission limits are often defined as conducted requirements. In active antenna systems, the RF requirements are defined as radiated requirement, which are measured over the air in the far-field of the antennas; thus they include the antenna patterns and directivity effects [1]. In OTA test procedures, the spatial characteristics of the BS including the antenna system are evaluated. In base stations equipped with AAS, where the active parts of the transceiver are integrated with the antenna system, it is not always possible to conduct the measurements at the antenna connector. For this reason, 3GPP Rel-13 specified the RF requirements for the AAS base stations in a set of separate RF specifications that are applicable to LTE and the previous generations. The radiated RF requirements and OTA testing are being specified for the new radio for the FR1 and FR2 bands, which have borrowed a large portion of the previously developed AAS specifications. Note that the term AAS is not used within the NR base station RF specifications, rather the requirements are specified for different BS types.

The AAS-type BS requirements are based on the generalized AAS BS radio architecture shown in Fig. 5.4 [5]. The architecture consists of a transceiver unit array that is connected to a composite antenna structure that contains a radio distribution network and an antenna array. The transceiver unit array comprises a number of transmitter and receiver units, which are connected to the composite antenna structure via a number of connectors on the transceiver array boundary (TAB). The TAB connectors correspond to the antenna connectors on a non-AAS BS and serve as a reference point for the conducted requirements. The radio distribution network is a passive unit which distributes or aggregates the transmitter outputs or the receiver inputs to the corresponding antenna elements, respectively. It must be noted that the actual implementation of an AAS BS may be different in terms of the physical location of different parts, array geometry, type of antenna elements, etc. Based on the architecture shown in Fig. 5.4, two types of requirements can be defined. Conducted requirements are defined for each RF characteristic at an individual or a group of TAB connectors. The conducted requirements are defined such that they are equivalent to the corresponding conducted requirement of a non-AAS BS, which implies that the performance of the system or the impact on other systems is expected to be the same. Radiated requirements, on the other hand, are defined based on OTA measurements conducted in the far-field of the antenna system. Since the spatial direction becomes relevant in this case, it is detailed for each requirement how it applies. The radiated requirements are defined with reference to a radiated interface boundary (RIB) in the far-field region of the antenna array [5,10].

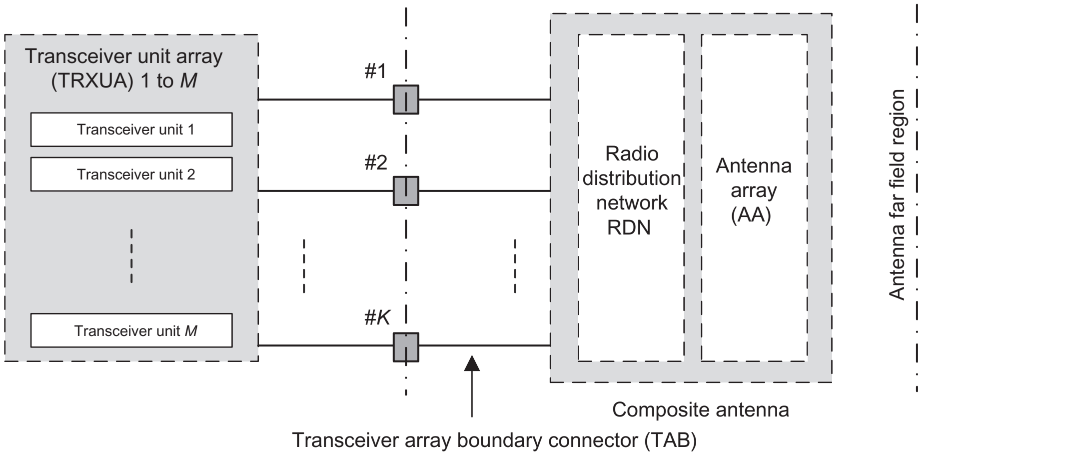

Figure 5.4
Radiated and conducted reference points for BS type 1-H [5]. *BS*, Base station.

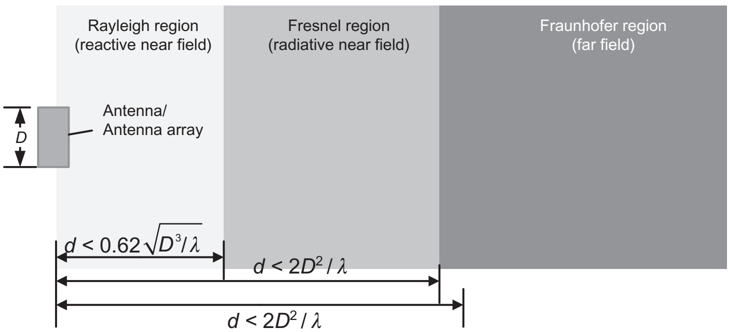

Figure 5.5
Illustration of the radiation regions of an antenna element/array.

In order to determine an appropriate point in the antenna array far-field region for conducting the OTA measurements, one must understand the radiation characteristics of the antenna systems and be able to define the minimum distance $d_{\min}$ in the far-field radiation region of an antenna array. In the theory of electromagnetics and antennas, the radiation regions surrounding an antenna element/array can be divided into three regions as follows (see Fig. 5.5):

- *Reactive near-field region*: In the immediate vicinity of the antenna, we have the reactive near-field. In this region, the fields are predominately reactive fields, which means the electric (E) and the magnetic (H) fields are out of phase by 90 degrees relative to each other. Note that for propagating or radiating fields, the fields are orthogonal and are in-phase. The boundary of this region is typically given as $d < 0.62\sqrt{D^3/\lambda}$, where

D is the maximum linear dimension of an antenna element/array, λ denotes the wavelength, and d is the distance from the antenna(s).

- *Radiating near-field (Fresnel) region*: The radiating near-field or Fresnel region is the region between the near- and far-fields. In this region, the reactive fields are not dominant, and the radiating fields begin to emerge. However, unlike the far-field region, the shape of the radiation pattern may vary noticeably with distance. This region is often identified by $0.62\sqrt{D^3/\lambda} < d < 2D^2/\lambda$. Note that depending on the values of d and the wavelength, this field may or may not exist.
- *Far-field (Fraunhofer) region*: The far-field is a region relatively far from the antenna element/array, where the shape of radiation pattern does not vary with distance, although the fields are attenuated proportional to d^{-1}, the power density degrades by d^{-2}. This region is dominated by radiated fields, with the E- and H-fields in a plane orthogonal to each other and to the direction of propagation which is the main characteristic of plane waves. If the maximum linear dimension of an antenna is D, the following three conditions must all be satisfied so that a point can be in the far-field region: $d > 2D^2/\lambda$, $d \gg D$ and $d \gg \lambda$. The first two inequalities ensure that the power radiated in a given direction from distinct parts of the antenna is approximately parallel. This helps ensure that the fields in the far-field region can be characterized as plane waves. Near a radiating antenna, there are reactive fields that typically have the E-fields and H-fields diminish with distance as d^{-2} or d^{-3}. The third inequality ensures that these near-fields are disappeared, and we are left with the radiating fields, which diminish with distance as d^{-1}.

The NR base stations can be categorized into three main configurations from RF requirements perspective as follows [5]:

- *BS type 1-C*: An NR BS operating in FR1 which is connected to the antennas via coaxial cables. The RF requirements for this BS type only consist of conducted requirements defined at individual antenna connectors. For BS type 1-C, the requirements are applied at the BS antenna connector (port A). If any external apparatus such as an amplifier, a filter, or the combination of such devices is used, the RF requirements apply at the far end antenna connector (port B).
- *BS type 1-H*: An NR BS operating in FR1 consisting of an integrated AAS with RF transceiver connected to antennas using a TAB connector. The requirement set for this BS type consists of conducted requirements defined at individual TAB connectors and OTA requirements defined at the RIB. For BS Type 1-H, the requirements are defined for two points of reference, signified by radiated requirements and conducted requirements.
- *BS type 1-O*: A connector-less AAS-type NR BS operating in FR1 whose RF requirement set only consist of OTA requirements defined at the RIB.
- *BS type 2-O*: A connector-less AAS-type NR BS operating in FR2 whose RF requirement set only consists of OTA requirements defined at the RIB.

For BS Type 1-O and BS Type 2-O, the radiated characteristics are defined over the air, where the operating band–specific radiated interface is referred to as the RIB. The radiated requirements are also referred to as OTA requirements. The (spatial) characteristics in which the OTA requirements apply are detailed for each requirement. Fig. 5.6 illustrates the NR BS configurations and the reference points for conducting connected or OTA measurements. Note that one of the RF configurations for FR1 does not require connectors between the RF transceivers and antennas according to BS Type 1-O; thus smaller equipment size and improved power efficiency can be expected relative to the LTE transceivers with AAS specified in 3GPP Rel-13. While the operation in FR2 has the advantage of wideband transmission in high-frequency bands in terms of RF configuration, the higher frequencies result in larger power losses at the connectors and cables; increased path loss and reduced coverage due to lower power density over wider channels. Therefore, higher antenna gains are necessary to compensate for the losses and to maintain a certain link budget and coverage. Since it would be more difficult to design and implement RF signal transceivers and antennas with high density in FR2, if conventional RF configuration with connectors is used, BS type 2-O RF configuration without connectors is defined for FR2 operation. The BS Type 2-O would allow implementation of beamforming over wide channel bandwidths to maintain coverage and to achieve high spectral efficiencies [29].

The RF performance requirements for BS types 1-C and 1-H are based on LTE-advanced specifications [11], when NR radio parameters are applied. However, BS types 1-O and 2-O have integrated radio transceivers and antennas with no connectors to conduct measurements; thus OTA specifications have been extended such that in the overall RF performance specifications, a reference point in the radiated space, referred to RIB, can be defined. The output power for various BS classes and types is shown in Table 5.3. It must be noted that the factor $10\log(N_{TXU-counted})$ is used to derive the rated carrier output power from output power per TAB connector for BS Type 1-H and no upper limits for output power has been specified for BS Type 2-O in 3GPP Rel-15.

In addition to the equivalent isotropic radiated power (EIRP) and the equivalent isotropic sensitivity (EIS) including the antenna characteristics in the beam direction, which were specified in 3GPP Rel-13 LTE, the total radiated power (TRP) is introduced as a new metric in NR RF specifications (Fig. 5.7). The TRP definition makes it possible to specify OTA requirements for power-related RF performance requirements such as the BS output power and spurious emissions. Fig. 5.7 illustrates the visualization of the EIRP, EIS, and TRP definitions.

The main BS RF performance specifications of the LTE and NR in FR1 and FR2 have been summarized and compared in Table 5.4. The RF specifications in FR1 are based on LTE specifications with the maximum channel bandwidth of 100 MHz. In FR2, the NR radio specifications support wider bands, lower latency, and faster response with maximum

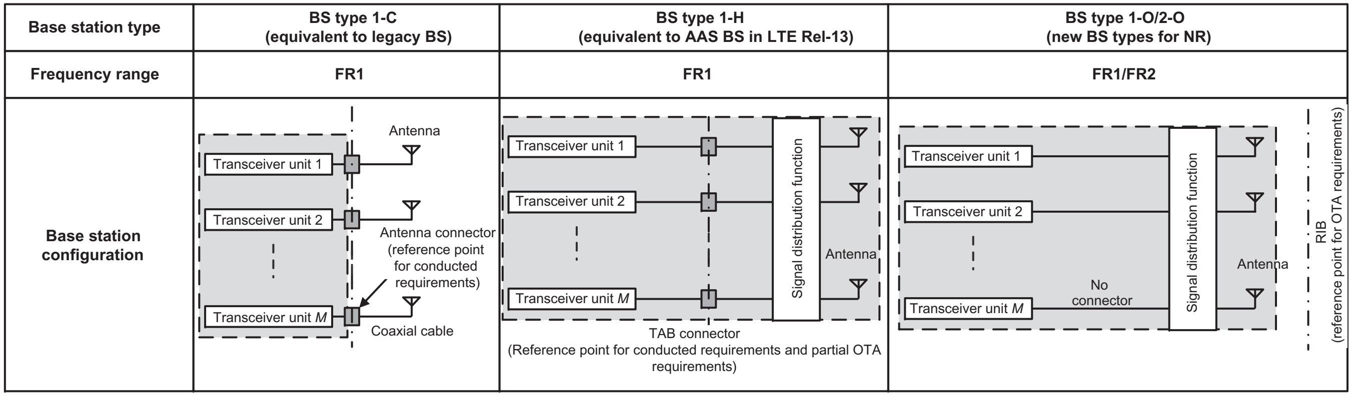

Figure 5.6
NR base station configurations [5,29].

Table 5.3: Output power per base station class and type (dBm) [5,14].

	BS Type 1-C	BS Type 1-H		BS Type 1-O
	The Rated Carrier Output Power per Antenna Connector	The Sum of Rated Carrier Output Power for All TAB Connectors for a Single Carrier	The Rated Carrier Output Power per TAB Connector	Rated Carrier TRP Output Power Declared per RIB
Wide area BS	No upper limit for wide area base station			
Medium range BS	≤ 38	$\leq 38 + 10\log(N_{TXU\text{-}counted})$	≤ 38	≤ 47
Local area BS	≤ 24	$\leq 24 + 10\log(N_{TXU\text{-}counted})$	≤ 24	≤ 33

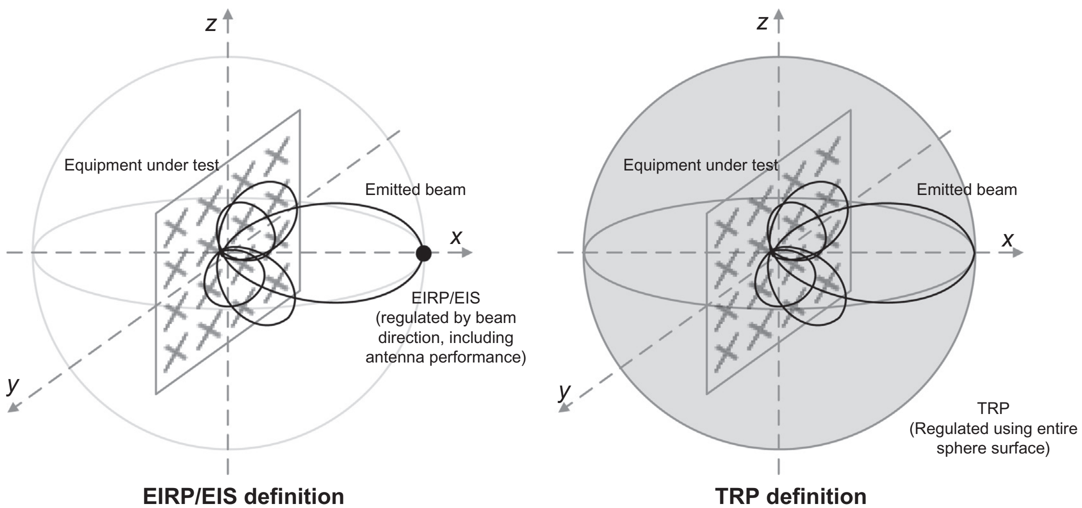

Figure 5.7
Illustration of the RF performance requirements for the NR base station and mobile station [29].

Table 5.4: Comparison of main base station RF performance specifications [5,29].

	LTE	NR FR1 BS Type 1-O	NR FR2 BS Type 2-O
Maximum channel bandwidth (MHz)	20	100	400
Transmitter transient period (μs)	< 17	< 10	< 3
ACLR (dB)	45	45	28
NF (dB)	5	5	10
Transmit power deviation (dB)	± 2.0	± 2.2 (EIRP accuracy) ± 2.0 (TRP accuracy)	± 3.4 (EIRP accuracy) ± 3.0 (TRP accuracy)

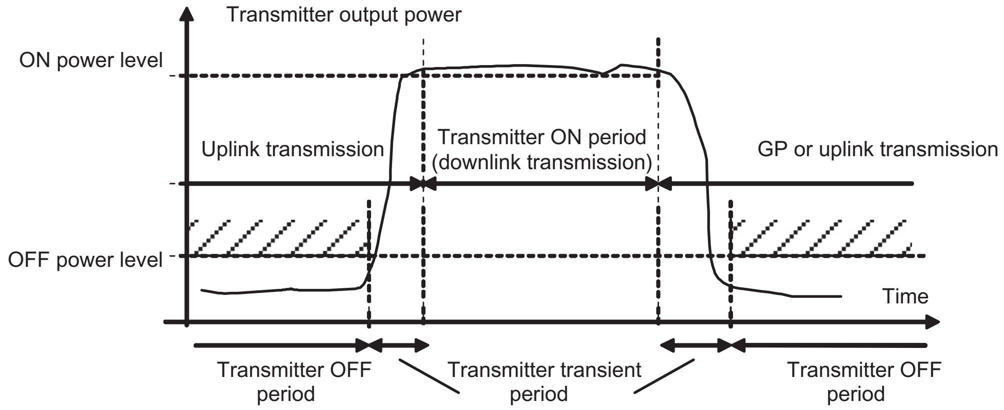

Figure 5.8
Illustration of transmitter transient period [5].

channel bandwidth of up to 400 MHz and transmitter transient period[3] of 3 μs, which reflects the ON/OFF switching time for a TDD system (see Fig. 5.8). The specifications further take into account the efficiency/performance degradation of gNB RF components in the mmWave bands, which are compensated by beamforming and high-gain antenna arrays, and by relaxing certain radio requirements such as adjacent channel leakage ratio (ACLR) in the transmitter side and noise figure (NF) in the receiver side. The transmit power fluctuation is controlled by TRP accuracy for per-carrier total power deviation and EIRP accuracy, accounting for the antenna performance.

5.2.2 Conducted and Radiated Base Station Transceiver Characteristics

The radiated RF requirements for the NR devices and base stations are primarily derived from the corresponding conducted RF requirements. Unlike the conducted requirements, the radiated requirements are inclusive of the antennas, thus the emission levels such as BS output power and unwanted emissions are defined either by incorporating the antenna gain as a directional requirement using an effective isotropic radiated power (EIRP) or by defining limits using TRP. The directional requirement is a requirement which is applied to a specific direction within the OTA coverage range for the transmitter and when the AoA of the incident wave of a received signal is within the *OTA reference sensitivity range of angles of arrival (REFSENS RoAoA)* (this is determined by the contour defined by the points at which the achieved EIS is 3 dB higher than the achieved EIS in the reference direction) or the

[3] The transmitter transient period is the time interval during which the transmitter is changing from the transmitter OFF period to the transmitter ON period or vice versa in a TDD base station [5].

minSENS RoAoA [the reference Range of Angles of Arrival (RoAoA) associated with the OTA sensitivity direction declaration (OSDD) with the lowest declared EIS] as appropriate for the receiver. The radiated transmit power is defined as the EIRP level for a declared beam at a specific beam peak direction. For each beam the requirement is based on declaration of a beam identity, reference beam direction pair, beam-width, rated beam EIRP, OTA peak directions set, the beam direction pairs at the maximum steering directions, and their associated rated beam EIRP and beam-width(s). For a declared beam and beam direction pair the rated beam EIRP level is the maximum power that the BS is declared to radiate at the associated beam peak direction during the transmitter ON period. For each beam peak direction associated with a beam direction pair within the OTA peak direction set, a specific rated beam EIRP level may be claimed. Any claimed value must be met within the accuracy requirements. The rated beam EIRP is only required to be declared for the beam direction pairs [5].

The TRP and EIRP metrics are related through the number of radiating antennas and are implementation specific, considering the geometry of the antenna array and the correlation between unwanted emission signals from different antenna ports. An EIRP limit results in different levels of total radiated unwanted emission power depending on the implementation. The EIRP limit per se cannot control the total amount of interference in the network, whereas a TRP requirement limits the total amount of interference in the network regardless of the specific BS implementation. In the case of passive systems, the antenna gain does not noticeably vary between the desired signal and the unwanted emissions. Thus EIRP is directly proportional to TRP and can be used interchangeably. For an active system such as NR, the EIRP may vary considerably between the desired signal and the unwanted emissions and also between implementations, thus EIRP (in this case) is not proportional to TRP and should not substitute TRP [10].

The RF requirements for BS type 1-O only consist of radiated (OTA) requirements, which are in general based on the corresponding conducted requirements with the exception of two additional radiated requirements; that is, *radiated transmit power* and *OTA sensitivity*. The BS type 1-H is defined with a set of requirements that are mainly related to conducted requirements and the additional radiated requirements, which are the same for BS type 1-O. The radiated transmit power is defined by considering the antenna array beamforming pattern in a specific direction as EIRP for each beam that the BS has declared to transmit, which in some sense relies on the accuracy of the declared EIRP level.

The OTA sensitivity requirement is based on declaration of one or more OSDDs,[4] related to an AAS BS receiver. The AAS BS receiver may optionally be capable of redirecting/changing the receiver target by means of adjusting BS settings, resulting in multiple

4 OTA sensitivity directions declaration is a set of manufacturer declarations comprising one or more equivalent isotropic sensitivity (EIS) values (with related RAT and channel bandwidth), and the directions where they apply.

sensitivity RoAoA. The sensitivity RoAoA resulting from the current AAS BS settings is referred to as the active sensitivity RoAoA. This definition reflects the antenna array beamforming pattern in a specific direction as declared by EIS[5] level toward a receiver target. The EIS limit must be met not only in a single direction but also within a Range of Angle of Arrival (RoAoA) in the direction of the receiver target. Depending on the level of adaptivity for the AAS BS, the following two alternative declarations can be considered. If the receiver is direction adaptable, in the sense that the receiver target can be redirected, then the declaration contains a receiver target redirection range in a specified receiver target direction. The EIS limit should be met within the redirection range, which is evaluated at five declared sensitivity RoAoA within that range. If the receiver is not adaptive to direction and thus one cannot redirect the receiver target, the declaration consists of a single sensitivity RoAoA in a specified receiver target direction, in which the EIS limit should be met. Note that the OTA sensitivity is defined in addition to the reference sensitivity requirement, which exists both as conducted for BS type 1-H and radiated for BS type 1-O [1,10].

The RF requirements for BS type 2-O are radiated requirements for base stations in FR2. These requirements are identical to the radiated RF requirements defined for FR1 bands, except that the limits for many requirements are different. As for the device, the difference in coexistence at mmWave frequencies leads to more relaxed requirements on ACLR and adjacent channel selectivity (ACS). The implementation of mmWave technologies is more challenging compared to the more mature technologies in sub-6 GHz frequency bands. In the following, we provide an overview of the radiated RF requirements in FR2 bands [5,29]:

- *Output power level*: The maximum output power is the same for FR1 and FR2; however, it is scaled from the conducted requirement and expressed as TRP. In addition, there is a directional radiated transmit power requirement. The dynamic range requirement is defined similar to that of FR1.
- *Transmitted signal quality*: Frequency error, error vector magnitude (EVM), and time-alignment requirements are defined similar to those in FR1 and with the same limits.
- *Radiated unwanted emissions requirements*: Occupied bandwidth, spectrum mask, ACLR, and spurious emissions are defined similar to those for FR1. The latter features/metrics are based on TRP and have less stringent limits relative to those in FR1. The ACLR is on the order of 15 dB and relaxed compared to FR1 due to more favorable coexistence conditions.
- *Reference sensitivity and dynamic range*: These are defined similar to those in FR1; however, the levels are not comparable. There is in addition a directional OTA sensitivity requirement.

[5] Equivalent isotropic sensitivity is the sensitivity for an isotropic directivity device equivalent to the sensitivity of the discussed device exposed to an incoming wave from a defined AoA.

- *Receiver susceptibility to interfering signals*: ACS, in-band, and OOB blocking are defined in the same way as it was done for FR1; however, no narrowband blocking scenario is defined since there are only wideband systems in FR2. The ACS requirement is relaxed compared to FR1 due to more favorable coexistence condition.

The RF characteristics of the receiver and transmitter of a BS or a device are defined by the RF requirements that are partly specified by 3GPP. The BS is the RAN node that transmits and receives RF signals on one or more antenna connectors. An NR BS, defined in RF specifications [5], is not the same as a gNB, which is the corresponding logical node in the radio access network. However, the device in RF specifications is the same as the UE defined in RAN specifications [2,3,4]. The conducted RF requirements are defined for operating frequencies in FR1, whereas only radiated (OTA) requirements are defined for FR2 bands. The set of conducted RF requirements defined for NR are essentially the same as those defined for LTE. Some of the requirements related to the frequency bands and/or the system deployment are technology-agnostic and are derived from regional regulatory requirements. The NR-specific requirements related to flexible channel bandwidths and multiple OFDM numerologies have certain implications on the transmitter requirements for unwanted emissions, where the definition of the limits depends on the channel bandwidth. Such limits would be more difficult to define for a system where the BS may operate with multiple channel bandwidths and where the device may change its channel bandwidth during operation. The properties of the flexible OFDM-based physical layer have implications on the way that the transmitter modulation quality and the receiver selectivity and blocking requirements are defined. Note that the channel bandwidth in general is different for the BS and the UE. The type of transmitter requirements defined for the device is very similar to those defined for the BS. However, the output power levels are considerably lower for a device, while the restrictions on the device implementation are much higher due to considerations for complexity and cost of implementation.

The RF requirements are grouped into transmitter and receiver requirements. In addition, there are performance characteristics for base stations and devices that define the receiver baseband minimum performance for all physical channels under different propagation conditions. Each RF requirement has a corresponding test defined in the NR test specifications for the BS and the device [2−5]. These specifications define the test setup, procedure, signals, tolerances, etc., which are required to demonstrate the compliance with the RF and performance requirements.

The transmitter characteristics define the RF requirements for the desired device/BS transmit signal as well as the unwanted signal emissions outside the transmitted signal bandwidth. These requirements are specified in three categories: (1) output power level requirements set limits on the maximum permissible transmit power, the dynamic variation of the power level, and the transmitter ON/OFF state; (2) transmit signal quality requirements define the authenticity of the transmitted signal and the relation between multiple transmitter branches; and (3) unwanted emissions requirements set limits on the emissions outside the transmit signal bandwidth which are associated with the regulatory requirements and coexistence (spectrum sharing) with other systems. A summary of the device/BS transmitter and receiver characteristics is provided in Tables 5.5 and 5.6, respectively.

The set of receiver requirements for NR are similar to those defined for other systems such as LTE. The receiver characteristics can be categorized into three groups: (1) sensitivity and dynamic range requirements for receiving the desired signals; (2) receiver susceptibility to interfering signals defines receiver susceptibility to different types of interfering signals at different frequency offsets; and (3) unwanted emissions limits are also defined for the receiver.

Table 5.5: Summary of conducted NR transmitter characteristics [2–5,10].

Requirement Category	Base Station Requirements	UE Requirements
Output power level	Maximum output power	Transmit power
	Output power dynamics	Output power dynamics
	ON/OFF power (TDD)	Power control
Transmitted signal quality	Frequency error	Frequency error
	EVM	Transmit modulation quality
	Time alignment between transmitter branches	In-band emissions
Unwanted emissions	Operating band unwanted emissions	Spectrum emission mask
	ACLR and CACLR	ACLR and CACLR
	Spurious emissions	Spurious emissions
	Occupied bandwidth	Occupied bandwidth
	Transmitter intermodulation	Transmitter intermodulation

ACLR, Adjacent channel leakage ratio; EVM, error vector magnitude.

Table 5.6: Summary of conducted NR receiver characteristics [2–5,10].

Requirement Category	Base Station Requirements	UE Requirements
Sensitivity and dynamic range	Reference sensitivity	Reference sensitivity
	Dynamic range	Power level
	In-channel selectivity	Maximum input level
Receiver Susceptibility to Interfering Signals	Out-of-band blocking	Out-of-band blocking
	In-band blocking	Spurious response
	Narrowband blocking	In-band blocking
	Adjacent channel selectivity	Narrowband blocking
	Receiver intermodulation	Adjacent channel selectivity
		Receiver intermodulation
Unwanted emissions from the receiver	Receiver spurious emissions	Receiver spurious emissions

The channel bandwidths supported by a UE are determined by the NR operating bands as well as the transmitter and receiver RF requirements for those frequency bands. This is due to fact that the combination of maximum power and a large number of transmitted and/or received resource blocks can make it difficult to meet some of the RF requirements. The concept of network signaling (NS) of RF requirements has been used in LTE and continued to be used in NR, where a device can be informed during call setup of whether some specific RF requirements apply when the device is connected to a network. Additional RF requirements apply to the UE when a specific NS value (NS_*m*) is signaled to the device as part of the handover procedure or via a broadcast message. These requirements are associated with limitations and variations of RF parameters such as device output power, maximum channel bandwidth, and the number of transmitted resource blocks. The variations of the requirements are defined along with the NS_*m* in the UE RF specifications [2], where each value corresponds to a specific condition. The default value for all bands is NS_01. The NS_*m* values are related to the permissible power reduction referred to as additional maximum power reduction (A-MPR) and may further apply to transmission using the minimum number of resource blocks, depending on the channel bandwidth [10]. More specifically, each additional emission requirement is associated with a unique NS value indicated via RRC signaling by an NR frequency band number and an associated value in the field *additionalSpectrumEmission.* In NR radio specifications, the notion of indication or signaling of an NS value refers to the corresponding indication of an NR frequency band number (i.e., the IE field *freqBandIndicatorNR* and an associated value of *additionalSpectrumEmission* in the relevant RRC information elements) [2].

In order to accommodate different deployment scenarios for base stations, there are multiple sets of RF requirements for NR base stations, each applicable to a BS class. The NR BS classes include macro-cell, micro-cell, and pico-cell deployment scenarios; however, 3GPP does not identify the BS classes based on the latter terms, rather it refers to them as wide area, medium range, and local area base stations. The wide area BS class is a type of BS for macro-cell scenarios, with minimum distance of 35 m between the BS and the UE. This is the typical large cell deployment with high-tower or above-rooftop installations, providing wide area outdoor or indoor coverage. Medium range BS is a type of BS for micro-cell scenarios, with the minimum distance of 5 m between the BS and the UE. Typical deployments are outdoor below-rooftop installations, providing outdoor hotspot coverage and outdoor-to-indoor coverage through walls. Local area BS is a type of BS that is intended for pico-cell scenarios, with the minimum distance of 2 m between the BS and the UE. Typical deployments are indoor offices and indoor/outdoor hotspots, with the BS mounted on walls or ceilings. The associated deployment scenarios for each BS class are exactly the same for the BS with and without connectors. More specifically, for BS type 1-O and 2-O, BS classes are defined as *wide area* base stations characterized by requirements derived from macro-cell scenarios with a BS to UE minimum ground distance of 35 m; *medium range* base stations characterized by requirements derived from micro-cell scenarios with a BS to UE minimum ground distance of 5 m; and *local area* base stations are characterized by requirements derived from pico-cell scenarios with a BS to UE minimum ground distance of 2 m.

For BS types 1-C and 1-H, the BS classes are defined as *wide area* base stations characterized by requirements derived from macro-cell scenarios with a BS to UE minimum coupling loss of 70 dB; *medium range* base stations are characterized by requirements derived from micro-cell scenarios with a BS to UE minimum coupling loss of 53 dB; and *local area* base stations are characterized by requirements derived from pico-cell scenarios with a BS to minimum coupling loss of 45 dB [5].

The local area and medium range BS classes have different requirements compared to wide area base stations, mainly due to shorter minimum BS to device distance and lower minimum coupling loss. The maximum BS output power is limited to 38 dBm for medium range base stations and 24 dBm for local area base stations. This power is defined per antenna and per carrier. There is no maximum BS output power defined for wide area base stations. The spectrum mask, that is, the limits on the operating band unwanted emissions (OBUE), has lower limits for medium range and local area, consistent with the lower maximum output power levels. The receiver reference sensitivity limits are higher and more relaxed for the medium range and local area base stations. The receiver dynamic range and in-channel selectivity (ICS) are also adjusted accordingly for these BS classes. The limits on co-location for medium range and local area base stations are relatively more relaxed compared to the wide area base stations. The co-location refers to the additional blocking requirement which may be applied for the protection of NR BS receivers when another GSM, CDMA, UTRA, E-UTRA, or NR BS operating in a different frequency band is co-located with an NR BS. The requirement is applicable to all channel bandwidths supported by the NR BS. These requirements assume a 30 dB coupling loss between interfering transmitter and NR BS receiver and are based on co-location with base stations of the same class. All medium range and local area BS limits for receiver susceptibility to the interfering signals are adjusted considering the higher receiver sensitivity limit and the lower minimum coupling loss (BS to device) [5].

The requirements for base stations that are capable of multi-band operation are defined at the multi-band connector or multi-band RIB, where the RF requirements apply separately to each supported operating band. For BS types 1-C and 1-H that are capable of multi-band operation, various structures in the form of combinations of different transmitter and receiver implementations (multi-band or single-band) with mapping of transceivers to one or more antenna connectors for BS type 1-C or TAB connectors for BS type 1-H in different ways are possible [5].

The conducted transmitter characteristics are specified at the antenna connector for BS type 1-C and at the TAB connector for BS type 1-H. For BS type 1-H, the minimum number of supported geographical cells (i.e., geographical areas covered by beams) must be identified.

The minimum number of supported geographical cells (N_{cells}) corresponds to the BS setting with the minimum amount of cell splitting when transmitting on all TAB connectors for the operating band, or with the minimum number of transmitted beams. Each TAB connector of the BS Type 1-H supporting transmission in an operating band is mapped to one *TAB connector TX min cell group*. The *TAB connector TX min cell group* is an operating band–specific declared group of TAB connectors to which BS type 1-H conducted transit requirements are applied. In the latter definition, the group corresponds to the group of TAB connectors which are responsible for transmitting a cell when the BS type 1-H setting corresponding to the declared minimum number of cells with transmission on all TAB connectors supporting an operating band, but its existence is not limited to that condition. Note that the mapping of the TAB connectors to cells/beams is implementation dependent. The number of active transmitter units that are considered when calculating the conducted transmitter emissions limits ($N_{TXU\text{-}counted}$) for BS type 1-H is given as $N_{TXU\text{-}counted} = \min(N_{TXU\text{-}active}, 8N_{cells})$ where $N_{TXU\text{-}countedpercell}$ is used for scaling of basic limits and is calculated as $N_{TXU\text{-}countedpercell} = N_{TXU\text{-}counted}/N_{cells}$. The parameter $N_{TXU\text{-}active}$ depends on the actual number of active transmitter units and is independent of N_{cells} [5].

Some NR features require the BS to transmit from two or more antennas, for example, transmit diversity and multi-antenna transmission modes. For carrier aggregation, the carriers may also be transmitted from different antennas. In order for the device to properly receive the signals from multiple antennas, the timing relationship between any two transmitter branches is specified in terms of the maximum time-alignment error between transmitter branches. The maximum allowed error depends on the feature or combination of features in the transmitter branches.

In general, there is no maximum output power requirement for wide area base stations. As mentioned earlier, there is the maximum output power limit of 38 dBm for the medium range and 24 dBm for the local area base stations. In addition, 3GPP RF specifications specify a tolerance, which defines the extent to which the actual maximum power may deviate from the power level declared by the BS manufacturer. Power control is used to limit the interference level. A total power control dynamic range per resource element is specified for a BS, which is the difference between the power of a resource element and the average resource element power for a BS at maximum output power ($P_{max,c,TABC}$) for a specified reference condition (Table 5.7). There is also a dynamic range requirement for the total BS power. The BS total power dynamic range is the difference between the maximum and the minimum transmit power of an OFDM symbol for a specified reference condition (Table 5.8). For TDD operation, a power mask is defined for the BS output power, defining the OFF power level during the uplink subframes and the maximum time for the transmitter transition time between the transmitter ON and OFF states [5].

Transmit OFF power requirements only apply to TDD mode of operation of an NR BS. The transmitter OFF power is defined as the mean power measured over $70/N(\mu s)$ filtered with

Table 5.7: Resource element power control dynamic range [5].

Modulation Scheme Used on the Resource Element	Resource Element Power Control Dynamic Range (dB)	
	Down	Up
QPSK (PDCCH)	−6	+4
QPSK (PDSCH)	−6	+3
16QAM (PDSCH)	−3	+3
64QAM (PDSCH)	0	0
256QAM (PDSCH)	0	0

Table 5.8: Total power dynamic range [5].

BS Channel Bandwidth (MHz)	Total Power Dynamic Range (dB)		
	SCS 15 kHz	SCS 30 kHz	SCS 60 kHz
5	13.9	10.4	N/A
10	17.1	13.8	10.4
15	18.9	15.7	12.5
20	20.2	17	13.8
25	21.2	18.1	14.9
30	22	18.9	15.7
40	23.3	20.2	17
50	24.3	21.2	18.1
60	N/A	22	18.9
70	N/A	22.7	19.6
80	N/A	23.3	20.2
90	N/A	23.8	20.8
100	N/A	24.3	21.3

a square filter of bandwidth equal to the transmission bandwidth configuration of the BS (BW_{config}) centered around the assigned channel frequency during the *transmitter OFF period*. The parameter $N = SCS/15$, where SCS is the subcarrier spacing in kHz. For multi-band connectors and for single-band connectors supporting transmission in multiple operating bands, the requirement is only applicable during the transmitter OFF period in all supported operating bands. For BSs supporting intraband contiguous carrier aggregation, the transmitter OFF power is defined as the mean power measured over $70/N$ (μs) filtered with a square filter of bandwidth equal to the *aggregated BS channel bandwidth* $BW_{channel\_CA}$ centered around $\left(f_{edge\text{-}high} + f_{edge\text{-}low}\right)/2$ during the *transmitter OFF period*. In this case, parameter $N = SCS/15$, where SCS is the smallest supported subcarrier spacing in kHz in the *aggregated BS channel bandwidth* [5].

The requirements for transmitted signal quality specify the extent to which the transmitted signal from the BS or the device deviates from a reference signal. The transmitted signal impairments are caused by the transmitter RF components and particularly the nonlinear response of the power amplifier (PA). The signal quality of the BS or the device can be

Table 5.9: Error vector magnitude requirements for base station type 1-C and BS type 1-H [5].

Modulation Scheme for PDSCH	Required EVM (%)
QPSK	17.5
16QAM	12.5
64QAM	8
256QAM	3.5

assessed based on the requirements on EVM and frequency error as well as the device in-band emissions. The EVM is a measure of the error in the modulated signal constellation, calculated as the RMS of the error vectors over the active subcarriers, considering all symbols of the modulation scheme under consideration. It is expressed as a percentage value in relation to the power of the ideal signal (Table 5.9). In other words, EVM defines the maximum SINR that can be achieved at the receiver, if there are no additional impairments to the signal between the transmitter and the receiver. Since a receiver can remove some impairments of the transmitted signal such as time dispersion, the EVM is assessed after cyclic prefix removal and the equalization modules. In this way, the EVM evaluation includes a standardized model of the receiver. The frequency offset resulting from the EVM evaluation is averaged and used as a measure of the frequency error of the transmitted signal. Frequency error is the measure of the difference between the actual BS transmit frequency and the assigned frequency, wherein the same source is used for RF frequency and data clock generation.

The EVM is a measure of the difference between the ideal symbols and the measured symbols after the equalization. The EVM is averaged over all allocated downlink resource blocks with the modulation scheme under consideration in the frequency domain, and a minimum of 10 downlink subframes. Although the basic unit ofmeasurement is one subframe, the equalizer is calculated over 10 subframe measurement periods to reduce the impact of noise in the reference symbols. The boundaries of the 10 subframe measurement periods do not need to be aligned with radio frame boundaries. In FDD mode, the averaging in the time domain is performed over 10 subframes (the same frame) of the measurement period from the equalizer estimation step, whereas in TDD mode, the averaging in the time domain can be calculated from subframes of different frames and has a minimum of 10 subframes averaging length. The TDD frame special fields such as guard period are not included in the averaging window. This can be mathematically expressed as follows [5]:

$$\overline{\mathrm{EVM}_{frame}} = \left(\frac{\sum_{k=1}^{N_{dl}} \sum_{l=1}^{N_k} \mathrm{EVM}_{kl}^2}{\sum_{k=1}^{N_{dl}} N_k} \right)^{1/2}$$

where N_k is the number of resource blocks with the considered modulation scheme in the kth subframe and N_{dl} is the number of allocated downlink subframes in one frame. The EVM measurements start on the third symbol of a slot, when the first symbol of that slot is a downlink symbol.

The unwanted emissions consist of OOB emissions and spurious emissions according to ITU-R definitions. In ITU-R definition, OOB emissions are unwanted emissions immediately outside the *BS channel bandwidth* resulting from the modulation process and nonlinearity in the transmitter analog components but excluding spurious emissions. Spurious emissions are emissions which are caused by unwanted transmitter effects, such as harmonics, parasitic, intermodulation products, and frequency conversion products, but exclude OOB emissions [5].

The ITU-R defines the boundary between the OOB and spurious emission domains as a frequency separation relative to the center of the carrier by 2.5× (or 250%) of the NR channel bandwidth. Any emission outside the channel bandwidth which occurs in the frequency range separated from the assigned frequency of the emission by less than 2.5 times the channel bandwidth of the emission will generally be considered an emission in the OOB domain. However, this frequency separation may be dependent on the type of modulation, the maximum symbol rate in the case of digital modulation, the type of transmitter, and frequency coordination factors. For example, in the case of some digital, broadband, or pulse-modulated systems, the frequency separation may need to differ from the 2.5× factor. The transmitter nonlinearities may also spread in-band signal components into the frequency band of the OOB frequency range. Furthermore, the transmitter oscillator sideband noise also may extend into that frequency range. Since it may not be practical to isolate these emissions, their level will tend to be included during OOB power measurements.

All emissions, including intermodulation products, conversion products, and parasitic emissions, which fall within frequencies separated from the center frequency of the emission by 2.5× or more of the necessary bandwidth of the emission will generally be considered as emissions in the spurious domain. However, this frequency separation may be dependent on the type of modulation, the maximum symbol rate in the case of digital modulation, the type of transmitter, and frequency coordination factors. For example, in the case of some digital, broadband, or pulse-modulated systems, the frequency separation may need to differ from the 2.5× factor. For multi-channel or multi-carrier transmitters, where several carriers may be transmitted simultaneously from the PA or an AAS, the center frequency of the emission is taken to be the center of either the assigned bandwidth of the BS or of the −3 dB bandwidth of the transmitter, using the lesser of the two bandwidths.

The above separation of the requirements can be easily applied to systems with fixed channel bandwidth. It does; however, become more difficult for NR, which supports flexible bandwidths, implying that the frequency range where requirements apply would then vary

Table 5.10: Maximum offset of operating band unwanted emission outside the downlink operating bands [5].

BS Type	Operating Band (MHz)	Δf_{OBUE} (MHz)
BS Type 1-H	$f_{DL-high} - f_{DL-low} < 100$	10
	$100 \leq f_{DL-high} - f_{DL-low} \leq 900$	40
BS Type 1-C	$f_{DL-high} - f_{DL-low} \leq 200$	10
	$200 < f_{DL-high} - f_{DL-low} \leq 900$	40

with the channel bandwidth. The approach taken for defining the boundary in 3GPP is slightly different for BS and the device requirements. With the recommended boundary between OOB emissions and spurious emissions set at $2.5\times$ the channel bandwidth, third-order and fifth-order intermodulation products from the carrier will fall inside the OOB domain, which will cover a frequency range of twice the channel bandwidth on each side of the carrier. For the OOB domain, two overlapping requirements are defined for the BS and the device, namely, spectrum emissions mask (SEM) and ACLR.

The OOB emission requirement for the BS transmitter is specified both in terms of ACLR and OBUE. The maximum offset of the OBUEs mask from the operating band edge is denoted as Δf_{OBUE}. The OBUE defines all unwanted emissions in each supported downlink operating band plus the frequency ranges Δf_{OBUE} above and Δf_{OBUE} below each band. The unwanted emissions outside of this frequency range are limited by a spurious emissions requirement. The values of Δf_{OBUE} are defined in Table 5.10 for the NR operating bands [5].

The occupied bandwidth is the size of a frequency band below the lower and above the upper frequency edges, where the mean power emitted is equal to $\beta/2$ (in percent) of the total mean transmitted power. The value of $\beta/2$ is equal to 0.5%. The occupied bandwidth requirement is applicable during the transmitter ON period for a single carrier. There may also be regional requirements to declare the occupied bandwidth according to the above definition. The occupied bandwidth is a regulatory requirement that is specified for the equipment in some regions. It was originally defined by the ITU-R as the maximum bandwidth, outside of which emissions do not exceed a certain percentage of the total emissions. The occupied bandwidth for NR is equal to the channel bandwidth, outside of which a maximum of 1% of the emissions are allowed (0.5% on each side).

The adjacent channel leakage ratio is the ratio of the filtered mean power centered on the assigned channel frequency to the filtered mean power centered on an adjacent channel frequency. The requirements apply outside the BS RF bandwidth or radio bandwidth regardless of the type of transmitter considered (single-carrier or multi-carrier) and for all transmission modes supported by the manufacturer's specification. These requirements are applied during the transmitter ON period.

The spectrum of an OFDM signal decays slowly outside of the transmission bandwidth. Since the transmitted signal for NR occupies up to 98% of the channel bandwidth, it is not possible to meet the unwanted emission limits directly outside the channel bandwidth, if certain time-domain windowing or frequency-domain shaping are not used. Filtering is always used both in the form of time-domain digital filtering of the baseband signal or analog-domain filtering of the RF signal.

The nonlinear characteristics of the PA must be taken into account, since it is the cause of intermodulation products outside the channel bandwidth. A power backoff can be used to operate in linear region of the PA at the cost of lower power efficiency, thus the amount of backoff should be minimized in practice. For this reason, additional linearization schemes can be employed [e.g., digital predistortion (DPD) for the PA]. These are especially important for the BS, where there are fewer restrictions on implementation complexity and use of advanced linearization schemes is an essential part of controlling unwanted emissions.

The emission mask defines the permissible OOB spectrum emissions outside the transmission bandwidth. Due to support of flexible channel bandwidths by NR, the determination of the frequency boundary between OOB emissions and spurious emissions is more involved for the NR BS and the devices. For the NR BS, the problem of implicit variation of the boundary between OOB and spurious domains with the varying channel bandwidth is addressed by not defining an explicit boundary. The 3GPP solution for this problem is the concept of OBUE for the NR base stations instead of the spectrum mask that is usually defined for OOB emissions. The OBUE requirement applies over the entire BS transmitter operating band, plus an additional 10−40 MHz on each side, as shown in Fig. 5.9. All requirements outside of that range are set by the regulatory spurious emission limits, based on the ITU-R recommendations. As shown in Fig. 5.9, a large part of the OBUE is defined over a frequency range that for smaller channel bandwidths can be both in spurious and OOB domains. This means that the limits for the frequency ranges that may be in the spurious domain also have to align with the regulatory limits from ITU-R. The shape of the mask is generic for all channel bandwidths, which has to align with ITU-R limits starting 10−40 MHz from the channel edges. The OBUE is defined with a 100 kHz measurement bandwidth and is aligned to some extent with the corresponding emission masks for the LTE systems. In the case of carrier aggregation for a BS, the OBUE requirement is applied as for any multicarrier transmission, where the OBUE will be defined relative to the carriers on the edges of the RF bandwidth. In the case of non-contiguous carrier aggregation, the OBUE within a subblock gap is partly calculated as the cumulative sum of contributions from each subblock [10].

In addition to a SEM, the OOB emissions are defined by the ACLR requirement. The ACLR concept is very useful for analysis of coexistence between two systems that operate on adjacent frequencies. The ACLR defines the ratio of the power transmitted within the

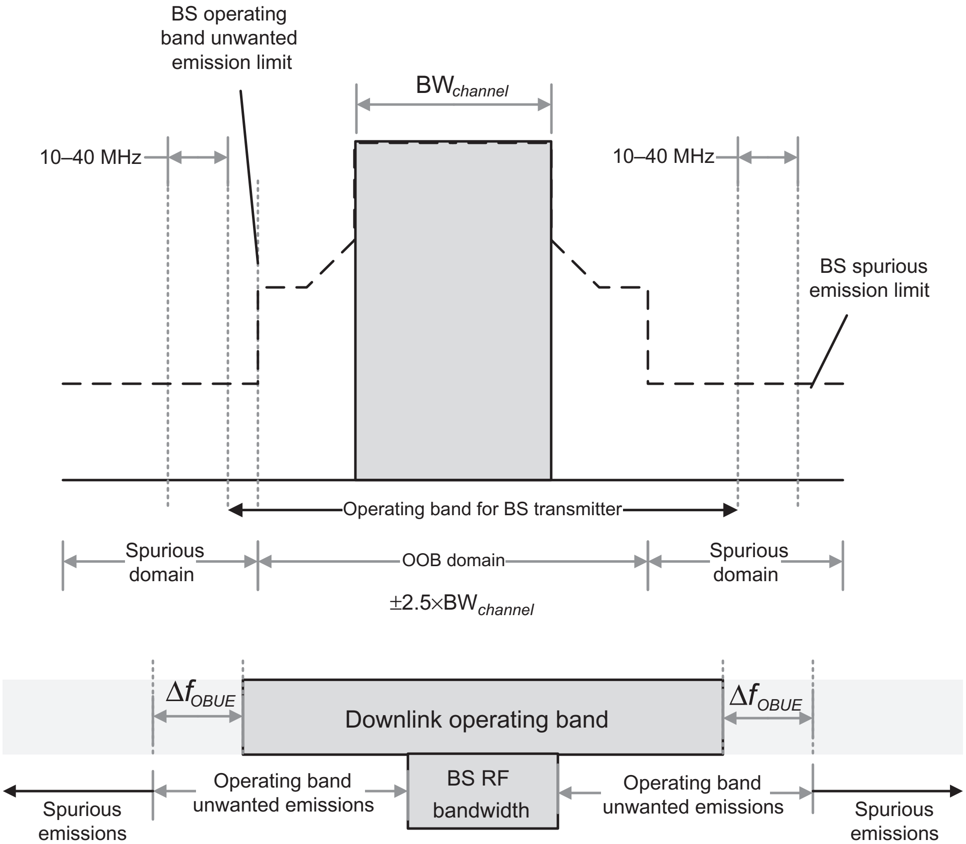

Figure 5.9
Illustration of the unwanted and spurious emissions applicable to NR base stations [5,8,9].

assigned channel bandwidth to the power of the unwanted emissions transmitted on an adjacent channel. There is a corresponding receiver requirement known as ACS, which defines the receiver's ability to suppress a signal on an adjacent channel. The definitions of ACLR and ACS are illustrated in Fig. 5.10 for a wanted and an interfering signal received in adjacent channels. The interfering signal's leakage of unwanted emissions at the wanted signal receiver is given by the ACLR and the ability of the receiver of the wanted signal to suppress the interfering signal in the adjacent channel is defined by the ACS. The two parameters when combined define the total leakage between two transmissions on adjacent channels. That ratio is called the adjacent channel interference ratio (ACIR) and is defined as the ratio of the power transmitted on one channel to the total interference received by a receiver on the adjacent channel, due to both the transmitter (ACLR) and the receiver (ACS). This relationship between the adjacent channel parameters is given as $\mathrm{ACIR} = 1/\left(1/\mathrm{ACLR} + 1/\mathrm{ACS}\right)$. The ACLR and ACS can be defined with different

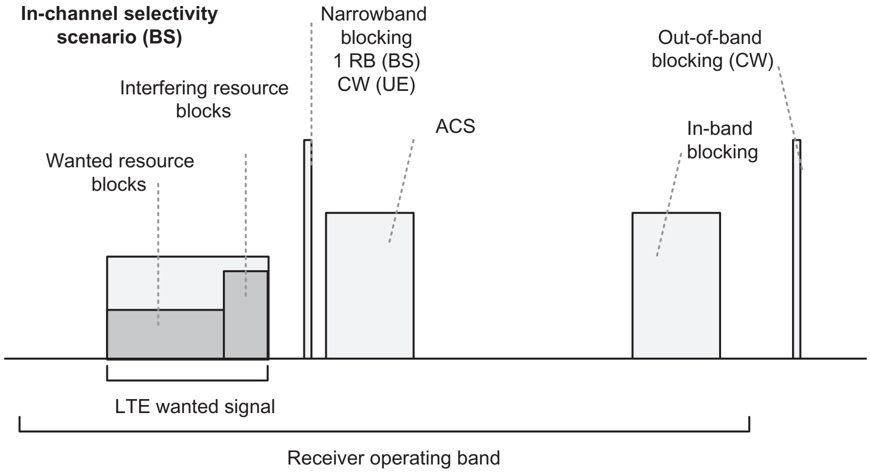

Figure 5.10
Illustration of the BS/UE requirements for receiver susceptibility to interference [5,10].

channel bandwidths for the two adjacent channels, which is the case for some requirements set for NR due to the bandwidth flexibility.

The ACLR limits for NR devices and base station have been derived based on extensive analysis of NR coexistence with NR or other systems on adjacent carriers. For an NR BS, there are ACLR requirements both for an adjacent channel with an NR receiver of the same channel bandwidth and for an adjacent LTE receiver. The ACLR requirement for NR BS is set to 45 dB. This is considerably more stringent than the ACS requirement for the device, which implies that in the downlink, the device receiver performance will be the limiting factor for ACIR and consequently for coexistence between the base stations and the devices. From a system perspective, this choice is cost-efficient since it moves implementation complexity to the BS, instead of requiring all devices to have high-performance RF components. In the case of carrier aggregation in a BS, the ACLR limits are applied as for any multi-carrier transmission, where the ACLR requirement will be defined for the carriers on the edges of the RF bandwidth. In the case of non-contiguous carrier aggregation where the subblock gap is very small that the ACLR requirements at the edges of the gap will overlap, a special cumulative ACLR (CACLR) requirement is defined for the gap. For CACLR, contributions from carriers on both sides of the subblock gap are taken into account in the calculation of CACLR limit. The CACLR limit is the same as the ACLR for the BS at 45 dB. The ACLR limits for the device are set based on the assumption that an NR or a UTRA receiver on the adjacent channel is used. In the case of carrier aggregation, the device ACLR requirement is applied to the aggregated channel bandwidth as opposed to per carrier. The ACLR limit for NR devices is set to 30 dB. This is considerably more relaxed compared to the ACS requirement for the BS, implying that in the uplink, the device transmitter performance will be the limiting factor for the ACIR and consequently for the coexistence between the base stations and the devices [10].

The limits for the BS spurious emissions are derived from ITU-R and are only defined in the region outside the frequency range of the OBUE limits (see Fig. 5.9) and in the frequencies that are separated from the BS transmitter operating band by at least 10−40 MHz. There are also additional regional or optional limits for protection of other systems that may coexist with or even co-located with the NR.

One of the implementation aspects of an RF transmitter is the possibility of intermodulation between the transmitted signal and other strong signals transmitted in the proximity of the BS or the device, which leads to the requirement for transmitter intermodulation. For the BS, the requirement is based on a stationary scenario with a co-located BS transmitter, with its transmitted signal appearing at the antenna connector of the BS being specified but attenuated by 30 dB. Since it is a stationary scenario, there are no additional unwanted emissions allowed, implying that all unwanted emission limits also have to be met in the presence of the interferer. For the device, there is a similar requirement based on a scenario with another device's transmitted signal appearing at the antenna connector of the device being specified but attenuated by 40 dB. The requirement specifies the minimum attenuation of the resulting intermodulation product below the transmitted signal [10].

The reference sensitivity requirement is meant to verify the receiver noise figure, which is a measure of how much the receiver's RF signal chain degrades the SNR of the received signal. For this reason a low-SNR transmission scheme using QPSK is chosen as a reference channel for the reference sensitivity test. The reference sensitivity is defined at a receiver input level where the throughput is 95% of the maximum throughput for the reference channel. For the device, reference sensitivity is defined in terms of the full channel bandwidth signals and with all resource blocks allocated to the wanted signal [5].

The reference sensitivity power level $P_{REFSENS}$ is the minimum mean power received at the antenna connector for BS Type 1-C or TAB connector for BS Type 1-H at which certain throughput requirement is met for a specified reference measurement channel. For NR, this throughput must be greater than 95% of the maximum throughput of the reference measurement channel. The reference sensitivity is a range of values that can be calculated as $P_{REFSENS} = 10\log(kT_0W_N) + \mathrm{NF} + \mathrm{SINR} + L_{IM}(\mathrm{dBm})$ for the BS, where kT_0W_N is the thermal noise power in dBm, W_N is the noise equivalent bandwidth (approximately the system bandwidth), NF is the maximum receiver noise figure, SINR is the minimum SINR for the chosen MCS, and L_{IM} is the implementation margin. Note that the noise floor for any receiver is defined as $\mathrm{Noise\_Floor} = 10\log(kT_0W_N) + \mathrm{NF} = -174 + \mathrm{NF} + 10\log W_N$ (dBm), where $\mathrm{NF} = \mathrm{SNR}_{in} - \mathrm{SNR}_{out}$ (in dB) is a measure of SINR degradation by the components in the RF signal path including the RF filters and low-noise amplifier (LNA). The thermal noise power density $kT_0 = -174$ dBm/Hz is measured at typical room

Table 5.11: NR wide area/medium range/local area base station reference sensitivity levels [5].

BS Channel Bandwidth (MHz)	Subcarrier Spacing (kHz)	Reference Measurement Channel	NR Wide Area BS Reference Sensitivity Power Level (dBm) $P_{REFSENS}$	NR Medium Range BS Reference Sensitivity Power Level (dBm) $P_{REFSENS}$	NR Local Area BS Reference Sensitivity Power Level (dBm) $P_{REFSENS}$
5, 10, 15	15	G-FR1-A1-1	− 101.7	− 96.7	− 93.7
10, 15	30	G-FR1-A1-2	− 101.8	− 96.8	− 93.8
10, 15	60	G-FR1-A1-3	− 98.9	− 93.9	− 90.9
20, 25, 30, 40, 50	15	G-FR1-A1-4	− 95.3	− 90.3	− 87.3
20, 25, 30, 40, 50, 60, 70, 80, 90, 100	30	G-FR1-A1-5	− 95.6	− 90.6	− 87.6
20, 25, 30, 40, 50, 60, 70, 80, 90, 100	60	G-FR1-A1-6	− 95.7	− 90.7	− 87.7

temperature 290 K degree. The LTE specifications require NF ≤ 9 dB for the UE and NF ≤ 5 dB for the eNB, nevertheless, commercial RF components and UE receivers can achieve lower than these limits. In NR the implementation margin is 2 dB and the NF is 5 dB for wide area BS, 10 dB for medium range BS and 13 dB for local area BS [14].

In Table 5.11, $P_{REFSENS}$ is the power level of a single instance of the reference measurement channel. This requirement must be met for each consecutive application of a single instance of the reference measurement channel mapped to disjoint frequency ranges with a width corresponding to the number of resource blocks of each reference measurement channel, except for one instance that might overlap with one other instance to cover the full *BS channel bandwidth*.

The dynamic range is specified as a measure of the capability of the receiver to receive a wanted signal in the presence of an interfering signal at the antenna connector for BS Type 1-C or TAB connector for BS Type 1-H within the received *BS channel bandwidth*. Under this condition, certain throughput requirement must be met for a specified reference measurement channel. The interfering signal for the dynamic range requirement is an additive white Gaussian noise. For NR, this throughput must be greater than 95% of the maximum throughput of the reference measurement channel (Table 5.12). The dynamic range requirement for the device is specified as the maximum signal level at which the throughput requirement can be met. The wanted signal mean power is the power level of a single instance of the corresponding reference measurement channel. This requirement must be

Table 5.12: Wide area/medium range/local area base station dynamic ranges [5].

BS Channel Bandwidth (MHz)	Subcarrier Spacing (kHz)	Reference Measurement Channel	Wide Area BS Wanted Signal Mean Power (dBm)	Wide Area BS Interfering Signal Mean Power (dBm)/BW_{config}	Medium Range BS Wanted Signal Mean Power (dBm)	Medium Range BS Interfering Signal Mean Power (dBm)/BW_{config}	Local Area BS Wanted Signal Mean Power (dBm)	Local Area BS Interfering Signal Mean Power (dBm)/BW_{config}	Type of Interfering Signal
5	15	G-FR1-A2-1	− 70.7	− 82.5	− 65.7	− 77.5	− 62.7	− 74.5	AWGN
	30	G-FR1-A2-2	− 71.4		− 66.4		− 63.4		
10	15	G-FR1-A2-1	− 70.7	− 79.3	− 65.7	− 74.3	− 62.7	− 71.3	AWGN
	30	G-FR1-A2-2	− 71.4		− 66.4		− 63.4		
	60	G-FR1-A2-3	− 68.4		− 63.4		− 60.4		
15	15	G-FR1-A2-1	− 70.7	− 77.5	− 65.7	− 72.5	− 62.7	− 69.5	AWGN
	30	G-FR1-A2-2	− 71.4		− 66.4		− 63.4		
	60	G-FR1-A2-3	− 68.4		− 63.4		− 60.4		
20	15	G-FR1-A2-4	− 64.5	− 76.2	− 59.5	− 71.2	− 56.5	− 68.2	AWGN
	30	G-FR1-A2-5	− 64.5		− 59.5		− 56.5		
	60	G-FR1-A2-6	− 64.8		− 59.8		− 56.8		
25	15	G-FR1-A2-4	− 64.5	− 75.2	− 59.5	− 70.2	− 56.5	− 67.2	AWGN
	30	G-FR1-A2-5	− 64.5		− 59.5		− 56.5		
	60	G-FR1-A2-6	− 64.8		− 59.8		− 56.8		
30	15	G-FR1-A2-4	− 64.5	− 74.4	− 59.5	− 69.4	− 56.5	− 66.4	AWGN
	30	G-FR1-A2-5	− 64.5		− 59.5		− 56.5		
	60	G-FR1-A2-6	− 64.8		− 59.8		− 56.8		
40	15	G-FR1-A2-4	− 64.5	− 73.1	− 59.5	− 68.1	− 56.5	− 65.1	AWGN
	30	G-FR1-A2-5	− 64.5		− 59.5		− 56.5		
	60	G-FR1-A2-6	− 64.8		− 59.8		− 56.8		
50	15	G-FR1-A2-4	− 64.5	− 72.2	− 59.5	− 67.2	− 56.5	− 64.2	AWGN
	30	G-FR1-A2-5	− 64.5		− 59.5		− 56.5		
	60	G-FR1-A2-6	− 64.8		− 59.8		− 56.8		
60	30	G-FR1-A2-5	− 64.5	− 71.4	− 59.5	− 66.4	− 56.5	− 63.4	AWGN
	60	G-FR1-A2-6	− 64.8		− 59.8		− 56.8		
70	30	G-FR1-A2-5	− 64.5	− 70.8	− 59.5	− 65.8	− 56.5	− 62.8	AWGN
	60	G-FR1-A2-6	− 64.8		− 59.8		− 56.8		
80	30	G-FR1-A2-5	− 64.5	− 70.1	− 59.5	− 65.1	− 56.5	− 62.1	AWGN
	60	G-FR1-A2-6	− 64.8		− 59.8		− 56.8		
90	30	G-FR1-A2-5	− 64.5	− 69.6	− 59.5	− 64.6	− 56.5	− 61.6	AWGN
	60	G-FR1-A2-6	− 64.8		− 59.8		− 56.8		
100	30	G-FR1-A2-5	− 64.5	− 69.1	− 59.5	− 64.1	− 56.5	− 61.1	AWGN
	60	G-FR1-A2-6	− 64.8		− 59.8		− 56.8		

met for each consecutive application of a single instance of the reference measurement channel mapped to disjoint frequency ranges with a width corresponding to the number of resource blocks of each reference measurement channel, except for one instance that might overlap another instance to cover the full *BS channel bandwidth* [5].

There are a set of requirements for the BS and the device, which define the receiver ability to receive a wanted signal in the presence of a stronger interfering signal. There are different interference scenarios depending on the frequency offset of the interferer from the wanted signal, where different types of receiver impairments will affect the performance.

A set of combinations of the interfering signals with different bandwidths can be defined which model the practical scenarios that are encountered within and outside of the BS/device receiver bandwidth. While the types of requirements are very similar for the BS and the device, the signal levels are different since the interference scenarios for the BS and device can vary. The following requirements are defined for NR base stations and the devices. In all cases where the interfering signal is an NR signal, it has the same or smaller bandwidth (less than 20 MHz) relative to the desired signal [5]:

- *In-band/Out-of-band Blocking*: Blocking refers to the scenario where strong interfering signals are present outside the operating band (OOB blocking) or inside the operating band (in-band blocking), but not adjacent to the desired signal. The in-band blocking characteristics is a measure of the receiver's ability to receive a desired signal in its assigned channel at the antenna connector for BS Type 1-C or TAB connector for BS Type 1-H in the presence of an unwanted interferer, which is an NR signal for general blocking or an NR signal with one resource block for narrowband blocking (see Fig. 5.10). In-band blocking includes interferers within 20−60 MHz range outside of the operating band for the base stations and within 15 MHz range of the operating band for the devices. These scenarios are modeled with a continuous wave (CW) signal for the out-of-band case and an NR signal for the in-band case. There are additional BS blocking requirements for the scenario in which the BS is co-located with another BS in a different operating band. For the UEs, a number of exceptions from the out-of-band blocking requirements are allowed, where the device is required to comply with more relaxed spurious response requirements.

  For a BS operating in non-contiguous spectrum in any operating bands, the in-band blocking requirements further apply in subblock gaps, if the subblock gap size is at least twice as wide as the interfering signal minimum offset. The interfering signal offset is defined relative to the subblock edges inside the subblock gap. For a multi-band connector, the blocking requirements apply in the in-band blocking frequency ranges for each supported operating band. For a BS operating in non-contiguous spectrum within any operating bands, the narrowband blocking requirements are further applied within the subblock gap, if the subblock gap size is at least as wide as the channel bandwidth of the NR interfering signal. The interfering signal offset is defined relative to the subblock edges inside the subblock gap. For a multi-band connector the narrowband blocking requirement is further applied inside any inter-RF bandwidth gap,[6] in case the inter-RF bandwidth gap size is at least as wide as the NR interfering signal. The interfering signal offset is defined relative to the base station RF bandwidth edges inside the inter-RF bandwidth gap.

- *Adjacent Channel Selectivity (ACS)*: This is a measure of the receiver's ability to receive a desired signal in its assigned channel frequency at the antenna connector for BS type 1-C or TAB connector for BS type 1-H in the presence of an adjacent channel signal with a specified center frequency offset from the interfering signal to the edge of the band of a victim system. For a BS operating in non-contiguous spectrum within any operating bands, the ACS requirement is further applied within the subblock gaps, in case the subblock gap size is at least as wide as the NR interfering signal. The interfering signal offset is defined relative to the subblock edges inside the subblock gap. For a multi-band connector, the ACS requirement is applied within the inter-RF bandwidth gap, in case the inter-RF bandwidth gap size is at least as wide as the NR interfering signal. The interfering signal offset is defined relative to the base station RF bandwidth edges inside the inter-RF bandwidth gap. This corresponds to a scenario where there is a strong signal in the channel adjacent to the desired signal and is closely related to the corresponding ACLR requirement. The adjacent interferer is an NR signal. For the device, the ACS is specified for two cases with a lower and a higher signal level.
- *In-Channel Selectivity (ICS)*: This is a measure of the receiver ability to receive a desired signal in its assigned resource block locations at the antenna connector for BS type 1-C or TAB connector for BS type 1-H in the presence of an interfering signal received with a larger power spectral density. Under this condition, certain throughput requirement (throughput must be greater than 95% of the maximum throughput of the reference measurement) must be met for a specified reference measurement channel. The interfering signal is an NR signal which is time-aligned with the desired signal. In other words, in this scenario, multiple signals with different power levels are received within the channel bandwidth, where the performance of the weaker desired signal is verified in the presence of the stronger interfering signal. The ICS requirement is only specified for the base stations (see Fig. 5.10).
- *Receiver Intermodulation*: The third and higher order harmonics resulted from mixing of the two interfering RF signals can produce an interfering signal in the band of the desired channel. Intermodulation response rejection is a measure of the capability of the receiver to receive a desired signal in its assigned channel bandwidth at the antenna connector for BS type 1-C or TAB connector for BS type 1-H in the presence of two

[6] Inter-RF bandwidth gap is the gap between the RF bandwidths in the two bands. Note that the inter-RF bandwidth gap may span a frequency range where other mobile operators can be deployed in bands *X* and *Y*, as well as the frequency range between the two bands that may be used for other services.

interfering signals which have a specific frequency relationship to the desired signal. In this scenario, there are two interfering signals in the proximity of the desired signal, where the interferers are a continuous wave signal and an NR signal. The purpose of this requirement is to test the receiver linearity. The interferers are placed at frequencies in such a way that the main intermodulation product falls within the desired signal channel bandwidth. There is also a narrowband intermodulation requirement for the BS where the CW signal is very close to the desired signal and the NR interferer is a single-resource-block signal (narrowband).

- *Receiver Spurious Emissions*: The receiver spurious emissions power is the power of emissions generated or amplified in a receiver unit that appear at the antenna connector (for BS Type 1-C) or at the TAB connector (for BS Type 1-H). The requirements are applicable to all base stations with separate RX and TX antenna connectors/TAB connectors. For systems operating in FDD mode, the test is performed when both TX and RX are ON, with the TX antenna connectors/TAB connectors terminated. For antenna connectors/TAB connectors supporting both RX and TX in TDD, the requirements are verified during the transmitter OFF period. For antenna connectors/TAB connectors supporting both RX and TX in FDD mode, the RX spurious emissions requirements are superseded by the TX spurious emissions requirements. For RX-only multi-band connectors, the spurious emissions requirements are subject to exclusion zones in each supported operating band. For multi-band connectors which transmit and receive in operating band supporting TDD mode, the RX spurious emissions requirements are verified during the TX OFF period and are subject to exclusion zones in each supported operating band. The number of active receiver units that are considered when calculating the conducted RX spurious emission limits ($N_{RXU-counted}$) for BS Type 1-H is given as $N_{RXU-counted} = \min(N_{RXU-active}, 8N_{cells})$ where $N_{RXU-counted-per-cell}$ is used for scaling of basic limits and is derived as $N_{RXU-counted-per-cell} = N_{RXU-counted}/N_{cells}$, where N_{cells} was defined earlier. The $N_{RXU-active}$ is the number of active receiver units and is independent of the declaration of N_{cells}. Since the receiver emissions are dominated by the transmitted signal, the receiver spurious emission limits are only applicable when the transmitter is not active, and also when the transmitter is active for an NR base station operating in FDD mode that has a separate receiver antenna connector [5].

5.3 UE Transceiver RF Characteristics and Requirements

5.3.1 General UE RF Requirements

An NR UE should support both standalone and non-standalone operational modes. For the non-standalone operation, the LTE carrier(s) is required to be the anchor carrier for the NR UEs. In addition to LTE and NR, there may be operators around the globe which rely on 3G and other 2G services. Those radio access technologies may utilize different frequency bands/channels for operation. The discrete (non-integrated) RF front-end architectures that support these RATs and their associated frequency bands may provide good RF performance; however, there would be implications related to cost, form factor, and hardware complexity. Multi-mode multi-band integrated RF architectures may overcome many of the challenges of the discrete design. Such architectures would combine several bands into a single transceiver chain, independent of radio access technology. The RF front-end subsystem comprises PAs, filters, duplexers, switches, oscillators, mixers, LNAs and other passive and active RF components that enable the device to conform to 3GPP and regional regulatory emission specifications. It is understood that various OEMs may combine RF front-end components in different groupings depending on component selection and functionality provided by those components for carrier aggregation and multi-connectivity purposes. Various architectures combine mid-frequency bands (1.7−2.1 GHz and 2.3−2.7 GHz) into a single module to simplify the implementation of carrier aggregation. Similarly, high-frequency bands (in sub-6 GHz) can be combined with Wi-Fi and share antennas.

Due to different transmit/receive configurations (e.g., antenna configurations), network deployments, and different downlink and uplink transmit power, the LTE coverage in general was limited in the uplink. According to the evaluation done for LTE Band 41 Power Class 2 (+26 dBm), the coverage asymmetry could be up to 5 dB based on the network deployment parameters. To improve the uplink coverage, an effective way is to increase the transmit power. Based on some initial analysis of the NR link budget, it appeares that it would suffer from the same issue. Therefore, operators proposed to specify Power Class 2 (+26 dBm) and Power Class 3 (+23 dBm) UE in 3GPP RAN Working Group 4 [2,25].

In contrast to the NR BS, the UE implementations utilize beamforming in FR2 and the RF front-end configurations in FR1 are not significantly different from LTE UEs except in the new 3.7 and 4.5 GHz band implementations. The early deployments of NR use non-standalone mode, thus the compliant UEs will have to implement both LTE and NR radio transceivers. In FR1, the RF performance is specified for the new radio parameters such as the maximum transmit power and receiver sensitivity. In this frequency range, the conducted requirements for antenna connectors are used similar to those of LTE. On the contrary, the FR2 implementations use integrated transceivers and antennas and the measurements cannot be conducted at the connectors, which was the case in the base stations. Therefore, the OTA requirements were introduced to allow UE RF performance measurements. The requirements are defined for EIRP and maximum transmit power in FR2 using the cumulative distribution function of the EIRP values obtained/measured, when performing beamforming, over a full sphere with the UE at the center of the sphere (see Fig. 5.11). These requirements were introduced to ensure (in statistical sense) that the beam can be aimed at the intended direction and toward the serving BS while fulfilling the uplink link budget requirement [29].

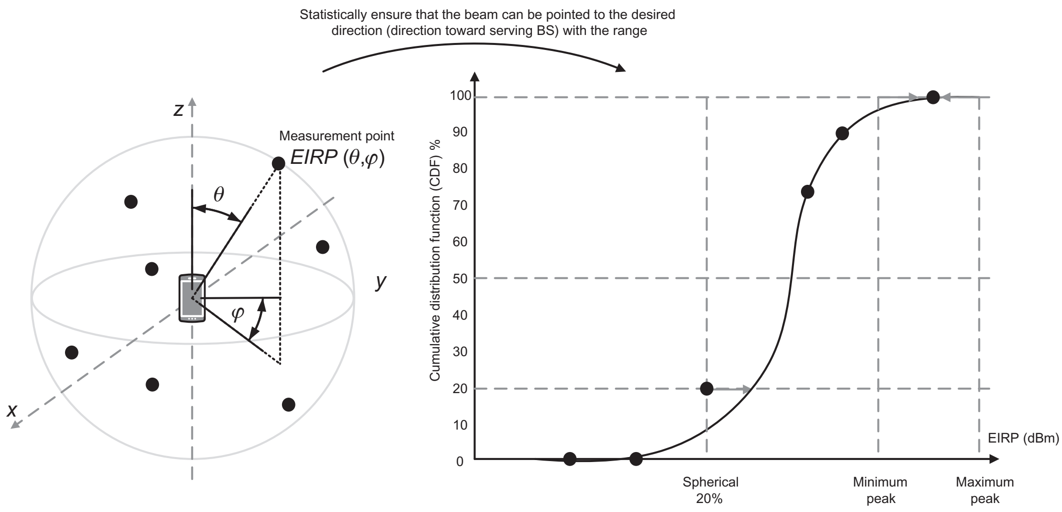

Figure 5.11
OTA EIRP evaluation using cumulative distribution function [29].

Table 5.13: UE power classes in FR2 [3,29].

FR2 Power Class	UE Type	EIRP (dBm)			TRP (dBm)
		Maximum	Minimum Peak	Spherical	Maximum
1	Fixed wireless terminal	55	40	32 (85%)	35
2	Vehicle-mounted terminal	43	29	18 (60%)	23
3	Mobile terminal	43	22.5	11.5 (50%)	23
4	Fixed wireless terminal	43	34	25 (20%)	23

In Fig. 5.11, the minimum peak value is defined as the value that at least one of the measured EIRP values would be exceeded. The spherical coverage is defined as $X\%$ CDF, where the value must be maintained over $(100-X)\%$ of the surface area of the sphere. The maximum value is defined as the value that the maximum measured EIRP must not exceed. The maximum EIRP is set based on regional regulatory requirements. The NR Rel-15 further considers methods to increase the effective range as well as increasing data rates for fixed wireless terminals. The NR UEs are categorized based on the power class in the standards specifications. There are four power classes for the NR devices operating in FR2, which are characterized by the transmit power and the spherical coverage as shown in Table 5.13.

The UE power class 1 is meant for transportable terminals and allows a maximum TRP of 35 dBm compared to other UE power classes with maximum permitted TRP of 23 dBm. It assumes use cases where an equipment at fixed position is able to emit a strong signal with a narrow beam in a particular direction. The UE power class 2 is intended for vehicular use cases. The target for spherical coverage is 60% and the maximum TRP is the same as power class 3 for mobile terminals; however, it requires a higher EIRP value. The UE power class 3 is defined for mobile handsets where the orientation of the terminal antennas with respect to the BS is random. Therefore, the target for spherical coverage is 50%. The UE power class 4 requires wider spherical coverage (20%) and higher EIRP than power class 2. Since this type of the terminal can be installed with the knowledge of the location of the BS, in contrast to the narrow-beam operation of power class 1, it can be used more flexibly in the scenarios involving moving platforms such as vehicles and trains [29].

5.3.2 Conducted and Radiated UE Transceiver Characteristics

The radiated RF requirements for the NR devices are primarily derived from the corresponding conducted RF requirements. Unlike the conducted requirements, the radiated requirements are inclusive of the antennas, thus the emission levels such as BS output power and unwanted emissions are defined either by incorporating the antenna gain as a directional requirement using an EIRP or by defining limits using TRP. The TRP and EIRP metrics are related through the number of radiating antennas and are implementation specific, considering the geometry of the antenna array and the correlation between unwanted emission signals from different antenna ports. An EIRP limit results in different levels of total radiated unwanted emission power depending on the implementation. The EIRP limit per se cannot control the total amount of interference in the network, whereas a TRP requirement limits the total amount of interference in the network regardless of the specific implementation. In the case of passive systems, the antenna gain does not noticeably vary between the desired signal and the unwanted emissions. Thus EIRP is directly proportional to TRP and can be used interchangeably. For an active system such as NR, the EIRP may vary considerably between the desired signal and the unwanted emissions and also between implementations, thus EIRP (in this case) is not proportional to TRP and should not substitute TRP [10].

The RF requirements in FR2 bands are specified separately for devices, because of the higher number of antenna elements for operation in FR2 and the high level of integration when mmWave technologies are used. In general, the RF requirements in FR2 are the same as the conducted RF requirements for FR1; however, the limits for certain requirements might be different. The difference in coexistence in mmWave bands leads to relatively relaxed requirements on ACLR and spectrum mask, which was demonstrated through coexistence studies performed in 3GPP. The mmWave implementations are more challenging than the more mature technologies in the frequency bands below 6 GHz. It should be noted that the channel bandwidths and numerologies supported in FR2 (see Table 5.1) are different from those used in FR1, making the comparison of the requirement levels, especially those related to the receiver somewhat difficult. The following is a list of the radiated RF requirements in FR2 [3,10]:

- *Output power level*: The maximum output power is expressed as TRP and/or EIRP. In FR2, the minimum output power and transmitter OFF power levels are higher than those in FR1. The radiated transmit power is an additional requirement, which unlike the maximum output power, is directional.
- *Transmitted signal quality*: Frequency error and EVM requirements are defined similar to those in FR1 and with the same limits.
- *Radiated unwanted emissions*: Occupied bandwidth, ACLR, spectrum mask, and spurious emissions are defined in the same way as for FR1. The latter two are based on TRP. The limits are relatively less stringent than those in FR1. The ACLR is approximately 10 dB more relaxed compared to FR1, due to more suitable coexistence conditions.
- *Reference sensitivity and dynamic range*: These are defined similar to the counterparts in FR1; however, the levels are not comparable.
- *Receiver susceptibility to interfering signals*: ACS, in-band, and OOB blocking are defined similar to FR1. There is no requirement for narrowband blocking, since there are only wideband systems in FR2. The ACS is approximately 10 dB more relaxed relative to its counterpart in FR1 due to better coexistence conditions (Fig. 5.12).

The device output power level is set by taking the following considerations into account [10]:

1. *UE Power Class* defines a nominal maximum output power for QPSK modulation. It may be different in different operating bands, but the main device power class is currently set to 23 dBm for all bands.
2. *Maximum Power Reduction (MPR)* defines the allowed reduction of maximum power level for certain combination of modulation scheme and resource block allocation. The

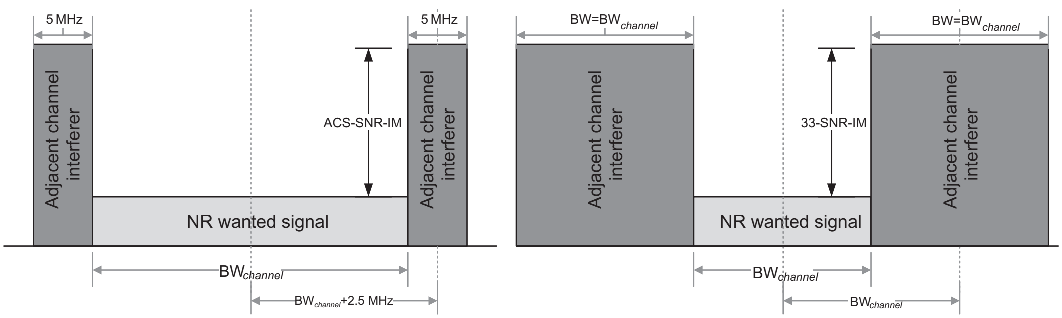

Figure 5.12
Illustration of UE ACS requirements [14].

UE is allowed to reduce the maximum output power due to higher order modulations and transmit bandwidth configurations. For UE power class 2 and 3 the allowed MPR are specified in [3] for channel bandwidths that meets both of the following criteria: Channel_Bandwidth $\leq$ 100 MHz and Relative Channel Bandwidth $\leq$ 4% (TDD bands) and $\leq$ 3% (FDD bands), where Relative Channel Bandwith $= 2\mathrm{BW}_{channel}/\left(f_{UL\_low} + f_{UL\_high}\right)$.

3. *Additional Maximum Power Reduction (A-MPR)* may be applied in some regions and is usually related to specific transmitter requirements such as regional emission limits and to certain RF carrier configurations. For those requirements, there is an associated NS value that identifies the permissible A-MPR and the associated conditions.

The minimum output power level configuration defines the device dynamic range. There is a definition of the transmitter OFF power level, which is applicable to conditions where the device is not allowed to transmit. There is also a general ON/OFF time mask specified, plus specific time masks for PRACH, PUCCH, SRS, and for PUCCH/PUSCH/SRS transitions. The device transmit power control is specified through requirements for the absolute power tolerance for the initial power setting, the relative power tolerance between two subframes, and the aggregated power tolerance for a sequence of power control commands. In-band emissions are emissions within the channel bandwidth. The requirement limits a device's transmission leakage into non-allocated resource blocks within the channel bandwidth. Unlike the OOB emissions, the in-band emissions are measured after cyclic prefix removal and FFT blocks, representing how a device transmitter affects a BS receiver in practice. For implementation purposes, it is not possible to define a generic device spectrum mask that does not vary with the channel bandwidth, thus the frequency ranges for OOB limits and spurious emissions limits do not follow the same principle as for the BS. The SEM extends to a separation Δf_{OOB} from the channel edges, as illustrated in Fig. 5.13. For 5 MHz channel bandwidth, this point corresponds to 250% of the operating bandwidth as recommended by ITU-R, but for wider channel bandwidths, it is set closer than 250%. The SEM is defined as a general mask and a set of additional masks that can be applied to reflect different regional requirements. Each additional regional mask is associated with a specific NS value. The device spurious emission limits are defined for all frequency ranges outside the frequency range covered by the SEM. The limits are not only based on regional regulatory requirements, but also additional requirements considered for coexistence with other frequency bands when the device is roaming. The additional spurious emission limits can have an associated NS value. Furthermore, there are BS and device emission limits defined for the receiver. Since the receiver emissions are dominated by the transmitted signal, the receiver spurious emission limits are only applicable when the transmitter is not active, or when the transmitter is active for an NR FDD base station that has a separate receiver antenna connector [10].

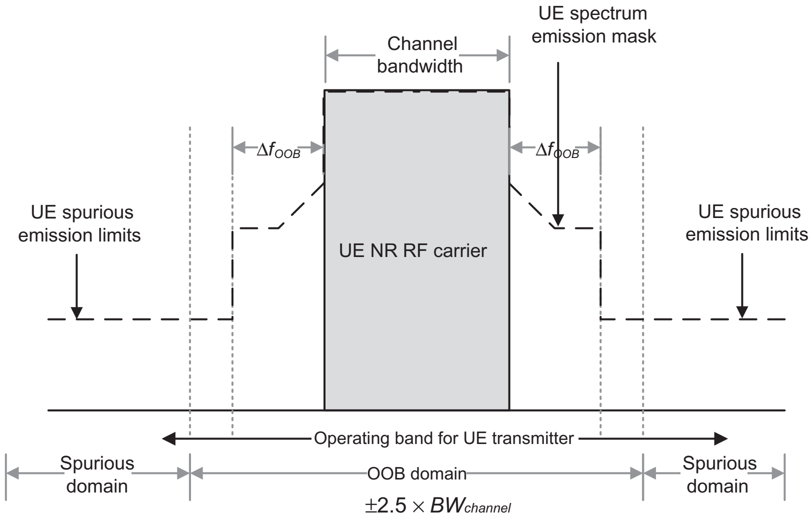

Figure 5.13
Illustration of the spectrum emission mask and spurious emissions requirements for the NR devices [3].

The UE output RF spectrum consists of the following regions that are regulated by 3GPP and regional regulatory requirements [14]:

- *Out-of-Band Emissions*: These are unwanted emissions occurring immediately outside of the assigned channel bandwidth, resulting from the modulation process and nonlinearity in the transmitter RF components. The OOB emissions are characterized in terms of the following metrics (see Fig. 5.14):
  - *Spectrum Emission Mask (SEM)*: The SEM is defined starting from each edge of the assigned NR channel bandwidth to $\pm(\text{BW}_{channel} + 5\text{MHz})$ for FR1 and starting from each edge of the assigned NR channel bandwidth to $\pm 2\text{BW}_{channel}$ for FR2.
  - *Adjacent Channel Leakage Ratio (ACLR)*: The NR ACLR and UTRA ACLR requirements used for FR1 and the NR ACLR requirements are applied for FR2 bands.
- *Spurious Emissions*: These are caused by unwanted transmitter effects such as harmonics emission, parasitic emissions, intermodulation products, and frequency conversion products which cover a frequency range up to the fifth harmonic or 26 GHz for FR1 and a frequency range up to the second harmonic or the uplink operating band for FR2.

The UE reference sensitivity in FR1 is defined as a power level $P_{REFSENS} = 10\log(kT_0W_N) + \text{NF} + \text{SINR} + L_{IM} - G_{DIV}$ (dBm) where kT_0W_N is the thermal noise power in dBm, W_N is the noise equivalent bandwidth (approximately the UE bandwidth), NF is the maximum receiver noise figure, SINR is the minimum SINR for the chosen MCS, and L_{IM}

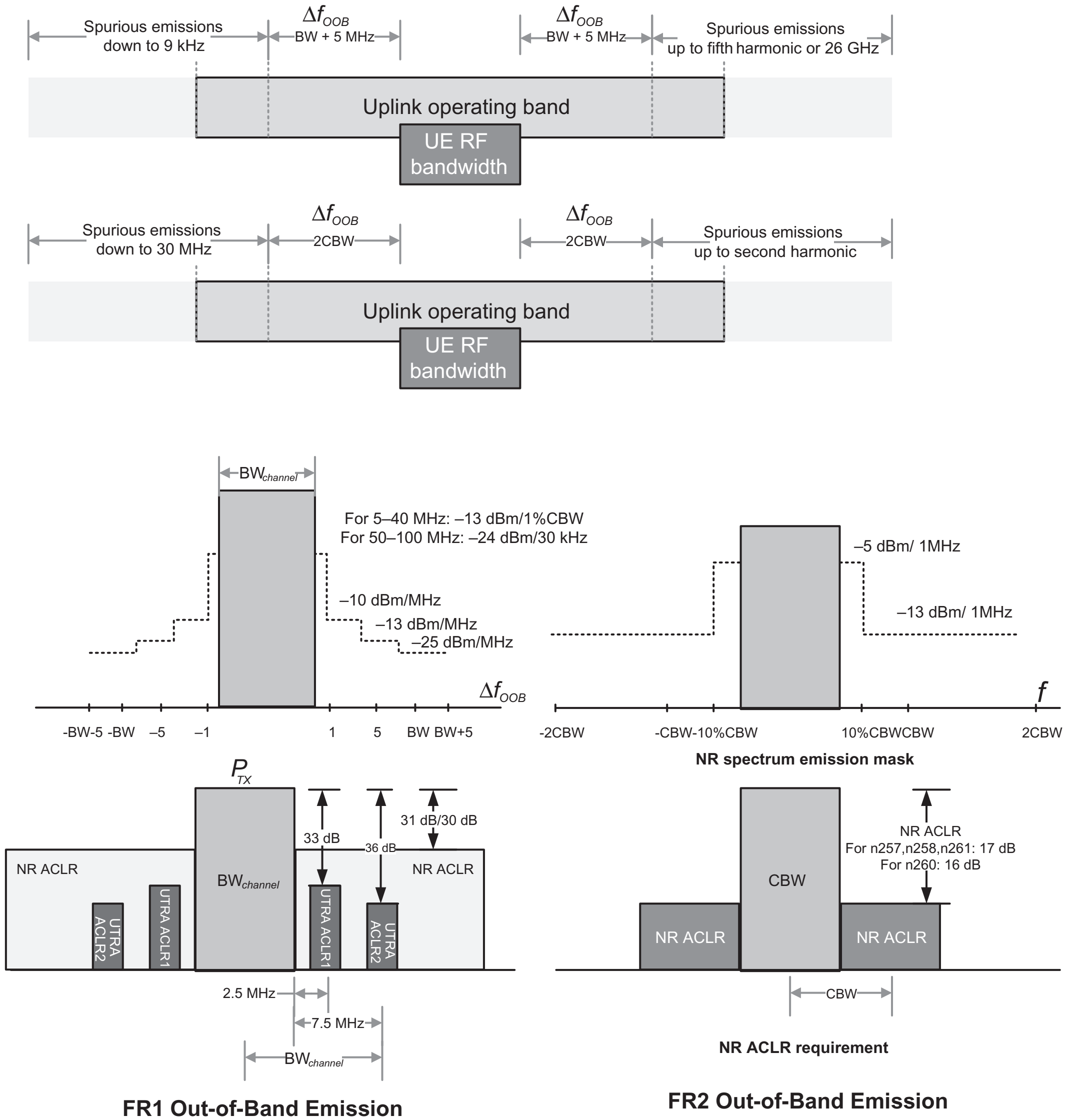

Figure 5.14
UE output RF spectrum and out-of-band emissions [2,3].

is the implementation margin. Note that the noise floor for any receiver is defined as $\text{Noise\_Floor} = 10\log(kT_0 W_N) + \text{NF} = -174 + \text{NF} + 10\log W_N$ (dBm), where $\text{NF} = \text{SNR}_{in} - \text{SNR}_{out}$ (in dB) is a measure of SINR degradation by the components in the RF signal path including the RF filters and LNA. The thermal noise power density $kT_0 = -174$ dBm/Hz is measured at typical room temperature 290 K degree. 3GPP assumes minimum SINR $= -1$ dB for the UE, the implementation margin of $L_{IM} = 2.5$ dB and a diversity gain of $G_{DIV} = 3$ dB for a UE for two RX antennas. The noise figure NF is 9 dB for LTE refarming bands, 10 dB for n77 (3.3−3.8 GHz), n78, and n79 bands, and 10.5 dB for n77 (3.8−4.2 GHz) band. In FR1, both 2RX REFSENS (for all bands) and 4RX REFSENS (for n7, n38, n41, n77, n78, n79) bands are defined. In FR2, REFSENS power level is the EIS level in the RX beam peak direction. In carrier aggregation and NR-LTE dual-connectivity scenarios, an additional relaxation is defined for the following cases: (1) band-combination-specific Δ_{RIB} for FR1 inter-band carrier aggregation; (2) aggregated *Channel BW* specific Δ_{RIB} for FR2 intra-band continuous carrier aggregation, that is, 0.5 dB for aggregated bandwidths greater than 800 MHz; and (3) band-combination-specific maximum sensitivity degradation (MSD)[7] for the band impacted by harmonic interference and intermodulation interference for the carrier aggregation and E-UTRA-NR dual-connectivity (EN-DC) scenarios [14]. Note that 3GPP allows certain amount of self-desensitization when TX emissions fall into the receiver band in non-contiguous carrier aggregation transmissions. This is done by relaxing the reference sensitivity requirements by an amount known as MSD when the UE transmits at the maximum power. However, these approaches (MPR and MSD) adversely impact the uplink link budget and throughput and are not the most appealing solutions. Alternatively, one could utilize higher quality RF components with good linear characteristics, which may considerably increase the overall radio implementation cost and size.

5.4 Radio Resource Management Specifications

Radio resource management (RRM) specifications are developed by 3GPP to ensure that the mobile terminals can maintain robust and reliable connection(s) to one or more base stations. The RRM specifications include mobility management, handover, cell quality measurements based on reference signals, or physical synchronization signals. For the purpose of handover to a neighbor cell or adding a new component carrier in carrier aggregation, the UE is required to conduct measurements on the neighboring cells signal quality using reference signal received power (RSRP) or reference signal received quality metrics (RSRQ). In LTE systems, all eNBs continuously transmit cell-specific reference signals (CRS), this would enable the terminals to measure the cell quality of the neighboring cells. However, NR eliminated always-on reference signals to reduce inter-cell interference and the overhead.

[7] The most basic receiver requirement is reference sensitivity. This is similar in concept to LTE/UMTS although there is a new dimension added with the introduction of the MSD parameter. This is a relaxation in the reference sensitivity level that applies when the UE is transmitting at maximum power (with MPR applied) using the maximum number of resource blocks allowed for the channel bandwidth. Under these conditions, it is expected that there will be a loss of sensitivity in the UE receiver.

In NR, the cell quality is measured by using SSBs. Each SSB comprises two synchronization signals and the physical broadcast channel which has a longer transmission periodicity compared to CRS. The SSB periodicity can be configured in the range of 5, 10, 20, 40, 80, and 160 ms for each cell; however, terminals do not need to measure cell quality with the same periodicity as the SSB and the appropriate measurement periodicity can be configured according to the channel conditions. This would avoid unnecessary measurements and conserve terminal power. The new SSB-based RRM measurement time configuration (SMTC)[8] window has been introduced to notify the terminals of the periodicity and the timing of the SSBs that the terminal must use for cell quality measurements. As shown in Fig. 5.15, the SMTC window periodicity can be set in the same range as the SSBs (i.e., 5, 10, 20, 40, 80, and 160 ms) and the duration of the window can be set to 1, 2, 3, 4, or 5 ms according to the number of SSBs transmitted on the cell that is being measured. When a UE is notified of the SMTC window by the gNB, it detects and measures the SSBs within that window and reports the measurement results back to the serving base station [6,29].

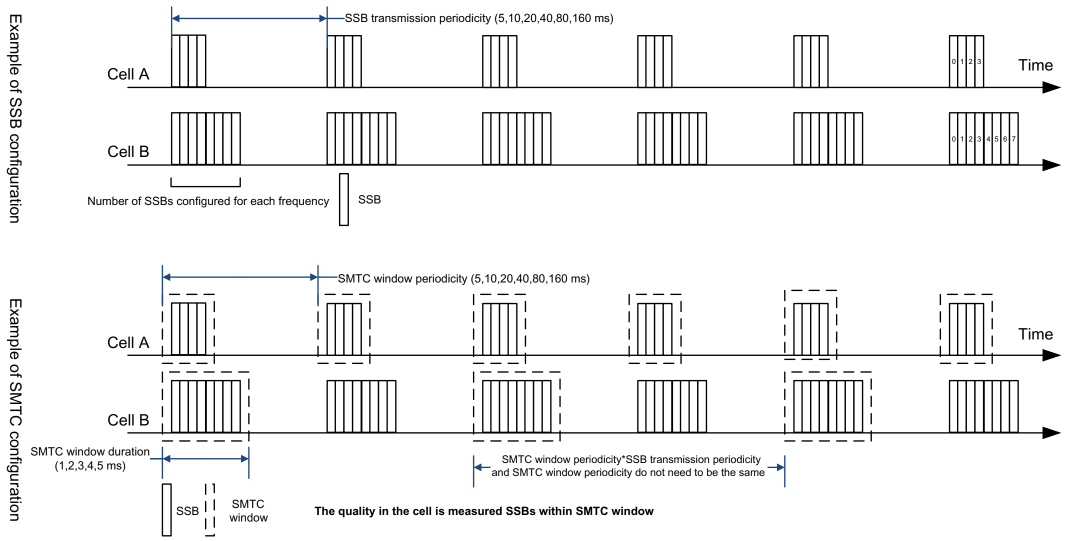

Figure 5.15
Example illustration of SSB and SMTC correspondence in NR [29].

8 SS-block-based RRM measurement timing configuration or SMTC is the measurement window periodicity/duration/offset information for UE RRM measurement per-carrier frequency. For intra-frequency connected mode measurement, up to two measurement window periodicities, can be configured. For the idle mode measurements, a single SMTC is configured per carrier frequency. For interfrequency connected mode measurements, a single SMTC is configured per-carrier frequency.

Using the same RF transceiver for measuring neighbor-cell quality or other component carriers as well as transmitting/receiving data in the serving cell would make it possible to reduce implementation cost. Nevertheless, this would imply that data cannot be transmitted/received in the serving cell during measurement of other cells or component carriers at different frequencies. In LTE, the UE data transmission in the serving cell is suspended during a measurement gap, providing the UE the opportunity to tune its RF transceivers to conduct neighbor-cell quality measurements or measurement of other component carriers at different frequencies. The concept of measurement gaps is used in NR; however, the measurements are conducted on SSBs and the measurement gap configurations have been improved compared to LTE. In LTE, the measurement gap length (MGL) is fixed such that at least one primary/secondary synchronization signal occasion can be observed within the gap. In LTE, the primary/secondary synchronization signals are transmitted every 5 ms; thus the MGL of LTE is 6 ms, allowing 0.5 ms for RF tuning in the beginning and in the end of the measurement gap. The terminals detect synchronization signals within the MGL and identify the cell ID and the reception timing. The terminal later performs measurements on the CRS. In NR, the SMTC window duration can be set to match the SSB transmission. However, a fixed MGL could cause potential degradation of the serving cell throughput. As an example, if the SMTC window duration is 2 ms and the MGL is 6 ms, a 4 ms interval will not be available for data transmission and reception in the serving cell [6,29].

In order to minimize such performance degradation, the MGL in NR can be configured and can be set to 1.5, 3, 3.5, 4, 5.5, or 6 ms. Fig. 5.16 illustrates the concept of measurement gap configuration in NR, which shows different SMTC window and MGLs. The measurement gap repetition period (MGRP) in NR can be more flexibly configured compared to LTE (Table 5.14). The NR supports 20, 40, 80, and 160 ms MGRPs compared to LTE that only supports 40 and 80 ms. As we mentioned earlier, an RF transceiver tuning time is provisioned in the beginning and the end of the measurement gap, during which the terminal cannot conduct measurements or transmit/receive data. As shown in Fig. 5.17, the SMTC window and the measurement gap start at the same time. The start of the SMTC window may overlap with the RF retuning time and the UE cannot perform measurements during that time. As such, the new measurement gap timing advance function has been introduced, which enables all SSBs (to be measured) within the SMTC window to be used by advancing the start of the RF retuning gap.

The SMTC window and the measurement gaps are configured based on the SSB transmission timing for the candidate cell. However, in some cases when measuring other cells, the serving BS may not have such information. In such cases, it is not possible to properly set the SMTC window and measurement gap for the UE. Therefore, NR defines an SFN and frame timing difference (SFTD) measurement function, where the terminal measures the timing difference of the SFN and the frame boundary between the serving cell and the

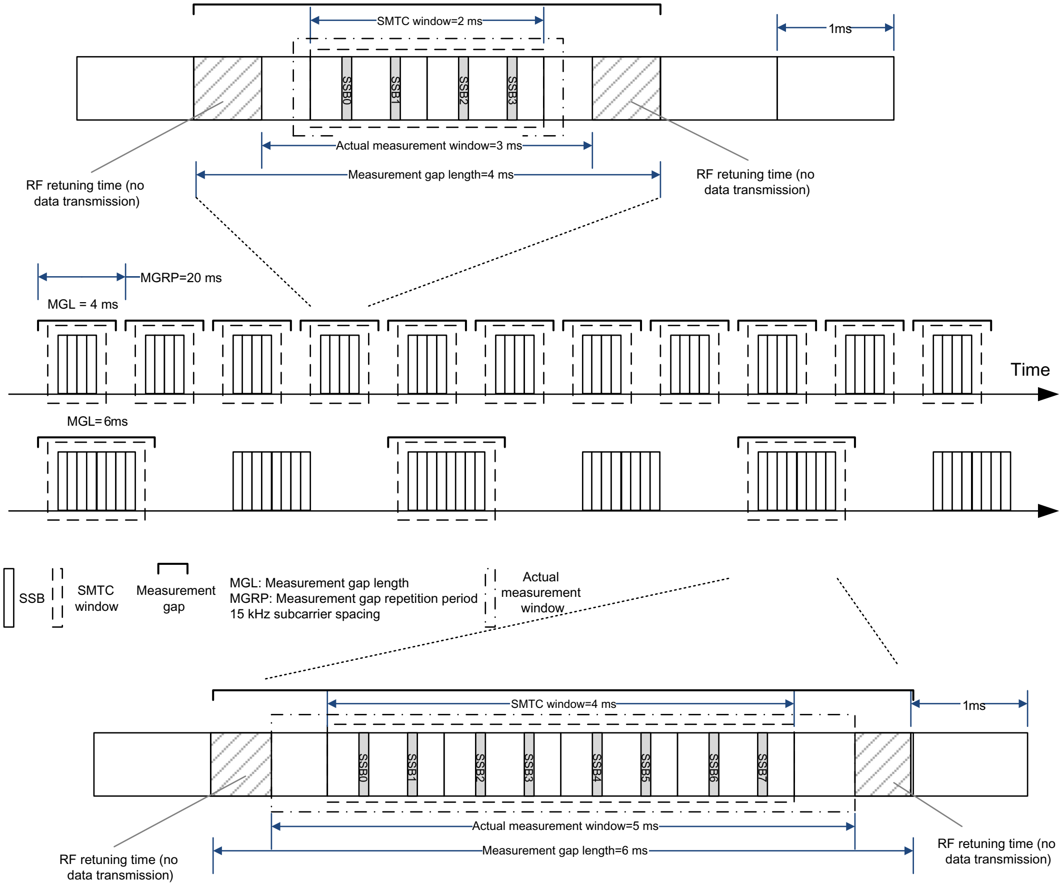

Figure 5.16
Illustration of the NR measurement gap configuration [29].

candidate cell and reports it to the serving gNB. The observed SFTD between an LTE PCell and an NR PSCell consists of two components as follows [7]:

- SFN_offset = $(\mathrm{SFN}_{PCell} - \mathrm{SFN}_{PSCell}) \bmod 1024$, where SFN_{PCell} is the SFN of an LTE PCell and SFN_{PSCell} is the SFN of the NR PSCell from which the UE receives the closest starting time to the time when it receives the start of the PCell radio frame.
- Frame_boundary_offset = $\lfloor (T_{Frame\_Boundary\_PCell} - T_{Frame\_Boundary\_PSCell})/5 \rfloor$, where $T_{Frame\_Boundary\_PCell}$ is the time when the UE receives the start of a radio frame from the PCell, $T_{Frame\_Boundary\_PSCell}$ is the time when the UE receives the start of the radio frame, from the PSCell, that is closest in time to the radio frame received from the PCell. The unit of $(T_{Frame\_Boundary\_PCell} - T_{Frame\_Boundary\_PSCell})$ is the NR sampling time.

The BS configures SFTD measurements for an NR UE neighbor cells. For a terminal to perform SFTD measurements on neighboring cells to which it is not connected, measurement gaps are required. However, if the serving gNB does not know the SSB transmission timing

Table 5.14: NR measurement gap pattern configurations [6].

Gap Pattern ID	Measurement Gap Length (ms)	Measurement Gap Repetition Period (ms)
0	6	40
1	6	80
2	3	40
3	3	80
4	6	20
5	6	160
6	4	20
7	4	40
8	4	80
9	4	160
10	3	20
11	3	160
12	5.5	20
13	5.5	40
14	5.5	80
15	5.5	160
16	3.5	20
17	3.5	40
18	3.5	80
19	3.5	160
20	1.5	20
21	1.5	40
22	1.5	80
23	1.5	160

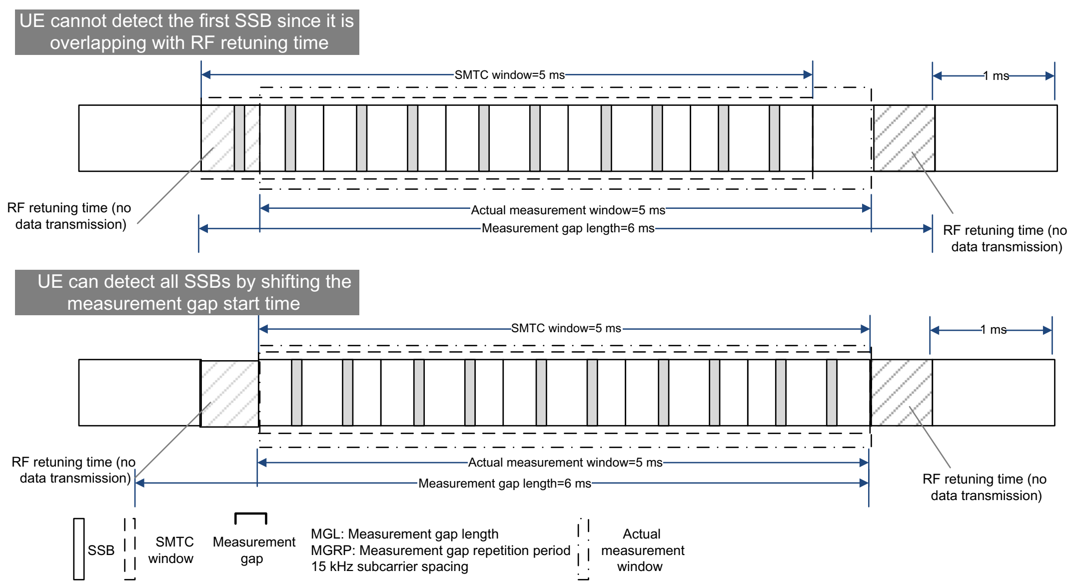

Figure 5.17
Illustration of the concept and usage of measurement gap timing advance [6,29].

of the candidate cell, it cannot configure such measurement gaps. In that case, the BS would need to configure the measurement gaps with different lengths until the terminal can detect the SSB of the candidate cell. This would increase the signaling overhead and the delay until the UE is able to find and connect to the target gNB. In order to avoid such inefficiency, a new procedure for performing SFTD measurement of candidate gNBs without using measurement gap is defined while maintaining UE's transmission/reception in the current cell. Fig. 5.18 shows an example of SFTD measurement which incorporates the aforementioned procedure. Instead of stopping transmission/reception of data by the RF device to perform measurements, the method uses separate RF device on the terminal, which is used if the UE supports carrier aggregation. When the RF transceiver that is used for measurement is turned on or off, the RF transceiver used for data transmission/reception is affected, thus the UE cannot transmit or receive in one subframe before and after the measurement window. However, the terminal can transmit/receive data in the current cell while SFTD measurements for neighboring gNBs are being performed.

5.5 Direct RF Sampling Transceivers

As analog-to-digital converter (ADC) and digital-to-analog converter (DAC) designs and architectures continue to advance using smaller geometry process nodes, a new class of giga-sample-per-second (GSPS) ADC and DAC products have begun to emerge. The ADCs/DACs that can directly sample the RF signal at gigahertz frequencies without the interleaving artifacts provide new solutions to systems for direct RF digitization of communications systems, instrumentation, radar applications, etc. The earlier solutions required multiple stages of filtering, synthesizers, and mixers to translate the input signal to a reference frequency that then could be digitized by an ADC at about 100-Msamples/s conversion rate. However, now direct RF sampling is achievable with state-of-the-art wideband ADC technology. It must be noted that the speed while important is not the only performance factor to consider in the designs as the dynamic range and spectral noise level are other important considerations.

In the past, the only monolithic ADC architectures that could run at GSPS speeds were flash converters with 6 or 8 bits of resolution. They were consuming excessive power and typically could not provide an effective number of bits (ENOB) beyond 7 bits due to the geometric size and power-constraint tradeoffs of flash architectures. The ENOB provides a measure of the performance of the ADC that is expressed in bits. ENOB is most accurately measured using a sine wave, curve-fitting method. The most common method for computing ENOB is to use the following equation based on the signal-to-noise-plus-distortion (SINAD)[9] at the full scale of the converter $\mathrm{ENOB} = (\mathrm{SINAD} - 1.76)/6.02$.

[9] The SINAD is the ratio of the rms signal amplitude to the rms value of the sum of all spectral components, including harmonics but excluding DC. The difference between SNR and SINAD is the energy contained in the first six harmonics.

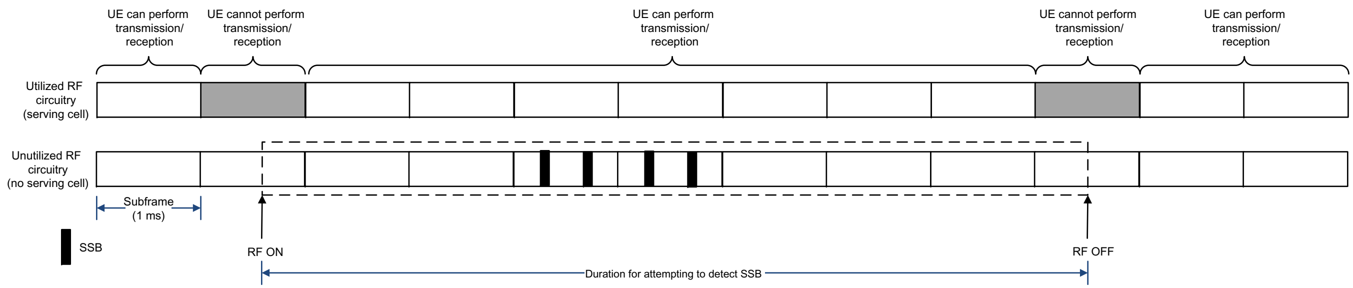

Figure 5.18
Example illustration of SFTD measurement without measurement gap [29]. *SFTD*, SFN and frame timing difference.

Therefore, the only way to sample higher dynamic range analog input signals above 1 GHz was to interleave multiple high-speed ADC cores with a sampling clock that had a staggered phase to each core with the required accuracy, or duty cycle. The analog input needed to be split and multiplexed to each ADC, which provided an opportunity for noise to enter the signal chain and reduce the input power. While this method may provide adequate results for some applications, the design was complex and yielded unwanted interleaving artifacts in the output frequency domain that needed digital filtering. The interleaving spurs can be seen in the frequency response of an FFT block, where the input offset, gain, bandwidth, and sample timing are not exactly matched across each of the internal interleaved ADC cores. This creates additional planning complexity for the system engineer to predetermine where interleaving artifacts will be seen in frequency and either avoid or remove them in digital postprocessing. Because each ADC core is discrete, the potential is high for manufacturing mismatch variance among these performance parameters during the life of a system in production. These mismatches cause imbalances in the periodicity of the incoming signal, and spurious frequencies are seen at the output of interleaved ADCs.

The new ADC technologies can take advantage of advanced architectures and algorithms that prevent the issues seen in dual and quad interleaved ADCs. Instead of using two interleaved ADCs at half speed, with added artifacts, the performance can be achieved in a single ADC at full speed without the interleaving spurs. Factory-trimmed algorithms and on-chip calibration ensure that each ADC operates to the expected high-performance standards, as opposed to being exposed to the mismatch variances seen from multiple discrete interleaved cores. When spurious frequencies are observed in an otherwise spectrally pure FFT, this reduces the available spurious free dynamic range (SFDR) of the carrier signal relative to other noise. To improve the SFDR of GSPS ADCs, new architectures and algorithms are now emerging beyond the use of interleaved cores. This removes the burden for system engineers to have dedicated ADC postprocessing routines that must identify and remove unwanted interleaving spurs (see Fig. 5.19).

The SFDR is defined as the ratio of the rms value of the signal to the rms value of the worst spurious signal regardless of where it falls in the frequency spectrum. The worst spur may or may not be a harmonic of the original signal. SFDR is an important specification in communications systems because it represents the smallest value of signal that can be distinguished from a large interfering signal (blocker). The SFDR can be specified with respect to full-scale (dBFS) or with respect to the actual signal amplitude (dBc). The definition of SFDR is shown graphically in Fig. 5.19. In other words, SFDR is the ratio of the rms value of the signal to the rms value of the peak spurious spectral component for the analog input that produces the worst result.

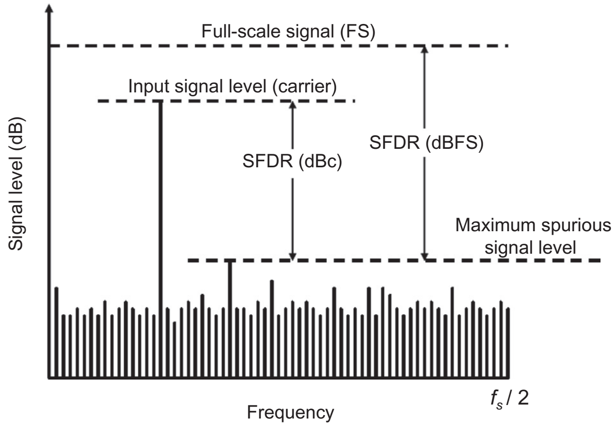

Figure 5.19
Illustration of SFDR concept.

Multi-GSPS converters with 10, 12, or 14-bit resolution generate multiple output data lanes. The use of low-voltage differential signaling (LVDS)[10] data could require 30 parallel lanes of 1-Gbps data for a 2.5 GSPS, 12-bit ADC, in which managing 30 differential LVDS pairs per ADC can be very challenging in terms of routing and maintaining matched lengths on a system layout. Equivalent data can be sent with only six or eight differential lanes using JESD204B/C,[11] a high-speed serializer/de-serializer (SERDES) standard designed specifically for converter interfaces. JESD204B/C provides a means to output data at high speeds on fewer data lines without the matched timing board complexities of many high-speed LVDS lanes. Since the data sent over JESD204B/C is framed, based on an embedded clock and control characters, the routing of the lower count serial lanes is much more tolerant of timing skew than LVDS. This eliminates the need to adjust output timing on every I/O of the system

[10] Low-voltage differential signaling (LVDS) is a high-speed digital interface that has become the solution for many applications that demand low-power consumption and high noise immunity for high data rates. Since its standardization under ANSI/TIA/EIA-644, LVDS has been implemented in a diverse set of applications and industries. The LVDS standard provides guidelines that define the electrical characteristics for the driver output and receiver input of an LVDS interface but does not define a specific communication protocol, required process technology, media, or voltage supply. The general, non-application-specific nature of the standard has been conducive to the adoption of LVDS across a wide variety of commercial and military applications.

[11] A standardized serial interface between data converters (ADCs and DACs) and logic devices (FPGAs or ASICs). JESD204B supports serial data rates up to 12.5 Gbps, a mechanism to achieve deterministic latency across the serial link, and uses 8B/10B encoding for SerDes synchronization, clock recovery, and DC balance. In Revision A, the main goal was to support both single and multiple lanes per convertor device. In Revision B, the added features were programmable deterministic latency, usage of device clock as main clock source, and data rate up to 12.5 Gbps. In the latest Revision C, the data rate is increased up to 32 Gbps, and three link layers have been introduced as 64B/66B, 64B/80B, and 8B/10B where 8B/10B link layer is same as JESD204B link layer. JESD204C defines layer-wise relationship with the IEEE Ethernet model.

board. Additionally, JESD204B offers informational control-bits of auxiliary data that can be appended to each analog sample to help characterize the downstream processing. In this manner, trigger time-stamping and over-range conditions can be tagged per sample so that a back-end FPGA can have further intelligence about data alignment and its validity.

The adaptive gain algorithms are important in terms of adjusting the amplitude of an analog input signal, since a saturated ADC input essentially makes the system blind in its ability to decipher signals. Ideally, the gain adaptation feedback loop should be as fast as possible. Whether the high-speed ADC output is LVDS-based or uses JESD204B/C, the added latency of this digital output often can be too long to wait to receive the saturated data, detect the issue, and react to the condition. One solution to this issue is to use a variable-level comparison within the ADC core itself and directly send an immediate output flag upon occurrence of an over-range condition. This technique bypasses the latency of the longer back-end output stage, which shortens the feedback time to the amplifier, allowing for a faster adaptive gain cycle. In addition to this fast, over-range detection output the over-range samples can be appended with alert bits, using the JESD204B interface, to let downstream system processing make appropriate decisions about the data.

A wideband ADC not only offers the benefits of broadband sampling, but also may provide more data than needed in some applications. For high-sample-rate systems that do not need to observe a large frequency spectrum, digital down-conversion (DDC) allows a subsampling and filtering strategy for decimating the amount of data output from the GSPS ADC. Downstream processing then observes a smaller portion of the frequency spectrum. The DDCs are often implemented after the ADC in the signal chain. This not only consumes more resources in an FPGA but also requires the full bandwidth to be transmitted between the ADC and FPGA. Instead of transmitting and processing the sampled data in an FPGA, the DDC filtering can be done within the ADC to see just one-eighth or one-sixteenth of the total bandwidth. When used in conjunction with a synthesized numerically controlled oscillator (NCO), the precise placement of the converter's DDC filter in the band can be tuned with accurate resolution. This permits a lower output rate and eliminates the need to move and process large amounts of unwanted data in an FPGA. When two DDCs are available, each with a unique NCO, they can alternately be stepped across the spectrum to sweep for expected signals, without loss of visibility. This is often typical in some radar applications.

5.5.1 Nyquist Zones and Sampling of Wideband Signals

In the theory of sampling, a signal must be sampled at more than twice the maximum frequency $f_{\max}$ in the spectrum of the signal, in order to be able to regenerate the signal from its sampled values. This sampling frequency is known as the Nyquist rate $f_s \geq 2f_{\max}$. If the signal is sampled below the Nyquist rate, aliasing occurs. As a general rule, for a sample rate of f_s, one would not be able to differentiate between any sine waves with frequencies

$Nf_s/2 \pm \Delta f$, where N is an integer. In this case, the signal power for these frequencies will concentrate at $f_s/2 - \Delta f$. As long as, for each Δf, there is only a signal at one of the aliasing frequencies, the signal has not been irreversibly corrupted. That means the signal can still be reconstructed, if one knows which of the frequencies that alias occurs $f_s/2 \pm \Delta f$ were occupied by the desired signal. In this case, the aliasing decisively can be useful, which leads to the concept of the Nyquist zones. However, if the desired signal has significant power at both $f_s/2 - \Delta f$ and $f_s/2 + \Delta f$ frequencies, then aliasing occurs and you no longer have the information to distinguish these two components since these formerly distinct signal frequencies now appear as a single frequency, thus it is not possible to determine the power of each component. For a sampling frequency f_s, various zones exist where, if a signal is band-limited to that zone, the original spectrum can be recovered. As shown in Fig. 5.20, at the sampling frequency f_s, one can define zones where a band-limited signal in a higher Nyquist zone will alias down into the first Nyquist zone. For even-numbered Nyquist zones, the spectrum will appear in reversed order, and for odd-numbered zones, it appears in the original order. This means, for example, a 100 MHz band ranging from 700 to 800 MHz need only be sampled at 100 MHz to be characterized. As long as you know which Nyquist zone is occupied, you have all information to reconstruct the original signal. One issue is that the boundaries of the Nyquist zones are fundamentally tied to the sampling rate, which implies that the bandwidth of the desired signal must fit within a Nyquist zone.

According to Poisson summation formula,[12] the samples of the Fourier transform of function $v(t)$ are sufficient to create a periodic extension of $V(f)$ as $V(f) = \sum_{k=-\infty}^{\infty} V(f - kf_s)$. The latter equation can be interpreted as the sum of the replicas of $V(f)$ shifted to integer multiples of f_s (see Fig. 5.20). Fig. 5.20 also illustrates an example of the Nyquist frequency $f_s/2$. Since these signals are repeated at multiples of f_s, then for bandwidths greater than $f_s/2$ aliasing and information loss will occur. The integer multiples of the Nyquist frequency determines the Nyquist Zones. In the literature, the Nyquist Zones are only defined for the positive frequency spectrum.

In other words, the FFT of a discrete-time signal can be divided into an infinite number of $f_s/2$ frequency bands, also known as Nyquist zones. The frequency spectrum between DC (zero frequency) and $f_s/2$ is known as the first Nyquist zone. The frequency spectrum repeats itself over different Nyquist zones. Note that the even-numbered Nyquist zones appear as mirror images of the odd-numbered Nyquist zones. Aliasing in ADCs is a ramification of the sample-and-hold processing of the analog signal at the input stage. In the

[12] In mathematics, the Poisson summation formula is an equation that relates the Fourier series coefficients of the periodic summation of a function to values of the function's continuous Fourier transform. Consequently, the periodic summation of a function is completely defined by discrete samples of the original function's Fourier transform. Conversely, the periodic summation of a function's Fourier transform is completely defined by discrete samples of the original function.

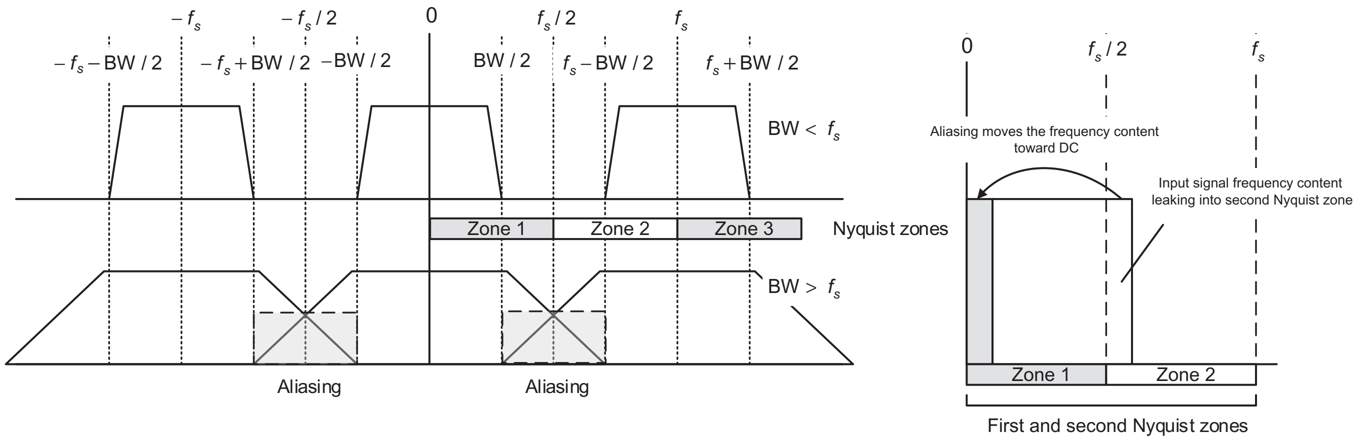

Figure 5.20

Illustration of the Nyquist zone concept [15].

digital signal processing domain, sample-and-hold processing is equivalent to convolution of the frequency spectrum of the impulse train (due to the sampling clock) with the frequency spectrum of the analog input. This convolution results in periodicity of the frequency spectrum that is observed over various Nyquist zones, as previously explained. When the input signal contains frequency components above the Nyquist frequency ($f_s/2$), the adjacent Nyquist zones start overlapping and result in aliasing. Aliasing in DACs is a consequence of the zero-order-hold processing (used to avoid code-dependent output glitches) of the discrete-time samples at the output stage. The zero-order-hold processing in the DSP domain is equivalent to convolution of the $\sin(x)/x$ type of frequency spectrum (of the rectangular function appearing due to holding discrete-time samples) with the DAC core's output-impulse train frequency spectrum (of varying amplitude, in general). As in the case of ADCs, the periodicity of the output spectrum over different Nyquist zones can be attributed to this convolution (see Fig. 5.21).

Data converters are divided into two main categories, namely, Nyquist rate and oversampling. In the Nyquist rate type, the input occupies a large fraction of the available bandwidth (Nyquist zone), whereas in the oversampling type, the input occupies only a small fraction of the Nyquist zone, simplifying the antialiasing filter (AAF) design and lowering the quantization noise at the expense of higher sampling rate (see Fig. 5.22).

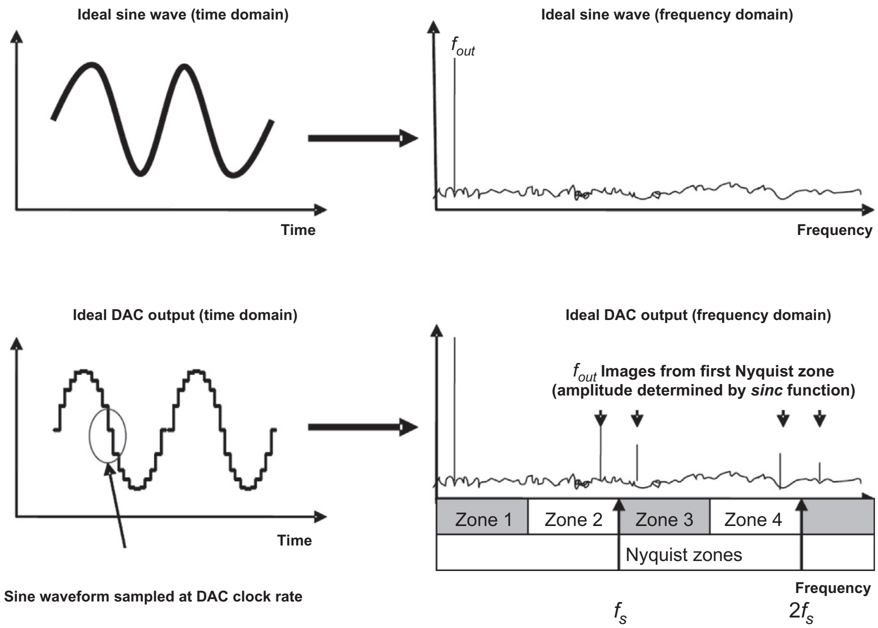

Figure 5.21
Example illustration of Nyquist zones [15].

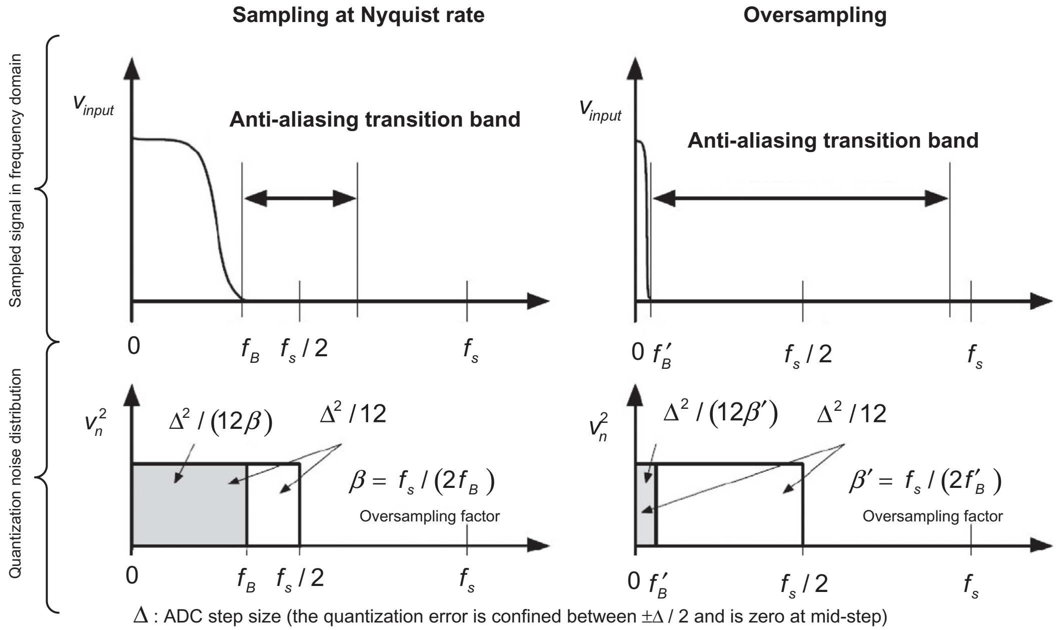

Figure 5.22
Illustration of Nyquist rate and oversampling data conversion [15].

5.5.2 Direct RF Sampling Transmitter Design

Direct digital domain to RF domain conversion with a high-speed RF DAC is a technology disruption for BS transmitters. The RF DAC partition uses direct digital synthesis to move the quadrature modulator, local oscillator (LO), and signal filtering analog functions into the digital domain (see Fig. 5.23). The RF DAC with direct digital synthesis partition capitalizes on the fact that digital processors scale better than their analog counterparts in terms of lower power consumption, faster speed, smaller die area, and lower cost. The RF DAC is the enabling technology that makes this possible because it bridges the digital-to-analog domains. An RF DAC is generally characterized as a mixed signal device that operates in multiple Nyquist zones with conversion rates above 1.5 GSPS to perform direct digital-to-RF signal synthesis. An RF DAC synthesizes output signals of at least 500 MHz signal bandwidth at carrier frequencies of 2.0 GHz or higher. Compared to conventional RF transmitter architectures such as zero-intermediate-frequency (ZIF), complex-IF and real-IF, the RF DAC architecture solution occupies less area with fewer components. It operates at lower power and delivers excellent dynamic performance.

In terms of RF performance, the RF DAC has significant benefits over other architectures. The digital up-conversion with digital filtering implemented in direct digital synthesis eliminates gain-phase errors and achieves perfect carrier suppression with no LO leakage. The result is excellent EVM performance when transmitting high-order modulations like

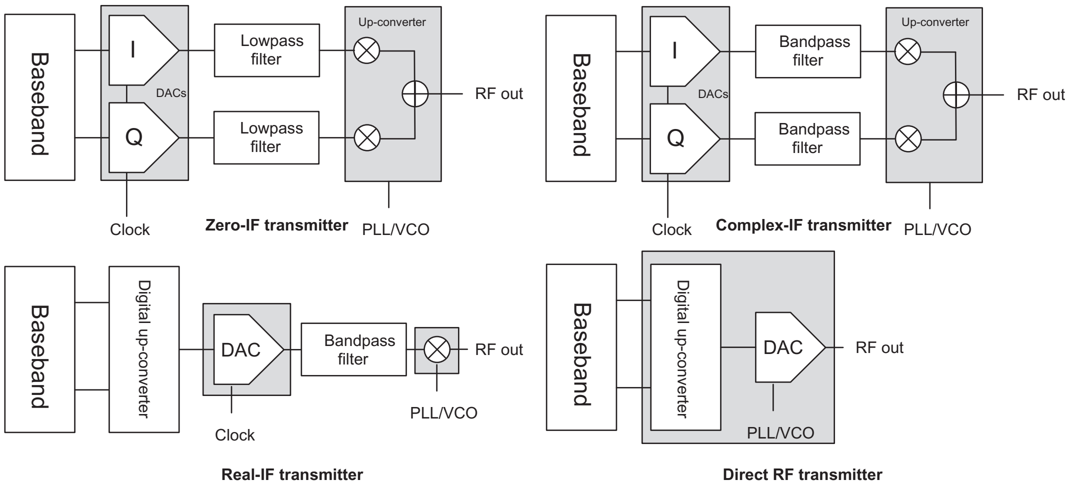

Figure 5.23
Illustration of RF transmitter architecture options [16].

64QAM. The quadrature NCO makes the RF DAC an agile transmitter capable of tuning across the entire spectrum of LTE/NR sub-6 GHz bands (and mmWave bands using an external mixer and filter). Since the RF DAC is broadband and frequency agile with high dynamic range, a single device can synthesize multi-carrier, multi-band, and multi-standard signals including GSM, WDMA, LTE and NR. As a result, a fully digital software-defined radio can be realized using common hardware across multiple BS hardware platforms [16].

Another advantage of direct-conversion RF DAC technology is enabling lower cost digital predistortion (DPD) observation receiver. Macro-cell base stations typically use DPD techniques for RF PA linearization. This requires a PA observation receiver channel to monitor the PA output. The observation receiver detects PA distortion products and works with a predistortion block to compensate for intermodulation and adjacent channel leakage power. The DPD bandwidth expansion requires the DPD observation receiver bandwidth to be five times the data bandwidth. As an example, in 100 MHz carrier aggregation scenarios, this means that the DPD bandwidth must be at least 500 MHz. Also, the observation receiver cannot add impairments to the observed signal because they cannot be discerned from the main transmitter (TX) path impairments. Consequently, the DPD observation path must have excellent linearity which adds cost and complexity. Conversely, if the main TX path has negligible impairments, then the DPD path impairments can be corrected. Since the RF DAC does not introduce gain or phase errors, there are negligible TX path impairments. Therefore, a low-cost and lower performance DPD receiver such as a ZIF receiver can be employed. There are three reasons that the ZIF architecture has lower cost compared to high-IF or direct RF sampling: (1) quadrature demodulation enables a lower conversion rate, baseband sampling, dual-channel high-speed ADC because it only needs to quantize one-half of the DPD expansion bandwidth; (2) the ADC samples baseband signals, which means that the ADC does not need pico- or femtosecond aperture jitter;[13] and (3) the baseband I/Q anti-alias filters are lower cost and easier to design compared to IF or RF filters used in high-IF or direct RF architectures. In summary, the RF DAC transmitter relaxes DPD receiver signal-path performance requirements, thus further reducing system cost and design complexity [16].

5.5.3 Direct RF Sampling Receiver Design

In a direct RF-sampling architecture the data converter digitizes a large segment of spectrum directly at RF frequencies and transfers the information to a digital signal processor. This is a paradigm shift that takes what has traditionally been handled by analog components (i.e., mixers, LOs and their associated filters and amplifiers) into the digital domain. A new class of direct RF-sampling ADCs is being designed in advanced complementary metal-oxide semiconductor (CMOS) processes that allows much higher conversion rates with lower power relative to the previous generations. Furthermore, this approach enables higher integration, which is used for a low-power, multi-giga-bit serial interface, and on-chip DDC, leading to very size- and power-efficient digital interconnect between the data converter and digital processor. As shown in Fig. 5.24, in a super-heterodyne receiver, the input signal, which resides at an RF frequency, is down-converted to a lower IF frequency. It is then digitized prior to digital filtering and demodulation. Depending on the application, the input signal can range from 700 MHz to several GHz, while IF typically ranges from zero to 500 MHz. The classic super-heterodyne receiver consists of bandpass filters (BPFs), a LNA, mixer and LO, an IF amplifier, and ADC AAF [15,17].

In a direct-conversion receiver, the RF-sampling ADC replaces the signal chain from the mixer, which greatly simplifies the overall receiver design. The first RF BPF is typically a

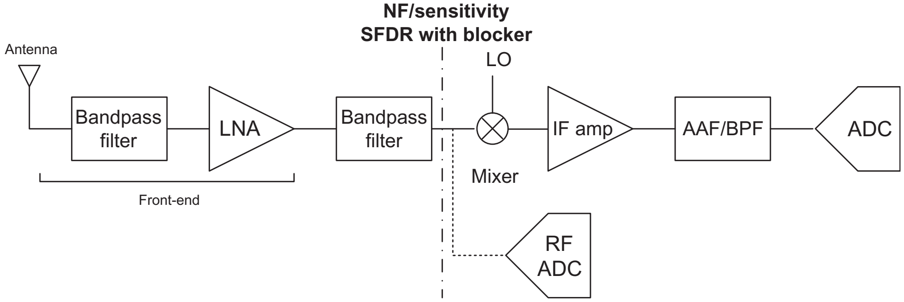

Figure 5.24
Performance comparison between super-heterodyne and direct RF-sampling receivers [15].

[13] Aperture jitter or aperture uncertainty is the sample-to-sample variation in aperture delay that can be manifested as frequency-dependent noise on the ADC input. Aperture delay is a measure of the difference in the delay between the analog path and the encode path. It is measured by observing the time from the 50% point of the rising edge of the sample clock to the time at which the input signal is actually sampled.

wideband, preselect filter, which provides most of the OOB rejection. It prevents signals that are far from the actual passband from saturating the analog front-end (AFE). The LNA increases the amplitude of weak signals. The lower its NF, the less it degrades the overall NF of the receiver based on the cascaded NF concept. The second RF filter is a narrowband filter prior to the mixer typically a surface acoustic wave (SAW) type. It suppresses nearby OOB interferers at the mixer image locations ($mf_{LO} \pm nf_{RF}$) that would fall into the IF passband. The mixer translates the RF signal to IF frequencies. The RF and LO signals mix to produce a difference frequency known as IF frequency ($f_{IF} = |f_{RF} - f_{LO}|$). The undesired spurs and images (e.g., a half-IF) need to be filtered out either before or after the mixer stage. It also can be used to convert the single-ended input signal to a differential signal for the ADC. The LO is tuned to the desired frequency spacing above or below the RF signal and is injected into the mixer.

In a direct RF-sampling receiver, the LO essentially turns into the ADC's sampling clock. However, similar to the IF-sampling converter, the RF ADC clock also requires very good phase noise properties. The IF amplifier adds gain to the input signal by reducing the ADC's impact to the receiver NF. Moreover, it compensates for the BPF attenuation. It can also be used as a buffer to drive the ADC's capacitive load. In many cases, this stage has multiple gain steps that are digitally controlled to provide an automated gain control loop, which enhances the system's dynamic range. The anti-aliasing filter limits noise and distortion contribution from the IF amplifier. More importantly, it filters the ADC's alias bands. This filter may need a sharp roll-off. The implementation is more feasible at IF than RF frequencies because the ratio of alias frequency to the desired signal is larger. Thus aliasing frequencies are further away from the filter cutoff frequency. The ADC digitizes the input signal. It needs a fast sampling rate to accommodate the filter roll-offs, typically at least three to five times the signal bandwidth. Its SNR needs to be good to ensure minimal impact on the receiver noise figure. The SFDR must be sufficient so that spurs caused by in-band and out-of-band interferers do not dominate the overall noise level.

When evaluating a transition from an IF-sampling-based to a direct RF-sampling-based receiver design, one must examine the overall impact on the receiver sensitivity, as well as performance in a blocking condition. Since the RF-sampling ADC replaces the signal chain from the mixer onward, the comparisons provided are focused at the mixer input, assuming the same AFE for both designs (see Fig. 5.24). The parameters also can be calculated at the antenna input for a given LNA in the same manner using the respective gains and losses of the amplifier and filters. The receiver sensitivity is a measure of how it can recover and process very weak input signals. Weak input signals cannot be demodulated, if the receiver noise within the demodulated bandwidth is larger than the received signal itself. The transmitter and receiver are often completely independent, thus raising the desired signal amplitude above the noise floor is not always possible. The only option to improve receiver sensitivity is to reduce its noise floor or, in other words, to improve its noise figure. In the sensitivity condition, the receiver operates at the maximum gain. In the presence of a strong blocker/interferer the gain needs to be reduced in order to avoid saturation of the input (desensitization). Therefore, the receiver NF will be much stronger, which impacts the minimum signal that can be recovered in a blocking condition.

The images from the mixer at $mf_{LO} \pm nf_{RF}$ can fall into the band of interest that requires adequate filtering at RF level. This ensures that an OOB blocker at the image locations does not generate a mixing product within the desired band. Furthermore, the phase noise of the LO frequency is also mixed with the RF signal (interferer). This increases the noise power with in-band blockers, increasing the amount of noise leaking from the blocker into the wanted carrier bandwidth. The IF gain amplifier generates second- and third-order harmonic distortion (HD2 and HD3) products from the strong interferer, which can fall in the signal band of interest. These low-order harmonics are typically either frequency controlled by choosing an appropriate IF frequency or attenuated by the following BPF.

Several impairments originate from the data converter. Similar to the IF amplifier, the ADC generates strong low-order harmonic distortions (HD2 and HD3) from the blocker. The IF frequency can be chosen such that these harmonics are located out-of-band for an in-band blocker. For an OOB blocker, adequate external filtering should be provided. High-order harmonics set a spur floor that cannot be controlled and needs to be taken into consideration. The phase noise of the sampling clock (jitter) also mixes with the interferer. It scales with $20\log(f_S/f_{IN})$, so the higher the input frequency range (IF) with a fixed ADC sampling rate, the larger its impact on the overall system noise. The RF-sampling ADC may achieve its fast clock rate using interleaving techniques. In this case, the interleaving spurs need to be considered not only for the noise analysis with in-band blockers but also for the filter design for out-of-band interferers.

5.6 Considerations for Implementation and Operation in mmWave Bands

5.6.1 Semiconductor Technologies for mmWave Operation

The mmWave RF analog implementations have been historically dominated by III−V[14] monolithic microwave integrated circuit (MMIC)[15] semiconductor technologies which primarily use

[14] A substance that can act as an electrical conductor or insulator depending on chemical alterations or external conditions. Examples include silicon, germanium, and gallium arsenide. It is called III−V materials since semiconductor elements are in groups III and V of the periodic table of chemical elements.

[15] An MMIC is a type of integrated circuit device that operates in microwave frequencies. These devices typically perform functions such as mixing, power amplification, low-noise amplification, and high-frequency switching. Inputs and outputs on MMIC devices are matched to a characteristic impedance of 50 ohms.

gallium arsenide (GaAs).[16] These are ideally suited for the RF front-ends of mmWave systems such as PAs and LNAs, as well as enabling oscillators with excellent phase noise characteristics. More recently, SiGe-heterojunction bipolar transistor (HBT)-based technologies have attracted increasing interest for emerging mmWave markets, as f_T or $f_{\max}$[17] of the HBT devices has exceeded 200 GHz. The performance of SiGe-HBT[18] is no longer the limiting factor for a mmWave transceiver front-end integration for small-cell scenarios with limited output power, rather the quality factors of the on-chip passive devices, such as inductor, capacitor, and transmission lines for matching and tuning and their accurate characterization in the mmWave frequency domain, are the limiting factors. However, the latest products in the market demonstrate sufficient quality, which is a trade-off between performance and cost. The E-band backhaul for macro-cells usually require support of high-order QAM modulation, which can be achieved with GaAs components or a combination of SiGe/BiCMOS[19] transmitter/receiver and GaAs PAs, LNAs, and voltage-controlled oscillators (VCOs) [23].

The CMOS[20] technology promises higher levels of integration at reduced cost, if volumes scale to multi-million parts per year. Several recent developments, particularly the need for the availability of chipsets operating in mmWave frequencies, have increased efforts to enable CMOS circuit blocks to operate in mmWave bands, as the CMOS transistor f_T approaches to 400 GHz. However, the performance for point-to-point links is worse than that achieved by SiGe[21] or GaAs components in terms of phase noise and NF for the same distances typically greater than 100 m [23].

[16] Gallium arsenide (GaAs) is a compound of the elements gallium and arsenic. It is a III−V direct bandgap semiconductor with a zinc blende crystal structure. GaAs is used in the manufacturing of devices such as microwave frequency integrated circuits, monolithic microwave integrated circuits, and infrared light-emitting diodes. GaAs is often used as a substrate material for the epitaxial growth of other III−V semiconductors including indium GaAs, aluminum GaAs, and others.

[17] The transit frequency is a measure of the intrinsic speed of a transistor and is defined as the frequency where the current gain reduces to one.

[18] HBT, a type of bipolar junction transistor (BJT) which uses differing semiconductor materials for the emitter and base regions, creating a heterojunction.

[19] BiCMOS is an evolved semiconductor technology that integrates two formerly separate semiconductor technologies, those of the BJT and the CMOS transistor, in a single integrated circuit device.

[20] CMOS is a technology for manufacturing integrated circuits. CMOS technology is used in microprocessors, microcontrollers, etc., as well as analog circuits such as image sensors and data converters. Two important characteristics of CMOS devices are high noise immunity and low static power consumption. Since one transistor of the pair is always off, the combination draws significant power only momentarily during switching between ON and OFF states. Consequently, CMOS devices do not dissipate heat as other forms of logic. CMOS also allows a high density of logic functions on a chip. It was primarily for this reason that CMOS became the most used technology to implement very-large-scale integration chips.

[21] Silicon germanium (SiGe) is an alloy with any molar ratio of silicon and germanium, which is commonly used as a semiconductor material in integrated circuits for HBTs or as a strain-inducing layer for CMOS transistors. This relatively new technology offers opportunities in mixed-signal and analog integrated circuit design and manufacturing. SiGe is also used as a thermoelectric material for high temperature applications process.

Fully depleted/partially depleted silicon on insulator (RF-SOI/FD-SOI) is a semiconductor technology which realizes fully/partially depleted CMOS devices on SOI and combines the flat structure of 2D CMOS transistors with the fully depleted mode of SOI. The SOI has a good power-performance-area-cost (PPAC) value. It uses a unique type of substrate material whose thickness is controlled to atomic scale and could provide excellent transistor performance. The FD-SOI technology can operate in mmWave frequencies, thus it is gaining more attention in practical 5G mmWave applications.

Packaging processes which are available and in production for RF analog components are Quad Flat No-leads (QFN) Package, embedded Wafer-Level-Ball (eWLB) Grid Array Package,[22] and Flip Chip.[23] However, the existing mmWave RF components are mostly bare die or modules. The eWLB examples with System in Package (SiP) demonstrate that assembly and packaging using the eWLB technology offer outstanding system integration capabilities. This includes the integration of different chips and the design of integrated passive components such as resistors, inductors, and transformers either in the redistribution layer (RDL)[24] or using through encapsulate via (TEV).[25] Antennas can also be integrated into the package. Other technologies where silicon wafer level technology and back-end merge are through silicon via (TSV)[26] and die embedding in laminate technologies. TSV technologies are typically combined with RDLs, for example, for silicon interposer. However, a major obstacle against their broad adoption is cost [23].

The evolution of RF analog components integration for mmWave applications depends on several factors including the following:

- Permissible output power and EIRP of the system including the antennas
- Phase noise required for standard modulation schemes (BPSK, QPSK, 16QAM, ..., 256QAM, etc.)
- Noise figure
- Power consumption
- Form factor and cost

[22] eWLB grid array is a packaging technology for integrated circuits. The package interconnects are applied on an artificial wafer made of silicon chips and a casting compound.

[23] Flip chip is a method for interconnecting semiconductor devices to external circuitry with solder bumps that have been deposited onto the chip pads. The solder bumps are deposited on the chip pads on the top side of the wafer during the final wafer processing step. In order to mount the chip to external circuitry, it is flipped over so that its top side faces down and aligned so that its pads align with matching pads on the external circuit, and then the solder is reflowed to complete the interconnect. This is in contrast to wire bonding, in which the chip is mounted upright and wires are used to interconnect the chip pads to external circuitry.

[24] An RDL is an extra metal layer on a chip that makes the I/O pads of an integrated circuit available in other locations. When an integrated circuit is manufactured, it usually has a set of I/O pads that are wire-bonded to the pins of the package.

[25] TEV technology is an interconnect technology process.

[26] TSV technology is interconnect technology process.

Silicon transistors cannot compete with III−V compounds (GaAs, InP, GaN) for low-noise performance, linearity, and output power at frequencies above 20 GHz. A GaAs mmWave LNA yields an average NF of 2.5 dB, which is much lower than state-of-the art SiGe LNA of 5 dB. The output power levels (P_{sat}) of over 30 dBm can be achieved with GaAs in E-Band, while SiGe-HBTs can reach 19 dBm (P_{sat}) [23].

Silicon RFICs allow the integration of multiple application-specific functionalities on a single silicon chip (RF application-specific integrated circuit or ASIC) with excellent yield and uniformity plus the possibility to integrate different calibration schemes to mitigate RF impairments, which are not possible or more complex to implement in GaAs. The level of integration is a factor to be considered. A high level of integration makes the chip very specific and could increase development time in a first design but reduces production test and simplifies module assembly. A good compromise for high-end applications (e.g., E-band high power, 256QAM) is to use compound semiconductors for the front-ends (LNA of the receiver input and PA of the transmitter output) and silicon semiconductors for the lower frequency mixed signal functions and control/digital elements. The III−V compound devices can realize systems to expand the use of the electromagnetic spectrum above 90 GHz. However, improvements in the high-frequency capability of CMOS/BiCMOS technology have made it possible to consider it as a low-cost, lower performance alternative to III−V compound devices.

5.6.2 Analog Front-End Design Considerations

The baseband modems utilize digital signal processing to perform channel coding and complex-valued modulation and demodulation in order to transmit extremely high data rate streams over RF carriers in mmWave bands. For backhaul applications with data rates ranging from 1 to 10 Gbps, the systems typically utilize channel bandwidths of 250 and 500 MHz which have recently increased to 1 and 2 GHz to support integrated access and backhaul applications. The baseband modulation schemes, generating complex-valued symbols, range from BPSK, QPSK, 16QAM and up to 256QAM (or higher) for high-performance systems. Therefore, it is necessary to perform real-time I/Q ADC (receive) and I/Q DAC (transmit) at the interface between the analog and digital domains. Moreover, such data conversions require low jitter baseband clocks in order to minimize inter-symbol interference. The combination of I/Q ADC, I/Q DAC, and baseband phase-locked loop (PLL)[27] is known as AFE as shown in Fig. 5.25. The performance of the AFE has a significant impact on the overall performance of the modem in terms of the receiver dynamic range, transmitter spurious emissions, and end-to-end bit error rate/packet error rate (BER/PER).

[27] A PLL is a closed-loop frequency-control system based on the phase difference between the input clock signal and the feedback clock signal of a controlled oscillator. The main blocks of the PLL are the phase-frequency detector, loop filter, voltage-controlled oscillator, and counters, such as a feedback counter, a prescale counter, and postscale counters.

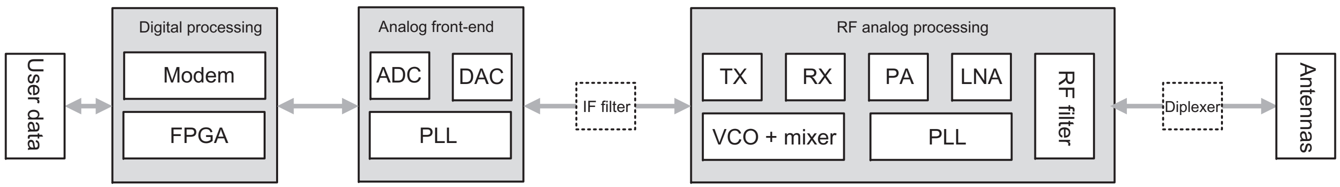

Figure 5.25
High-level system architecture of point-to-point mmWave radio [23].

A quadrature I/Q ADC is required in order to sample the IF (or baseband in case of zero-IF receiver) I/Q analog signals down-converted from the mmWave receiver. Typical channel bandwidths range from 250/500 MHz for traditional backhaul systems to 1 GHz or higher for the 5G systems, thus the sample rates would be in the excess of 500 MHz depending on the required over-sample ratio for a given modem implementation. The required resolution or ENOB is primarily a function of modulation scheme which ranges from 4 bits for QPSK to 5 bits for 16QAM, and 6 bits for 64QAM. There is a need for additional dynamic range to support OFDM waveform compared to single-carrier waveform due to power summation over multiple subcarriers (crest factor). This requires an additional 5−6 dB dynamic range for OFDM waveform compared to single-carrier waveform for a given modulation index. In order to determine the dynamic range of the ADC for a given modulation scheme and receiver, the noise floor of ADC and the receiver noise must be carefully modeled. The noise floor of the ADC is defined by the power level of the (quantization) noise produced by the ADC itself when referred back to the receiver input, that is, the actual ADC noise power divided by the gain of the front-end. Conventionally, this is set to a level approximately 10 dB below the total effective noise power at the input to the receiver so that the effect of the ADC is to increase the front-end NF by 0.5 dB. The receiver noise power is just thermal noise power (KTB) raised by the receiver NF in dB [23].

In addition to the receiver noise, a margin must be added for the minimum SNR requirement at the input to the demodulator for a given modulation and coding scheme. This must be followed by an additional margin related to the crest factor of the signal. A receiver extra power margin is defined, which is the amount by which the average signal power is allowed to increase before any gain reduction needs to be applied by the receiver automatic gain control (AGC). The receiver extra power margin determines the minimum AGC step requirement for any given mode. The larger the power margin, the larger the value that can be assigned to the AGC step size. Therefore, the ADC dynamic range is given by $\mathrm{DR}_{ADC} = 9.636 + \mathrm{SNR} + \mathrm{CF} + \mathrm{PH}\ (\mathrm{dB})$, where SNR is the minimum signal-to-noise ratio required at the demodulator, *CF* is the crest factor, and *PH* is the ADC power headroom. A typical dynamic range based on an example QPSK modulation would be approximately 36 dB, if the ADC power headroom is restricted to 14 dB. This analysis may be extended to higher order modulations in order to determine the optimum ADC dynamic range and resolution to support 16QAM, 64QAM, or 256QAM. The above ADC dynamic range is related to ENOB based on the following equation $\mathrm{ENOB} = \left[\mathrm{DR}_{ADC} - 10\log_{10}\left(1.5 f_s/\mathrm{BW}_{channel}\right)\right]/6\ (\mathrm{Bits})$, where f_s and $\mathrm{BW}_{channel}$ are the sampling frequency and channel bandwidth, respectively. In general, we can observe that doubling of I or Q resolution (i.e., $4\times$ in modulation domain) requires an additional 6 dB or one additional bit of ENOB.

A quadrature I/Q DAC is required to generate the baseband I/Q analog signals for up-conversion, filtering, and amplification in the mmWave transmitter. The typical bandwidths and sample rates are the same as those needed for the I/Q ADC. However, the key technical parameter for the transmitter is related to the spurious emission requirement. It can be shown that a DAC resolution of 7 bits or higher are typically required. The dynamic range requirement of the DACs is largely determined by the spurious emission requirement of the transmitter. The dynamic range requirement is also dependent on the power margin necessary to cope with both gain variations in the mmWave transmitter chain as well as the digital power control. The DAC dynamic range requirement can be calculated as $\mathrm{DR}_{DAC} = -S_p - L_{IF} + \mathrm{CF} + \mathrm{PH}$ (dB), where S_p is the spurious emissions requirement, L_{IF} is the IF filter attenuation, CF is the crest factor of the wanted signal, and PH is the required DAC power headroom. This equation applies to all modulation and coding schemes, thus if the power headroom is fixed, the dynamic range becomes dependent on only the crest factor.

A baseband PLL is required to generate the local I/Q sampling clocks required for the I/Q ADC and DAC. Depending on the channel bandwidth, the sample clock frequency can range from 500 MHz to 3 GHz or higher. The phase noise of this baseband clock may have an impact on the overall system performance. To evaluate this effect, consider the single-sideband phase noise characteristics of an example commercial ADC clock PLL with a 40 MHz crystal reference, which achieves an rms phase noise of 1.57 degrees in the bandwidth ranging from 100 Hz to 4.5 GHz. This phase noise can be translated into an equivalent SNR value of 31.24 dB in the Nyquist band between $\pm f_s/2$ using the following equation $\mathrm{SNR}_{\theta dB} = 10\log_{10}(180/\pi\theta)$ where f_s is the sampling frequency. The required phase noise performance of the ADC clock PLL can be further estimated from an analysis of the receiver SNR at reference sensitivity for all modulation schemes and later defining the acceptable additional reduction in SNR that could be tolerated as a result of the additional phase noise contribution of the sampling clock. If the additional reduction of SNR is set to 0.1 dB, it can be shown that the corresponding rms phase noise of the PLL would be 2.08 degrees and the rms time jitter for a sample clock of 2.64 GHz would be 2.2 ps. Therefore, a reasonable jitter requirement for the baseband PLL should be less than 2 ps [23].

In order to evaluate the performance impact of high-frequency operation on data converters, a figure of merit (FoM) is defined, which combines several performance metrics into a single value. For ADCs the construction of a fair FoM is non-trivial, not only because an ADC is rather a complex device with many attributes but also because the exact trade-off between the relevant metrics is often non-obvious. Most of the commonly used FoMs for ADCs share two basic attributes: (1) they are purposefully kept simple and mainly focus on power dissipation, conversion rate, and resolution and (2) they are constructed using first-order statistics, limited tradeoffs and avoid empirically derived parameters as much as possible. More complex FoMs (considering area, supply voltage, process technology, etc.) have been proposed in the literature [22]; nevertheless, their empirical and data-driven nature makes them difficult to use. Ultimately, it is important to realize that an FoM is not meant to capture every aspect of a specific data converter. A direct comparison between two converters should always be made using a table, where the FoM is just one of several entries. One of the widely used FoMs for quantifying the trade-off between ADC speed, resolution, and power dissipation is Walden FoM ($\text{FoM} = P2^{-ENOB}/f_s$). This FoM asserts that power P increases linearly with conversion rate f_s. This is justified from first-order physics (power = energy/time) but it tends to fall apart as the sampling speed approaches the technology limits. In terms of resolution, this FoM assumes that each added effective bit (ENOB) doubles the power dissipation. This trade-off was chosen empirically and tended to match experimental outcomes reasonably well for a number of years. In recent years, however, it has been observed that moderate-to-high-resolution converters tend to follow a $4\times$ power per effective bit trend, a trade-off that is consistent with noise-limited analog circuits. To capture the $4\times$ per bit trade-off, one can simply modify the denominator of the latter FoM expression and replace 2^{-ENOB} with 2^{-2ENOB}. However, a different way of formulating the same trade-off in logarithmic scale results in Schreier FoM ($\text{FoM} = \text{SNDR(dB)} + 10\log\left(f_s/2P\right)$) where in its original form takes the dynamic range DR of the data converter in dB and adds 10 times the logarithm of the bandwidth BW to power ratio. Thus a 6 dB increase in dynamic range is assumed to reflect the $4\times$ increase in power at the same efficiency. Also note that the proper unit of this FoM is dB/Joule. The dynamic range is replaced with SNDR (to capture distortion) and BW is replaced with $f_s/2$. While it is certainly true that the power dissipation of a data converter depends on its actual acquisition bandwidth (which may be different from $f_s/2$), it is difficult to argue that there should be a strict proportionality [22].

5.6.3 Power Consumption Considerations

Reducing power consumption is an important objective in the design of the network entities and the devices, but it is more critical for the devices due to their limited power supply. The new NR features affect the power consumption of the UE's major hardware components such as antenna tuning unit, RF front-end, RFIC, baseband modem, and application processor. In this section, we identify the main factors affecting the power consumption of UE's major hardware components and discuss methods to make their operation power-efficient. As we discussed earlier, the integration of RF front-end components can significantly reduce the [insertion] loss and improve the power efficiency of the AFE.

Table 5.15 shows the impact of NR features on UE key components' power consumption. The PA is one of the key components of the AFE. For PAs, the efficiency (usually noted as power-added efficiency or PAE[28]) is used to represent the power consumption

[28] PAE is a metric for rating the efficiency of a power amplifier that takes into account the effect of the gain of the amplifier. PAE will be very similar to efficiency when the gain of the amplifier is sufficiently high. But if the amplifier gain is relatively low, the amount of power that is needed to drive the input of the amplifier should be considered in a metric that measures the efficiency of the said amplifier.

Table 5.15: Impact of NR features on UE key components' power consumption [25].

5G Features	Antenna Tuning Unit	RF Front-End	RF Integrated Circuit	Baseband Processing	Application Processor
[Scalable] Bandwidth	x	x	x	x	x
UL-MIMO	x	x	x	x	x
High-power UE	x	x	x	x	x
Bandwidth part	x	x	x	x	
Discontinuous reception				x	
Cross-slot scheduling				x	
Multiple PDCCH monitoring occasions				x	
Measurement				x	
System information acquisition				x	
Paging				x	
RAN-based notification area update				x	

characteristic. The relationship between the efficiency and power consumption can be given as $\mathrm{PAE} = (P_{OUT} - P_{IN})/P_{dc} = (P_{OUT} - P_{IN})/(V_{dc}I_{dc})$. There are many parameters affecting the efficiency of the PA including backoff power, load line loss, topology, etc. For a typical load line, the loss is between 0.3 and 0.7 dB, which means it will degrade the PAE by 7%–15%. Another factor affecting the efficiency of the PA is the backoff power. For the topology of traditional linear PAs, for example, Class-A/Class-AB/Class-F, the PAE decease when the backoff from the peak power increases. There are several methods for improving the efficiency of the PAs including the following [25]:

- *Average power tracking (APT)*: This is a technique that can be utilized to adapt the supply voltage to a PA on a timeslot basis in order to reduce power consumption of the PA. In APT method, the supplied voltage of the PA is adjusted according to output power level so that the linearity of the PA is maintained while the efficiency is improved. Although limited by the speed of the DC–DC converter, the APT voltage must be kept constant in each time slot, which means the efficiency cannot be improved, if the PA constantly operates at the highest power.
- *Envelope tracking (ET)*: The ET is a power supply adaptation technique that maximizes the energy efficiency by keeping the PA in compression mode over the entire transmission cycle, as opposed to the peaks, by dynamically adjusting the supply voltage to the PA. The effect of OFDM-based signals on PAPR motivated the use of ET technique to achieve considerable energy saving in high-PAPR transmitters. More commercial ET PAs have been developed for 4G/5G UE transmitters.
- *High voltage supply*: With higher supply voltage, the portion of knee voltage of the transistors can be minimized, which means the PA will achieve higher efficiency. Another advantage of high supply voltage is that the load line can be increased.

With increasing of the load line, the associated loss will be lower, which further improves the efficiency. However, because the supply voltage of the handset is 3.8 V at nominal condition, another DC−DC converter is needed to achieve a higher voltage. It should be noted that the DC−DC's efficiency is always less than 100%.

- *Power combination techniques*: The PAs often need to support very large output power; therefore power combination techniques are used to combine the outputs of small-PAs. Since the combiners are always at the last stage of the power amplification, the efficiency of the combiners will affect the overall PA efficiency. There are many power combination techniques including voltage or current combination. However, regardless of the combination technique used, low-loss, high-Q components must be utilized. Among combination techniques the push−pull combiner has the advantage of doubling the impedance which as to be matched, lowering the loss of the impedance transformation.
- *Doherty amplifier*: The Doherty amplifier uses a configuration known as active load-pull technique, which can modify the RF load of the PA in different RF power levels. Thus when the power backs off from the peak power, the efficiency will not drop significantly comparing to the traditional PA topologies. Each Doherty amplifier includes two devices, which are referred to as the main and the auxiliary devices, where the RF output power is the combined power of both devices. As the input drive level is reduced, both devices contribute to the output power until a certain point is reached, typically 6 dB below the maximum composite power, where the auxiliary amplifier shuts down and generates no RF power, reducing the DC power consumption. The Doherty amplifier has some disadvantages. The AM−AM and AM−PM distortion of the amplifier cannot be maintained constant as the auxiliary amplifier turning ON and OFF with the changes of the input power. Therefore DPD method is always used in the Doherty PAs. Furthermore, the performance of the Doherty amplifier is very sensitive to the load. As a result, the Doherty amplifier topology is always adopted in the BS PA design.
- *Reconfigurable technique*: The reconfigurable technique allows the PA to adapt to different configuration in different bands and different modes. With this technique, the load line, bias point, and other configurations can be adjusted. Comparing with wideband design, this technique will improve the efficiency at specified modes.
- *Digital pre-distortion*: This technique can further reduce power consumption. The DPD technology improves both power consumption and linearity.

The baseband processing in an NR UE modem has to handle a significantly increased volume of data compared to its LTE counterpart, and this is reflected in increased power consumption at the highest data rates. Some early studies have estimated the baseband contribution to UE power consumption at peak throughput, as follows [25]:

- For UE configuration (2×100 MHz, 4×4 DL, 2×2 UL), the power at peak throughput for FDD and TDD modes are 4500 and 2970 mW, respectively.
- For UE configuration (1×100 MHz, 4×4 DL, 2×2 UL) the power at peak throughput for FDD and TDD modes are 2250 and 1485 mW, respectively.

That study was based on a very simple set of assumptions. In practice, the baseband processing of two aggregated RF carriers is not exactly twice the power of a single-carrier baseband. More recent studies suggest that these estimates were reasonably accurate for the single-carrier UE. However, power consumption at peak throughput is only one aspect of UE power consumption. In an active connection, a UE will spend a large percentage of time monitoring the downlink control channel in slots which may not contain any data for the UE. Therefore, it is important to minimize the power consumption during PDCCH monitoring occasions. This can be achieved by a combination of cross-slot scheduling, bandwidth part adaptation, and MIMO restriction. The UE can further reduce its average power consumption in the connected mode (at the expense of higher latency) by entering a DRX cycle to reduce the time spent on monitoring PDCCH.

The applications processor (AP) in a smartphone comprises a number of processing cores and graphical processors which cooperate to support the computational requirements of the active applications. Multiple cores support a mixture of clock rates and processing capabilities; thus power consumption can vary significantly depending on the number of cores that are active and the set of applications that are running. The highest processing loads are usually associated with display-intensive applications, and the combined power consumption of the applications processor and display can exceed 1000 mW in some scenarios. The AP power consumption can in many cases be considered independently from the UE modem power. Present-day application data requirements are relatively modest in NR, even streaming of high-definition video to the UE uses a comparatively small proportion of the data bandwidth that is available. However, the applications themselves can have a significant impact on modem power consumption. Interactive gaming applications can involve frequent transfer of small packets of data to give a real-time response, and if the update frequency falls within the inactivity timer period of the DRX cycle, the modem will be awake most of the time with increased power consumption. Updates from different applications, if their timing is not coordinated, would increase the proportion of DRX cycles in which the UE is active and thereby increase power consumption.

5.6.4 Antenna Design Considerations

Large antenna arrays at the BS are necessary for 5G systems and in particular for operation in mmWave bands. Large number of antennas at the UE side are also expected to guarantee acceptable received signal levels for 5G services. As the number of antennas increases, it is advantageous to use 2D and 3D structures for the arrays. This reduces the required space and also enables spatial separation and beamforming in two or three dimensions. For example, a 64-element uniform linear array with an inter-element spacing of $\lambda/2$ could occupy a horizontal span of 3 m at 2 GHz, which reduces to 1.5 m if dual-polarized antenna elements are utilized. In contrast an 8×4 dual-polarized array can be accommodated in a 0.6 m $\times$ 0.4 m space and can spatially resolve users and form beams in 3D space. Antenna arrays may be arranged in a number of different ways, the most common architectures being uniform linear arrays, uniform rectangular arrays, uniform circular arrays, and stacked uniform circular arrays. The system performance of these architectures is often measured in terms of beam gain and half-power beam-width both in azimuth and in elevation planes. For massive MIMO systems, it is important to capture as many degrees of freedom of the channel as possible with a high number of effective antennas. For compact antenna configurations, this often leads to antenna structures that have shown to be effective for multidimensional channel characterization and parameter estimation. The number of antennas on the UE will not be as large as the corresponding number at the BS; however, the number of UE antennas can be much larger than what is used today in mmWave frequencies. The studies suggest that the performance improves as the array aperture increases, but the impact of the aperture is mainly visible when the users are closely grouped (i.e., they have high correlation). Furthermore, there is in general a good channel resolvability in the sense that the larger the aperture, the greater the resolvability. One important aspect to remember is that, for physically large arrays, there can be large differences in received power levels over the array, which affects user resolvability. For beamforming solutions, the antenna elements in an array must be placed close together. All the analog components (phase shifters, LNAs, PAs, etc.) should be tightly packed behind the antenna elements (active antenna systems). The dense packing of antenna elements creates two main effects: (1) spatial correlation and (2) mutual coupling [21].

The NR operation in mmWave bands necessitates the use of active antenna systems and large antenna arrays in the base stations and the devices. While this goal is enabled by relatively small-sized antenna arrays in mmWave frequencies, it significantly increases the implementation complexity and pushes the limits of antenna design and semiconductor technologies. The fully integrated mmWave transceivers and antennas require sophisticated design and consideration of power efficiency, component placement and routing, mutual couplings, and thermal constraints both at the component level and system level within a small area or volume. These considerations directly affect the achievable performance and the RF requirements and are applicable to the NR base stations and the devices, given that in mmWave frequencies, the BS/UE transceiver design and development will have less differences compared to frequency bands below 6 GHz. The extremely wide bandwidths available at mmWave frequencies set forth serious challenges for the data converters and the data conversion interfaces between the analog and digital domains in both receivers and transmitters.

5.7 Sub-6 GHz Transceiver and Antenna Design Considerations

The evolution of 4G to 5G and the reliance on the existing technology allows several LTE solutions to be integrated into the initial rollout of the NR, providing immediate benefits without waiting for future releases. The early rollouts are also supported by the use of

EN-DC functionality where an NR gNB is always associated with an LTE eNB. In Rel-15 NR, 4×4 downlink MIMO, particularly at frequencies above 2.5 GHz, which include n77/78/79 and B41/7/38, will be mandatory. The presence of four MIMO layers not only enables extended downlink data rates but also means there will be four separate antennas in the UE. An additional feature that is driven by mobile operators, however not mandatory, is the deployment of 2×2 uplink MIMO. Having 2×2 uplink MIMO in UE requires two 5G NR transmit PAs to transmit from separate antennas. This is particularly beneficial in cases where higher frequency TDD spectrum (in sub-6 GHz) is used as is the case with n41, n77, n78, and n79 as well as other TDD bands. The effective doubling of the uplink data rate enables shorter uplink bursts and flexible use of the 5G frame timing to increase the number of downlink slots, potentially increasing downlink data rates. However, when the downlink data rate increases, the uplink is challenged by the requirement of fast feedback from the UE [28].

A further use of the available second transmit path is a new transmission mode known as 2TX coherent transmission. This effectively uses the principles of diversity, which are strongly leveraged on the downlink side of the network and enable up to 1.5−2 dB of additional transmit diversity gain, which is critical to address the fundamental uplink-limited network performance. Such improvements in uplink translate into an approximately 20% increase in the range at the cell edge. In addition to improving cell edge performance, 2×2 uplink MIMO improves spectrum efficiency. Since 5G NR is mostly a TDD technology above 2 GHz, and TDD cells are likely to be configured in a highly asymmetrical configuration with priority given to downlink, improving spectrum efficiency is the key to delivering high cell capacity.

In the initial-phase of Rel-15 deployment, mobile operators emphasized the need to establish a framework for the dual-connectivity non-standalone mode of operation. In principle, network deployment with dual-connectivity NSA means that the 5G systems are overlaid by an existing LTE network. Dual-connectivity implies that the control and synchronization between the BS and the UE are provided by a 4G network, while the 5G network is a complementary radio access network tied to the 4G anchor. In this model, the 4G anchor establishes the connection using the existing 4G network which later can be complemented with a 5G NR user-plane connection (see Fig. 5.26). The addition of the new radio, together with the existing LTE multi-band carrier aggregation mode, strains the system performance, size, and interference mechanisms, posing additional challenges to be resolved when designing the NR RF front-end. Depending on 1TX or 2TX carrier frequencies and their relative spacing, intermodulation distortion products may fall into the LTE anchor receiver frequency band and cause LTE receiver desensitization.

A simplified view of NSA option-3a network topology is shown in Fig. 5.26, which is expected to be used in early rollouts of 5G networks, where the mobility is handled by LTE

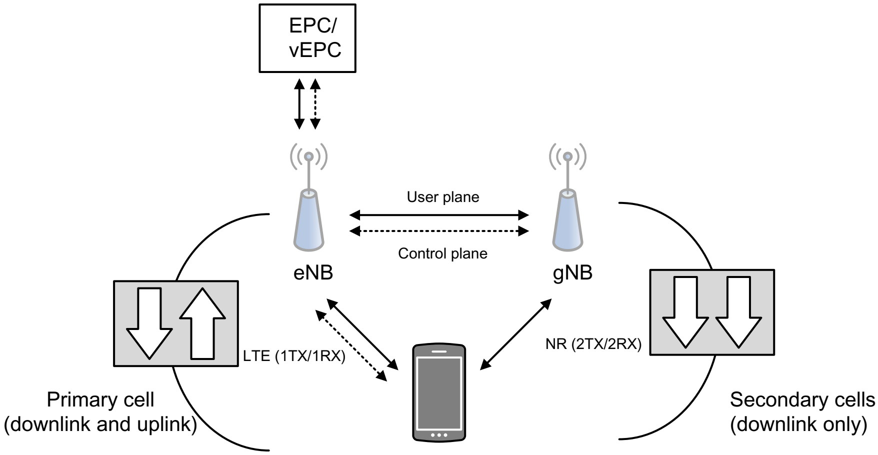

Figure 5.26
Non-standalone option-3a dual-connectivity network architecture [28].

anchors (control and user planes). This architecture leverages the LTE legacy coverage to ensure continuity of service delivery and the progressive rollout of 5G cells. It certainly seems the most plausible method of implementing 5G, while at the same time ensuring that the integrity of data connections is maintained in areas where the backhaul and network infrastructure have not yet upgraded to 5G. However, this requires UE by default to support simultaneous dual uplink transmissions of LTE (1TX/1RX) and NR (2TX/2RX) carriers in all possible combinations of standardized bands and radio access technologies (FDD, TDD, SUL, SDL), which would increase the implementation complexity where multiple separate radios and bands functioning in a small device form factor. When combined with an LTE TDD anchor point, the network operation may be synchronous, in which case the operating modes will be constrained to 1TX/2TX and 1RX/2RX, or asynchronous which will require 1TX/2TX, 1TX/2RX, 1RX/2TX, 1RX/2RX. When the LTE anchor is an FDD carrier, the TDD/FDD inter-band operation will require simultaneous 1TX/1RX, 1TX/2RX, and 2TX/1RX/2RX. In all cases, since control-plane information is transported over LTE radio bearers, it is critical to ensure that LTE anchor point uplink traffic is reliable [28].

The RF front-end developers try to mitigate the interference as much as possible to allow optimal signal transmission/reception in the UE. The complex nature of dual transmit LTE/NR concurrency and 5G-capable UE constitutes an even greater challenge for the NR RF front-end. The second phase of NR Rel-15 deployments will include standalone operation, which uses a 5G core network that will not require backward compatibility to LTE. However, the assumption is that the initial implementation of 5G NR starts with non-standalone, which is the main deployment strategy for refarmed bands, since the standalone systems are anticipated to be deployed the spectrum above 3 GHz.

The following is a list of key RF challenges that are foreseen as a result of the NR implementations in sub-6 GHz:

- *Wider channel bandwidth*: The new bands in the sub-6 GHz range feature much larger percent bandwidth[29] (n77 = 24%, n78 = 14%, n79 = 12.8%) than the existing bands (B41 = 7.5%, B40 = 4.2%, 5 GHz Wi-Fi = 12.7%). The instantaneous signal modulation bandwidth for NR is extended up to 100 MHz in bands n41, n77, n78, and n79. The contiguous intra-band EN-DC instantaneous bandwidth is 120 and 196 MHz for non-contiguous allocations. The conventional envelope tracking schemes would have difficulty going beyond 60 MHz bandwidth. Nevertheless, the NR will require new envelop tracking techniques to operate at 100 MHz bandwidth.
- High-power UE (power class 2 specific to TDD bands n41/77/78/79): As mentioned earlier, the HPUE or power class 2 (+26 dBm at a single antenna) will increase radiated output power by +3 dB relative to power class 3 operation. The PAs will need to be designed to meet higher output power with more complex waveforms. Optimized system design will be critical to achieving minimal post-PA loss in order to achieve HPUE benefits.
- *High-power transmission of NR inner-band allocations away from channel edge*: The NR would require less MPR or power backoff, when reduced waveform allocations are at a specified offset away from the channel edge. This enables much higher power across uplink modulation orders and addresses a fundamental coverage issue in LTE networks, which was related to the uplink power-limited transmission and SNR for reduced resource block allocations at the cell edge.
- *NR waveforms and 256QAM uplink*: The new NR waveforms, especially CP-OFDM, have a higher peak-to-average power ratio and will require more power backoff relative to the LTE waveforms. The 256QAM modulation is going to be used in the uplink to increase the data rates. This will challenge the RF front-end to maintain the total EVM below 3% including the PA and transceiver effects. Other issues such as in-band distortion, frame rate, and clipping must be managed to achieve maximal efficiency.
- *Cost-effective support for* 4×4 *downlink MIMO,* 2×2 *uplink MIMO and coherent 2TX transmission modes*: The 4×4 downlink MIMO is required in 3GPP for n7, n38, n41, n77, n78, n79 bands either operating as a standalone band or as part of a band combination. This feature has been prioritized due to the significant benefits of doubling the downlink data rate and spectral efficiency, as well as the up to 3 dB receive diversity gain versus 2×2 downlink modes.

[29] Percent bandwidth provides a normalized measure of how much frequency variation a system or component can tolerate. As we go higher in frequency, the absolute bandwidth will naturally increase, while its percent bandwidth will decrease. The percent bandwidth can be expressed as the ratio of the bandwidth over the carrier frequency, thus a filter with 1 GHz passband centered at 10 GHz will have 10% bandwidth, while a filter with 10 GHz bandwidth at 100 GHz will have the same percent bandwidth.

- *New 5G NR spectrum*: The new bands for sub-6 GHz will extend the frequency from 3 up to 6 GHz in the device. The increase in frequency will require improvements in the entire radio front-end as the industry tries to maintain current performance while operating at a higher frequency. There will also be new antenna multiplexing and tuning challenges, as well as in-device coexistence with 5 GHz Wi-Fi.

In the following, we focus on a practical example of a sub-6 GHz radio front-end supporting n77, n78, and n79 frequency bands. This example is intended to illustrate the important criteria for designing a 5G NR radio in sub-6 GHz (see Fig. 5.27). We will also discuss the impact of 5G spectrum, waveform, and modulation on the constituent components of the radio front-end module. The 3GPP requirements for n77 and n79 spectrum indicate 100 MHz instantaneous bandwidth for the component carrier in the uplink. This is much more rigorous than current LTE standards that use carrier aggregation of 20 MHz base channels to extend support for 40–60 MHz [28].

It is anticipated that early 5G systems will require the PA to operate with APT mode to accommodate the wider bandwidth signals. Therefore, users can expect a 100 MHz channel when the PA is operating under APT conditions. Conversely, conventional ET is challenged to perform in 40–60 MHz bandwidth. In order to extend the ET modulator bandwidth to reach 100 MHz, additional power consumption would be required, in addition to addressing amplitude/phase delay mismatch sensitivity, management of memory effects, limitations in

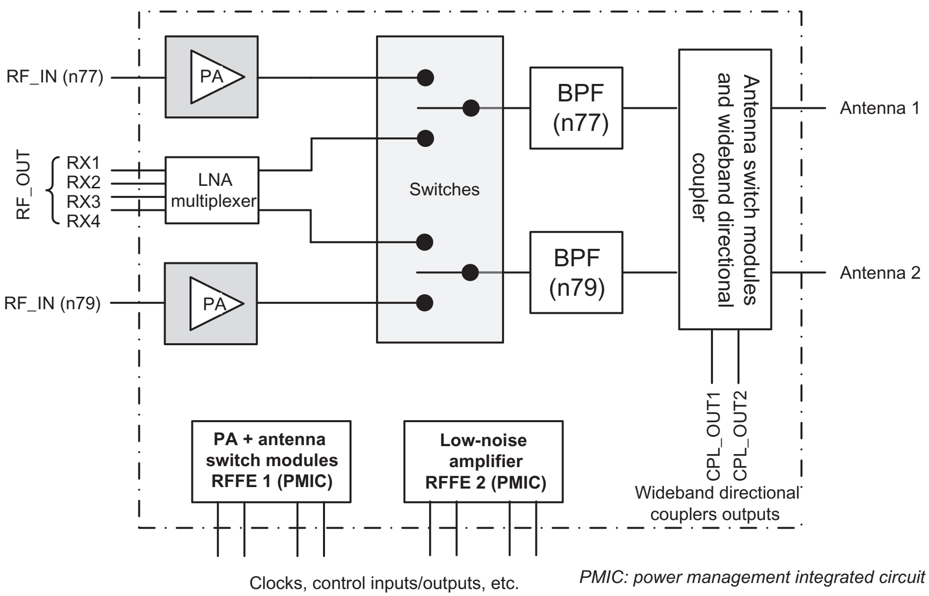

Figure 5.27
Example of sub-6 GHz 5G NR transmit/receive front-end architecture [28].

capacitive supply loading, OOB [transmitter] emissions, and intermodulation leakage into the LTE anchor band. Although, there are several promising techniques under development to extend ET to wider operating bandwidth, it will take several years before they can be commercialized. Designers are thus left with the challenge of delivering a better performing PA at two to three times the present state-of-the-art instantaneous bandwidth, while operating with a higher peak-to-average power ratio of CP-OFDM and over much larger passbands than present filters for LTE sub-3 GHz.

Aside from the wider channel bandwidth, operators have shown significant interest in high-power UE capabilities, especially as it pertains to TDD bands in the sub-6 GHz range. At present, there is some uncertainty regarding whether the bands will require 2×2 uplink (two transmitters operating at the same time) or a single transmitter would suffice, which means that the PAs will not only need to deliver industry-grade output powers compared to their LTE counterparts, but also they will have to achieve that over a wider bandwidth and at higher frequencies. Meeting higher output power at higher frequencies without ET scheme has created some serious design issues. In order to meet the new, challenging performance requirements of wider channel bandwidth and HPUE, new PA topologies have been developed that deliver linear PA performance at higher frequencies and over much wider channel bandwidths. These new architectures must be capable of significantly outperforming their LTE counterparts under more rigorous operating conditions.

When implementing the sub-6 GHz modules, integrating the receiver LNA functionality inside the module allows considerable flexibility and improves performance. In Fig. 5.27, there are two receive LNAs optimized for n77, n78, and n79 bands. Integrated LNAs have been shown to improve performance when overcoming system loss, especially in high-frequency regions where there is generally more insertion loss due to the high-frequency roll-off of various RF components. Integrated LNAs typically contribute about 1.5−2.0 dB system-wide noise figure enhancement, which translates directly to improved receive sensitivity when compared to alternative methods such as populating discrete LNAs at or near the transceiver.

In the case of sub-6 GHz applications utilizing new TDD spectrum, the legacy LTE systems are virtually non-existent. While many 3GPP specified bands exist today (e.g., B42/43/48), they have yet to be rolled out commercially in large volumes for LTE. The LTE bands only represent a small subset of the much larger NR band definitions. It is anticipated that n77, n78, and n79 RF front-end modules will be extensively deployed [28]. It must be noted that the passbands are significantly larger for these new NR bands. For example, n77 has a passband of 900 MHz; that is, 25% relative bandwidth, which is twice as large as the 5 GHz Wi-Fi band, and n79 has a passband of 600 MHz. In both instances, the conventional acoustic filters are not well suited for these extremely wide passbands. There are additional complexities that will determine the extent of the NR wideband filter requirements. For

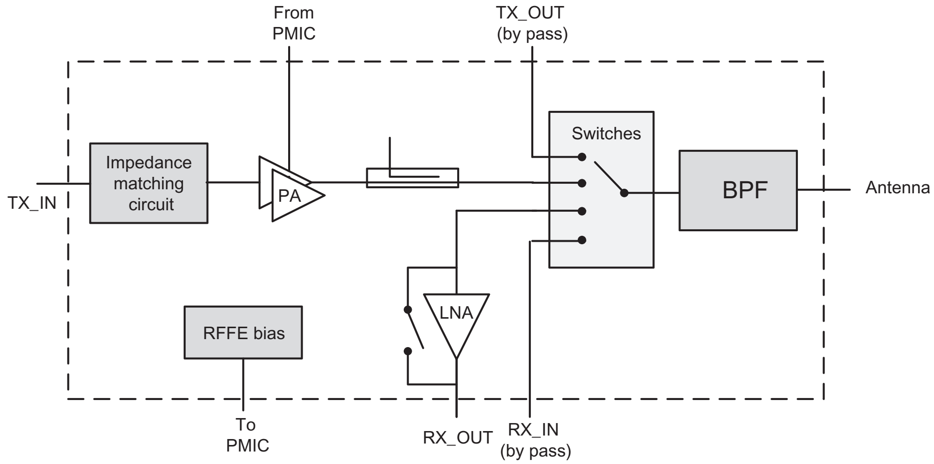

Figure 5.28
Example of 5G NR n77/n78 RF front-end block diagram [26].

example, one can derive a simple filter response, if an ideal environment with a separate high-band antenna and no coexistence requirements are assumed. On the other hand, if we consider a more complex radio environment, such as a multi-radio platform with simultaneous Wi-Fi transmission, the filter requirements become more stringent. Thus it is important to take note of the radio environment, antenna topology, and in-device coexistence requirements in order to specify the optimum filter. In other words, the filter design and antenna topology into which the front-end module will be subsequently integrated must be customized to the specific use case or application (see Fig. 5.28).

Antenna configurations play a significant role in the development and deployment of 5G products. While market requirements are slowly becoming clearer, there is already some uncertainty as to which of the optional features will be supported. One feature is the fast hopping sounding reference signal, which uses the transmitter to send a series of known symbols across all of the downlink receive antennas in the UE in order to better calibrate the MIMO channel and improve the downlink SNR. This process is key to enhanced MIMO and beamforming operations. The SRS carrier switching was recently introduced in LTE Rel-14 to assist the eNB in obtaining the channel state information of secondary TDD cells in LTE carrier aggregation scenarios. The SRS transmit switching allows the UE to route its SRS transmissions to all other available antenna ports. Assuming the channel reciprocity assumption holds, which should be the case for TDD operation, this feature enables gNB to estimate CSI on secondary downlink-only cells. By applying this concept to multi-user MIMO, the network performance is further enhanced, which in turn will improve the user experience.

5.8 mmWave Transceiver and Antenna Design Considerations

The RF PA plays an important role in wireless communication systems. The emerging 5G technologies have increased the demand for broadband, highly linear, and efficient PAs for various classes of base stations and devices. While amplifying the signal, the PA nonlinearity may cause some signal distortion which need to be minimized in order to meet the transmitter RF requirements. Therefore, PA attributes such as linearity, gain, flatness, efficiency, AM−AM/AM−PM distortion are important design parameters that must be taken into consideration. The GaAs E-mode pHEMT[30] technology inherits many advantages of D-mode pHEMT technology and exhibits single power, high span, and high linearity characteristics. The E-mode pHEMT device has a very low switching voltage, so there is no limit on the offset margin. An (E-mode) 0.15 μm GaAs pHEMT contains 17 masks and over 200 process steps. Each of these steps needs repeatable experiment to ensure the entire process is reliable and commercial grade [26].

The GaAs-based devices have dominated most of the market due to the maturity of the technology in the microwave frequency bands. However, an InP-device has a higher breakdown voltage, higher average electron velocity, larger discontinuity in the bandgap at the interface of InAlAs/InGaAs heterojunction, higher two-dimensional electron gas (2DEG) concentration,[31] and higher electron mobility in the channel, making it more suitable for

[30] A high-electron-mobility transistor (HEMT) is a field-effect transistor incorporating a junction between two materials with different band gaps (i.e., a heterojunction) as the channel instead of a doped region (as is generally the case for MOSFET). A commonly used material combination is GaAs with AlGaAs, depending on the application of the device. Devices incorporating more indium generally show better high-frequency performance, while in recent years, gallium nitride HEMTs have attracted attention due to their high-power performance. The HEMTs are used in integrated circuits as digital ON/OFF switches. HEMT transistors are able to operate at higher frequencies than ordinary transistors, up to mmWave frequencies. Ideally, the two different materials used for a heterojunction would have the same lattice constant (spacing between the atoms). In practice, the lattice constants are typically slightly different (e.g., AlGaAs on GaAs), resulting in crystal defects. A HEMT where this rule is violated is called a pHEMT or pseudo-morphic HEMT. This is achieved by using an extremely thin layer of one of the materials so thin that the crystal lattice simply stretches to fit the other material. This technique allows the construction of transistors with larger bandgap differences than otherwise possible, giving them better performance. Another way to use materials of different lattice constants is to place a buffer layer between them. This is done in the mHEMT or metamorphic HEMT, an advancement of the pHEMT. The buffer layer is made of AlInAs, with the indium concentration graded so that it can match the lattice constant of both the GaAs substrate and the GaInAs channel.

[31] A 2DEG is a scientific model in solid-state physics. It is an electron gas that is free to move in two dimensions, but tightly confined in the third dimension. This tight confinement leads to quantized energy levels for motion in the third direction, which can then be ignored for most problems. Thus, the electrons appear to be a 2D sheet embedded in a 3D coordinate system. The analogous construct of holes is called a two-dimensional hole gas, and such systems have many useful and interesting properties.

high-frequency applications. For low-voltage applications, InP HBT is expected to replace GaAs HBT. InP HBT is made of narrow bandgap material InGaAs. The bias voltage is as low as 0.5−0.6 V. In addition, InP HBT also exhibits better heat conduction characteristics, better heat dissipation capability (approximately 1.5 times of GaAs substrate), and higher peak electron velocity, allowing higher performance at lower voltages. Although InP/InGaAs HBT with lattice matching on the InP substrate has superior high-frequency and high-speed characteristics, and lower turn-on voltage than GaAs HBT, it suffers from small substrate size, high cost, and fragility of the InP substrate, limiting its large-scale and low-lost production. The capability to obtain high-quality and large-size semi-insulating InP substrate is the key to lower cost. Therefore, to meet the increasing demand for InP devices, the fundamental task is to grow high-quality and large-diameter InP single crystal substrate [26].

There is a growing interest in GaAs mHEMT technology due to the cost, fragility, and process compatibility issues of InP substrates. The GaAs mHEMT technology grows a relatively thick InAlAs layer between the channel and the GaAs substrate. The composition of indium changes from a positive value to zero, thus the mismatch of the lattice is alleviated. After adding a buffer layer with graded component, the component percentage of indium in the channel layer can be chosen almost arbitrarily between 30% and 60%, and the device performance is optimized with a great degree of freedom. The mHEMT can be considered as InP HEMT technology on GaAs substrate. It shows superior performance relative to InP in terms of low noise, allowing GaAs to solidify its position at the low-end of mmWave frequencies, and further its way into the high-end range of mmWave spectrum [26].

A 220−320 GHz mHEMT MMIC amplifier was reported in the literature using high indium composition in InGaAs as channel to achieve high-carrier density and mobility. Double InAlAs barrier layers and bilateral delta doping are designed to increase channel electron confinement and to reduce contact resistance. The buffer layer with linear component change is used to achieve lattice matching. The f_T and $f_{\max}$ are 515 and 700 GHz, respectively, for the device with 35 nm gate length and 20 μm gate width. The linear gain of the two-stage amplifier is more than 10.5 dB in the 220−320 GHz frequency band, and the small signal gain is 13.5 dB at 330 GHz. In another study, the device with 20 nm gate length was reported, and the f_T was 660 GHz for the mHEMT with 2×10 μm gate width. GaAs mHEMT has been able to work into the H-band and enter the low-end of the THz region. It has become an important component of the sub-mmWave applications with a cost advantage [26].

The mmWave RF components are expected to be deployed in large scale and high density and to meet the multi-band and high-frequency requirements. This would underline the importance of mmWave filter technologies. The acoustic filter technologies continue to evolve to meet the challenges of the global transition to 5G networks. SAW[32] and bulk acoustic wave (BAW)[33] filter technologies are widely used in today's mobile device filtering functions. Under the existing framework, SAW filters are widely used in 2G, 3G, and 4G wireless technologies, which has the advantages of design flexibility, good frequency selection characteristics, and small size; and can utilize the same production process as the integrated circuits. The frequency characteristic of the SAW filter is closely related to the pitch of the inter-digital transducers electrode, thus the higher the frequency, the smaller the distance required between the electrodes. The SAW filter quality drops rapidly when the frequency exceeds 2.5 GHz. Therefore, the SAW filter is mostly used in below 2.5 GHz applications. The BAW filters are gaining traction beyond 2.5 GHz. In contrast to SAW filters, they are insensitive to temperature changes, with low insertion loss and high OOB rejection, making them a good candidate for 4G/5G wireless systems' RF filtering applications. However, most BAW filters are used below 6 GHz, although they can reach to a maximum frequency of 20 GHz. The mmWave micro-electro-mechanical systems (MEMS) filter[34] on semiconductor substrate, uses the semiconductor manufacturing process, and constitutes a filter structure with high-Q factor and low-loss characteristics, as well as good performance in mmWave bands. However, regardless of whether it is using GaAs substrate or silicon substrate, the cost is relatively high. Under the conditions of ensuring a high degree of consistency, mass production is still more difficult [26].

As for the mmWave LNAs, GaAs pHEMT has excellent noise and linearity performance. Due to its high insulation rate, GaAs substrate improves chip inductor quality factor (Q), which helps reduce the overall NF of the terminal's receiver. While GaAs pHEMT has excellent noise and gain performance, it has not been widely used in the early years due to

[32] A SAW filter is a filter whereby the electrical input signal is converted to an acoustic wave by inter-digital transducers (IDTs) on a piezoelectric substrate such as quartz. The IDTs consist of interleaved metal electrodes which are used to launch and receive the waves, so that an electrical signal is converted to an acoustic wave and then back to an electrical signal. The most common group of SAW filters are bandpass filters, which are in very widespread use in radio systems. There are many types with differing advantages, such as low shape factor, low insertion loss, small size, or high-frequency operation. The wide variety of types is possible because almost arbitrary shapes can be defined on the surface with very high precision. SAW filters are limited to frequencies from about 50 MHz up to 3 GHz.

[33] Unlike SAW filters, the acoustic wave in a BAW filter propagates vertically. In a BAW resonator using a quartz crystal as the substrate, metal patches on the top and bottom sides of the quartz excite the acoustic waves, which bounce from the top to the bottom surface to form a standing acoustic wave. The frequency at which resonance occurs is determined by the thickness of the slab and the mass of the electrodes. At the high frequencies in which BAW filters are effective, the piezo layer must be only micro-meters thick, requiring the resonator structure to be made using thin-film deposition and micro-machining on a carrier substrate. The BAW filter size also decreases with higher frequencies, which makes ideal for most wireless applications. In addition, BAW design is far less sensitive to temperature variation even at broad bandwidths.

[34] MEMS are millimeter/sub-millimeter devices, realizing a certain transduction function between two (or more) distinct physical domains, among which the mechanical is always involved. More specifically, regardless of the specific function it is conceived for, a MEMS device always features tiny structural parts that move, bend, stretch, deform, and/or contact together. These characteristics make micro-system devices particularly suitable for the realization of a wide variety of micro-sized sensors and actuators.

its cost of production. However, with the recent improvement of the process, GaAs pHEMT has become the mainstream for the development of LNAs. For cost and performance considerations, the mainstream process is the gate length of 0.25 μm. GaAs pHEMT technology allows implementation of gate lengths in the order of 0.1 μm. As an example, the NF and gain of a prototype LNA based on the latter technology is 1.3 and 7 dB, respectively, at 40 GHz. Another prototype LNA developed by one of the leading RF companies has 0.13 μm gate length and a cutoff frequency of 110 GHz with an NF of less than 0.5 dB in 8−12 GHz band [26].

Switches are the smallest active electronic components and play a critical role in the overall operation and performance of wireless communication systems. Switches route the transmit and receive signals at different frequencies, enable monitoring and calibration, and are the baseline component for functions such as phase shifting and step attenuation. The 5G wireless systems will continue reliance on solid-state switching. The switch performance is evaluated by a number of different specifications. In general, a switch must have a low insertion loss, high linearity, high port-to-port isolation, high-power handling, and a fast switching time. In mmWave systems, broadband frequency operation becomes increasingly important. To facilitate the transition to mmWave, switches must be able to support multiple bands, such as dual bands (28 and 39 GHz) or even triple bands (24, 28, and 39 GHz). In addition to multiple bands, each band has rigorous requirements on frequency-response variation. For example, a 28 GHz mmWave application will have a wide bandwidth and the system frequency response (gain) may not meet the flatness requirement. Furthermore, a true wideband support implies minimum variation across the supported frequency bands [26].

5.9 Large Antenna Array Design and Implementation in Sub-6 GHz and mmWave Bands

It is desirable for an antenna system to be able to focus its radiated energy in a particular direction in order to maximize the signal power toward a particular user device. Using a simple $\lambda/2$ dipole as an example, the gain can be increased using two methods [18,19]:

- Antenna aperture: By increasing the aperture (or size of the antenna), the (larger) antenna becomes more directive due to the periodic current distribution across the antenna. Although this method does not require external circuitry for control, the direction of the beam is fixed and the number of sidelobes increases.
- Antenna array: If the single dipole element is repeated according to the periodicity of the current distribution, an antenna array is created. The amplitude and phase of the signals to individual elements can be adjusted to control both the beam direction and sidelobe levels, creating a phased-array. This results in a significantly more complex feeding network with relatively higher losses.

The beam steering capabilities of an antenna array can create both a high-gain beam toward a specific direction as well as creating a null in a specific direction in order to mitigate interference in a MU-MIMO system. Therefore, in addition to phase shifting, weighting of the signal amplitude is applied to reduce the side lobes. For example, a symmetric linear tapering of the signal amplitudes in a linear antenna array results in sidelobes that are 10−15 dB lower, but it increases the beam-width of the main lobe by approximately 5 degrees.

As we discussed in Chapter 4, there are three types of beamforming architectures used for antenna arrays (see Fig. 5.29):

- *Analog beamforming (ABF)*: The traditional way to form beams is to use attenuators and phase shifters as part of the analog RF circuitry where a single data stream is divided into separate paths. The advantage of this method is that only one RF chain (PA, LNA, filters, switch/circulator) is required. The disadvantage is the loss from the cascaded phase shifters at high power.
- *Digital beamforming (DBF)*: DBF assumes that there is a separate RF chain for each antenna element. The beam is formed by matrix-based operations in the baseband where

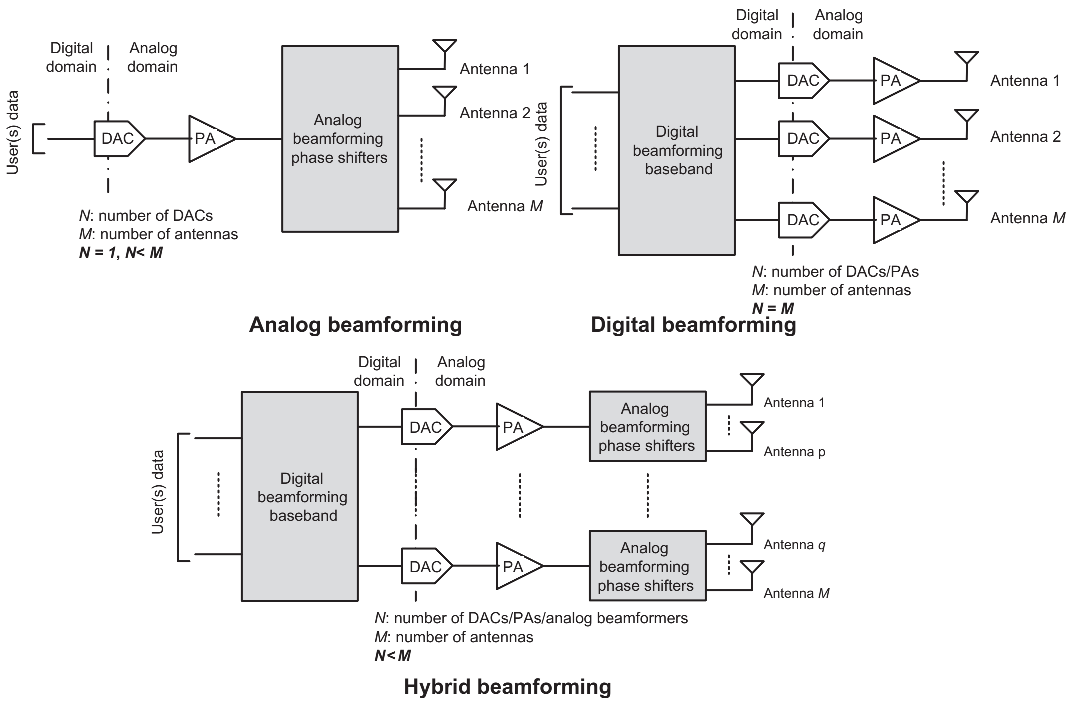

Figure 5.29
Beamforming architectures for active antenna systems [18].

digital amplitude and phase weighting is performed. For frequencies lower than 6 GHz, this is the preferred method since the RF chain components are comparatively inexpensive and can combine MIMO and beamforming into a single array. For frequencies of 28 GHz and above, the PAs and ADCs are very lossy for standard CMOS components. If alternative semiconductor technologies, such as GaAs and gallium nitrate, are used, the losses decrease at the expense of higher cost.

- *Hybrid beamforming (HBF)*: HBF combines DBF with ABF in order to allow the flexibility of MIMO plus beamforming while reducing the cost and losses of the beamforming unit (BFU). Each data stream has its own [separate] analog BFU with a set of M antennas. If there are N data streams, then there are $N \times M$ antennas. The analog BFU loss due to phase shifters can be mitigated by replacing the adaptive phase shifters with a selective beamformer such as a Butler matrix. One proposed architecture uses the digital BFU to steer the direction of the main beam while the analog BFU steers the beam within the digital envelop.

A 32-element linear antenna array (with $\lambda/2$ antenna spacing) is simulated and reported in [19] and the array factor is illustrated in Fig. 5.30, where the transceiver number is 4, and the antenna number per transceiver is 8. The maximum array gain is achieved at azimuth 90 degrees. The first curve in Fig. 5.30 depicts the array factor of ABF consisting of eight

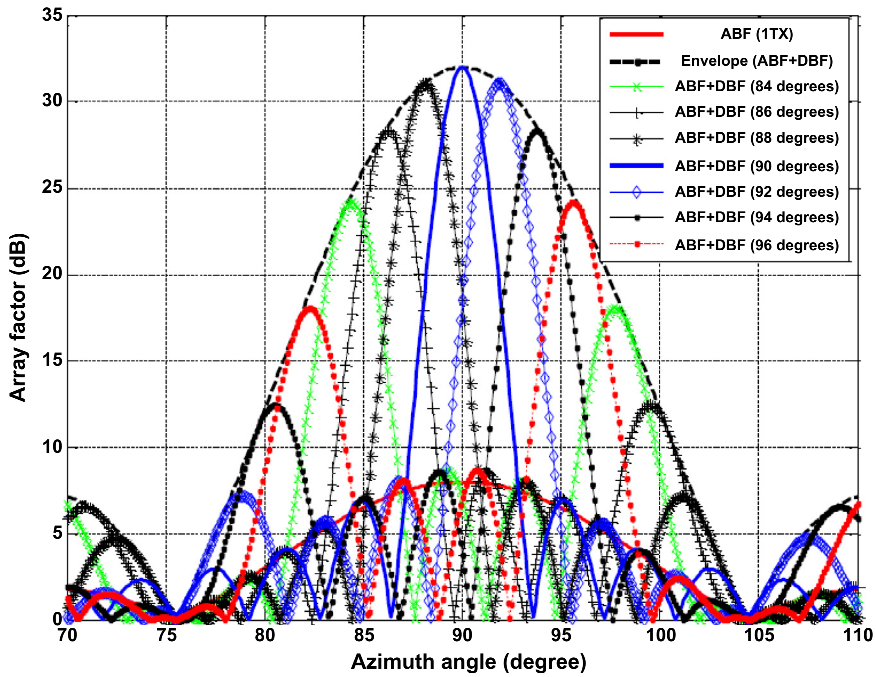

Figure 5.30
Comparison between analog, digital, and hybrid beamforming [18,19].

antenna elements with the main beam direction at azimuth 90 degrees. The second curve in the figure is the array factor of the 32-element hybrid analog and digital beamforming. It can be seen that the amplitude of the second curve is exactly four times of that of the first curve. With same ABF per transceiver, seven DBF designs are shown, with the main beam direction of azimuth 84, 86, 88, 90, 92, 94, and 96 degrees, respectively. As shown in Fig. 5.30, the main beam direction can be controlled with HBF design. Some observations are summarized as follows [19]:

1. The further the main beam direction of the HBF is away from the ABF main beam direction, the smaller the gain. Therefore, the coverage of the beams is limited.
2. The HBF designed for 90 degrees is actually the ABF with all antennas.
3. Users located at certain AoD φ within the range of 84−96 degrees will observe that the HBF designed for direction φ has the largest signal power.

Due to the sensitivity of the antenna array beam steering to the phase differences between the antenna elements, each array must be calibrated for the following tolerances (see Fig. 5.31) [18]:

- Phase: Phase error can have a significant effect on the antenna beam depending on its statistical properties. If the phase error is uniformly distributed across the array, then the main beam direction does not change. Instead, the nulls that are often used to block

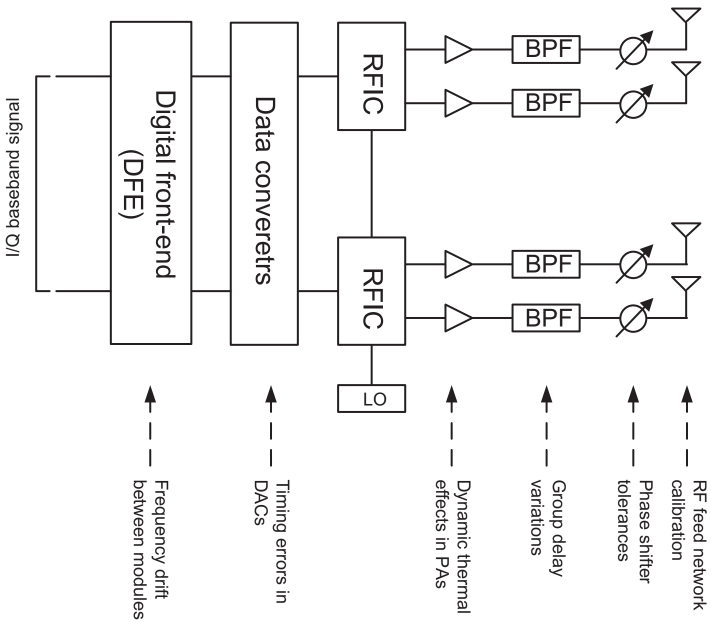

Figure 5.31
Array calibration: static and dynamic tolerances in active antenna systems [18].

interference are severely affected, resulting in 10−20 dB loss. If there is more deterministic phase error distribution, this will steer the beam in a different direction. Phase error can be caused by manufacturing tolerances in the RF feeding network, thermal effects in the PAs and LNAs, and group delay variations in the filters. It is recommended to keep the phase error between antenna elements below ±5 degrees (commercial specification for AAS).

- Amplitude: Amplitude error does not affect the direction of the beam, but rather the peak gain and the sidelobe levels and is generally due to the thermal effects on the active components (PA and LNA). Recommended error should be below ±0.5 dB (commercial specification for AAS).
- Timing/frequency: Depending on the circuit architecture, if a common LO network is not used between modules, there will be frequency drift in addition to the timing errors in the ADCs. Recommended level of frequency drift is 0.5 ppm (commercial specification for AAS).

The different phases in active antenna array development require different measurement and verification methods, thereby using various approaches to measure a massive MIMO system will require different test interfaces to both the complete antenna array and individual antenna elements. Antenna arrays with 64 or more elements (corresponding to an 8×8 cross-polarized antenna array) may not provide any individual antenna connectors in the final assembly. In earlier phases of the product design, however, antenna elements are typically accessible with connectors to verify the S-parameters of individual antennas. Verification and qualification of antenna arrays is required in all product development phases from initial R&D design to final production test. Mutual coupling between antenna elements has an adverse effect on network capacity. Therefore, simultaneous multi-port passive (conducted) measurements for accurate characterization are required [18].

In the design phase, when antenna connectors are still accessible, a vector network analyzer can be used to measure the mutual coupling between the antenna array elements. For antenna arrays, the most common measurement with a vector network analyzer is the S-parameter measurements (both transmission and reflection coefficients). The S-parameters include magnitude and phase information, which can be used to measure both near-field and far-field quantities. For a two-port system, S_{11} represents the reflection coefficient (reflected power at antenna 1 divided by injected power at antenna 1) and S_{21} denotes transmission coefficient (transmitted power at antenna 2 divided by injected power at antenna 1). An example of an antenna array is shown in Fig. 5.32 with 64 dual-polarized antennas and 128 antenna ports. Due to the large number of antenna ports, connecting cables and subsequent calibration of the vector network analyzer is difficult and time-consuming [18].

Fig. 5.33 illustrates the effect of antenna mutual coupling on the capacity of the cell. We compare a uniform linear array with a 2D planner array. It is shown that the element

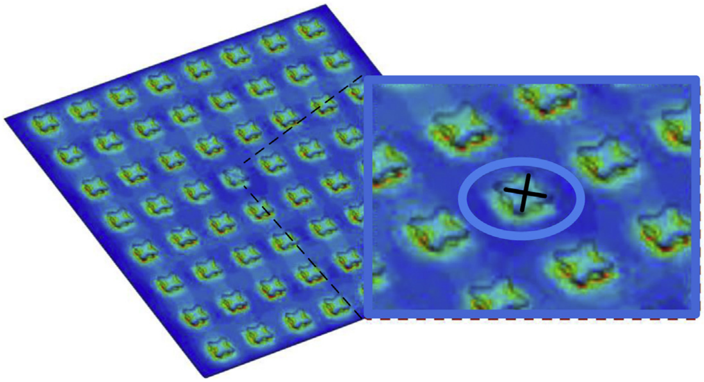

Figure 5.32
64-Element planar antenna array and mutual coupling between array elements [18].

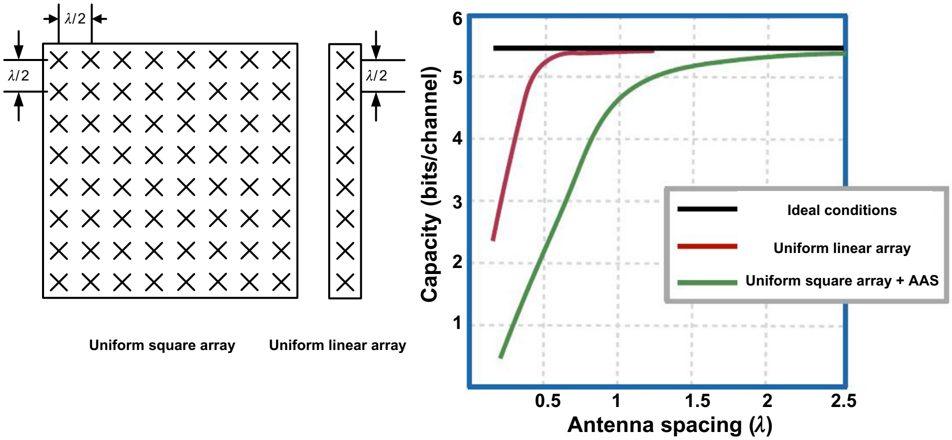

Figure 5.33
Network capacity as a function of mutual coupling in two antenna arrays [20].

spacing between elements on a 2D planner array needs to be three times greater than that between elements on a uniform linear array to maintain the same network capacity. A large antenna array to implement massive MIMO in both sub-6 GHz and mmWave frequency ranges does not provide antenna connectors due to added complexity and cost, physical size limitations, and resulting insertion loss. Consequently, the OTA tests are required. The OTA tests measuring the 3D antenna pattern can be performed either in the near-field or in the far-field. The measurements in the antenna near-field allow smaller anechoic chambers for the measurement; however, it requires an additional near-field to far-field transformation for antenna gain patterns.

As we mentioned in the previous sections, the OTA measurements can be used in pre-production and post-production phases. In the pre-production phase, the following OTA measurements need to be conducted [18]:

- Gain patterns: Gain patterns are either 2D from one of the three principal planes (E1, E2, or H-plane) or a complete 3D pattern. For antenna arrays with one or more beams, the 3D gain pattern is more useful.
- Radiated power: The EIRP is used to measure an AAS either as a UE or a BS. For UE testing, TRP is used instead where TRP is the weighted integral of the effective radiated power values over a sphere.
- Receiver sensitivity: Receiver sensitivity is characterized by the equivalent isotropic sensitivity (EIS) and measures the block error rate as a function of the receive power equal to the specified receiver sensitivity.
- Transceiver and receiver characterization: Each individual transceiver in the AAS needs to be verified through an OTA interface. This includes a range of measurements for both the transmitter (maximum output power, EVM, ACLR, spurious emissions, intermodulation) and the receiver (sensitivity, dynamic range, band selection, in-band/out-of-band/narrowband blocking, ACS). It is assumed that each transceiver will turn on for individual verification.
- Beam steering and beam tracking: Due to the high path loss and limited range of a mmWave wireless system, precise beam tracking and fast beam acquisition is required for mobile users. While in the antenna implementations of the existing cellular technologies static beam pattern characterization was sufficient, mmWave systems will further require dynamic beam measurement systems.

In the post-production phase the following tests need to be conducted [18]:

- Antenna calibration: In order to accurately form beams, the phase misalignment between RF signal paths needs to be less than ± 5 degrees. This measurement can be performed for both passive and active antenna systems using either a phase-coherent receiver to measure the relative difference between all antenna elements. This is then compiled into a lookup table for the AAS to use as a reference for beam generation or to calibrate the internal self-calibration circuits inside the AAS unit.
- Transceiver calibration: Due to lack of RF ports on some massive MIMO systems, the individual transceivers will need to be calibrated using OTA techniques. This includes both transmitters and receivers.
- Five-point beam test: The AAS manufacturer specifies a reference beam direction, maximum EIRP, and accordingly EIRP values for each declared beam. For conformance, the maximum EIRP point, and four additional points corresponding to the most extreme steering positions are measured.

- Functional tests: This is the final test performed on the completely assembled unit in production. It can consist of a simple radiated test, a five-point beam test, and aggregate transceiver functionality, such as an EVM measurement of all transceivers.

References

3GPP Specifications[35]

[1] 3GPP TS 37.105, Active Antenna System (AAS) Base Station (BS) Transmission and Reception (Release 15), December 2018.
[2] 3GPP TS 38.101-1, NR User Equipment (UE) Radio Transmission and Reception; Part 1: Frequency Range 1 Standalone (Release 15), December 2018.
[3] 3GPP TS 38.101-2, NR User Equipment (UE) Radio Transmission and Reception; Part 2: Frequency Range 2 Standalone (Release 15), December 2018.
[4] 3GPP TS 38.101-3, NR User Equipment (UE) Radio Transmission and Reception; Part 3: Frequency Range 1 and Frequency Range 2 Interworking Operation With Other Radios (Release 15), December 2018.
[5] 3GPP TS 38.104, NR Base Station (BS) Radio Transmission and Reception (Release 15), December 2018.
[6] 3GPP TS 38.133, NR Requirements for Support of Radio Resource Management (Release 15), January 2019.
[7] 3GPP TS 38.215, NR, Physical Layer Measurements (Release 15), December 2018.

ITU-R Recommendations[36]

[8] ITU-R Recommendation SM.328, Spectra and Bandwidth of Emissions, May 2006.
[9] ITU-R Recommendation SM.329, Unwanted Emissions in the Spurious Domain, September 2012.

Articles, Books, White Papers, and Application Notes

[10] E. Dahlman, S. Parkvall, 5G NR: The Next Generation Wireless Access Technology, Academic Press, August 2018.
[11] S. Ahmadi, LTE-Advanced: A Practical Systems Approach to Understanding 3GPP LTE Releases 10 and 11 Radio Access Technologies, November 2013, Academic Press.
[12] Global Mobile Suppliers Association (GSA) Report, Spectrum for Terrestrial 5G Networks: Licensing Developments Worldwide, January 2019.
[13] Ian Beavers, Giga-Sample ADCs Promise, Direct RF Conversion, Electronic Design, September 2014.
[14] 3GPP RWS-180011, NR Radio Frequency and Co-existence, Workshop on 3GPP Submission Toward IMT-2020, October 2018.
[15] T. Neu, Direct RF Conversion: From Vision to Reality, Texas Instruments Incorporated Application Note, May 2015.
[16] D. Anzaldo, Application Note, LTE-Advanced Release-12 Shapes New eNB Transmitter Architecture: Part 2, Analog Integration Challenge, Maxim Integrated Products, Inc., March 2015.
[17] C. Pearson, High Speed, Digital to Analog Converters Basics, Texas Instruments Incorporated, Application Note, October 2012.
[18] M. Kottkamp, C. Rowell, Antenna Array Testing—Conducted and Over the Air: The Way to 5G, Rohde & Schwarz White Paper, November 2016.

[35] 3GPP specifications can be accessed at the following URL: http://www.3gpp.org/ftp/Specs/archive/.
[36] ITU-R recommendations can be access at the following URL: https://www.itu.int/rec/.

[19] S. Han, et al., Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G, IEEE Commun. Mag. 53 (1) (January 2015).
[20] F. Rusek, et al., Scaling up MIMO: opportunities and challenges with very large arrays, IEEE Signal Proc. Mag. 30 (1) (January 2013).
[21] M. Shafi, et al., 5G: a tutorial overview of standards, trials, challenges, deployment, and practice, IEEE J. Sel. Areas Commun. 35 (6) (June 2017).
[22] B. Murmann, The race for the extra decibel, a brief review of current ADC performance trajectories, IEEE Solid-State Circ. Mag. 7 (3) (2015).
[23] U. Rüddenklau, mmWave Semiconductor Industry Technologies: Status and Evolution, ETSI White Paper, (July 2016).
[24] B. Peterson, D. Schnaufer, 5G Fixed Wireless Access Array and RF Front-End Tradeoffs, *Microwave Journal*, February, (2018).
[25] Global TD-LTE Initiative, GTI 5G Device Power Consumption White Paper, (February 2019).
[26] Global TD-LTE Initiative, GTI 5G Device RF Component Research Report, (January 2019).
[27] G. Lloyd, Linearization of RF Front-ends, Rohde & Schwarz White Paper, (September 2016).
[28] K. Walsh, White Paper 5G New Radio Solutions: Revolutionary Applications Here Sooner Than You Think, Skyworks Solutions, Inc., (September 2018).
[29] Y. Sano, et al., 5G radio performance and radio resource management specifications, NTT DoCoMo Tech. J. 20 (3) (January 2019).