The world’s thirst for wireless communication keeps increasing as users are attracted to new broadband services for accessing data in the cloud, video conferencing, and streaming videos using various user equipment. This growing demand for higher data rates (≥ 1 G/bs) is motivating vigorous research activities worldwide to develop wide-band and multiband systems. The fifth generation (5G) of wireless standards [Release (Rel)-14 in 2017, Rel-15 due in 2018, and Rel-16 due in 2019] are being developed for cellular communication by the 3rd Generation Partnership Project (3GPP) to directly address these issues [1]. We will first briefly review these developments to assess some of the new testing and calibration challenges they will introduce.

Review of Recent Developments

At low frequencies (<6 GHz), carrier aggregation and multiband scheduling can provide the means to increase the bandwidth for selected users. However, while the demand for increased bandwidth keeps on growing, the ever-increasing pool of wireless devices must continue to be supported, and more bandwidth must eventually be identified.

One solution, besides the new spectrum allocation implemented in advanced fourth-generation long-term evolution (LTE) Rel-10, which is being actively pursued in 5G, is to increase user capacity by making use of spatial diversity/multiplexing with the help of antenna arrays. This approach, originally referred to as space-division multiple access (SDMA), relies on multiple...
transmit and receive antennas in base stations (eight to 64 antennas) and mobile units (two to four antennas) and can be implemented using multiuser multiple-input/multiple-output (MU-MIMO) [2], [3].

In LTE Rel-10, up to eight antenna ports per transmission node are used to theoretically increase user capacity by a factor of 4 to 8 (in theory). In massive MIMO, a much larger number of antennas (hundreds) is being considered. In LTE Rel-13, 64 antennas providing another factor of 4 to 8 in capacity increase (in theory) are allowed in base stations. When both horizontal and vertical beam control are implemented, it is referred to as three-dimensional or full-dimension (FD) MIMO, with FD-MIMO being the term adopted by 3GPP [1].

The use of higher carrier frequencies—above 6 and 10 GHz [millimeter-wave (mmW)]—additionally provides a natural means to increase bandwidth and is also the subject of active investigation for 5G. However, at mmW frequencies, the signal attenuation in the channel is greatly increased, leading to a poor link budget. Furthermore, generating high-power mmW signals is more difficult and costly. Thus smaller (~200 m) and more closely spaced cells must be used.

To alleviate signal attenuation, beamforming synthesized by antenna arrays provides the means to increase the signal power delivered to each user while increasing the capacity (number of users) via spatial diversity/multiplexing. Thus, beamforming is investigated as an effective implementation of MU-MIMO for line-of-sight (LOS) channels. Beamforming is also being investigated below 6 GHz for future LTE applications (Rel-13) [1].

New techniques will be needed to test, calibrate, and operate 5G MIMO equipment. Given the complexity of the future 5G wireless systems involved and the impairments associated with the physical hardware in the base stations and mobile devices, 5G systems will need to be able to calibrate and linearize themselves. In the field of radar applications, self-calibration and adaptive schemes for beamforming have already been successfully deployed in phased-array antennas [4]. Cost-efficient solutions are required, however, for successful commercial deployment. Commercial test equipment such as the PXB baseband generator and channel emulator from Keysight Technologies is available to study the impact of the MIMO channel on the transmitter (Tx)/receiver (Rx) chain [5]. The MIMO channel emulation is performed at baseband, and the PXB can be inserted directly between the Tx and the Rx baseband modems. Upconversion and downconversion to RF are also an option if the nonlinear impairments of the RF Tx chain are also to be included.

At frequencies below 6 GHz, low-cost software-defined-radio (SDR) MIMO test beds are now available from a number of vendors, typically covering 50 MHz to 6 GHz with bandwidth up to 100 MHz. SDR test beds operating at mmW frequencies are also becoming available [6]. The use of these frequency-agile SDR MIMO systems for communication enables us to conceive systems capable of adapting to varying radio environment (cognitive radio) while supporting multiple standards with the use of a field-programmable gate-array (FPGA) baseband modem. Given their frequency-agility, SDR MIMO systems can also be used to study the increase/enhancement in data rate via concurrent carrier aggregation and multiband transmission.

Spectrum flexibility is also a target of 5G, and new frequency bands, including some below 1 GHz, are being investigated. The multiplication of bands that mobile devices might be called upon to support is also creating a bottleneck. Ideally, in an SDR system, a single ultrawide-band power amplifier (PA) is targeted, whereas as many PAs as bands are used in conventional cellular phone systems. As a hybrid solution, research is being actively pursued on multiband PAs that cover multiple noncontiguous bands while remaining power efficient (see [7] for a review of this field).

In this article we focus on the characterization and mitigation of the impairments associated with the PA nonlinearities in SDR systems when the same PA is used for the amplification of concurrent dual-band LTE/wide-band code-division multiple-access (WCDMA) signals. Both predistortion [8] and feedforward [9], [10] approaches will be presented. Our review will conclude with a discussion of nonlinear impairments in MU-MIMO systems and the advanced configurations for self-testing and adaptation that might be called upon for their mitigation.

**Dual-Band DPD in the Presence of Unfilterable Harmonic Interference**

In our first example of self-calibration, we consider a broadband PA excited by a dual-band signal, as shown in Figure 1. This dual-input/single-output (DISO) test bed consists of two channel signals: \( x_1(n) \) at the frequency \( \omega_1 \) and \( x_2(n) \) at the frequency \( \omega_2 \). For a wide range of separation between the bands, the various intermodulation products do not overlap with bands 1 and 2 and can be removed with the help of filters. Readers are referred to a previous review of frequency selective two-dimensional (2-D) DPD for the linearization of PAs under such a scenario [7]. In commercial SDR systems deployed in the 1–6-GHz frequency range, when the separation between the two bands becomes large enough that we have \( x_2 \) at a frequency verifying \( \omega_2 \approx n_2 \omega_1 \), the \( n \)th harmonic of the lower-band signal overlaps with the higher-band signal. That is, the input \( x_1 \) for band 1 at the frequency \( \omega_1 \) is generated via the \( n \)th-order nonlinearity of the PA: a signal \( x_1^n \) at the frequency \( n_1 \omega_1 \) that is interfering with the signal \( x_2 \) at the frequency \( \omega_2 \approx n_2 \omega_1 \). Similarly, for \( n = 2 \), the second-order nonlinearity of the PA induces a signal.
\( x_1 \) at the frequency \( \omega_2 - \omega_1 \approx \omega_1 \) that is interfering with the signal \( x_1 \) at the frequency \( \omega_1 \).

To better visualize the contribution of these nonlinear terms, let us consider the case where the separation between the two bands \( \omega_2 - \omega_1 \) is not equal to \( \omega_1 \) but rather to \( \omega_1 - \Delta \omega \), so that the second harmonic of the first band at \( 2\omega_1 \) does not quite overlap with the second band at \( \omega_2 \). The resulting spectrum at the PA output is plotted in Figure 2. The second harmonic of the lower-band signal \( x_1 \) term at the frequency \( 2\omega_1 \) is now on the right side of the higher-band \( \omega_2 = 2\omega_1 - \Delta \omega \) as is shown in Figure 2(b). Conversely, the second-order nonlinearity of the PA induces a signal \( x_1 \) at the frequency \( \omega_2 \) as \( \omega_2 - \omega_1 = \omega_1 - \Delta \omega \), which is now to the left side of the signal \( x_1 \) at the frequency \( \omega_1 \), as shown in Figure 2(a).

The continuous wave (CW) behavior of the PA system can be accurately modeled using the harmonic

**Figure 1.** Dual-band scenarios in which the nth harmonic of the first band (a) does not interfere and (b) does interfere with the second band signal transmission (from [8]).

**Figure 2.** (a) and (b) Identification of the side bands generated by the concurrent operation of two bands with the upper band frequency near the second harmonic frequency of the lower band. PSD: power spectral density.
**Harmonic Volterra Theory**

Here, we review a simple but general behavioral model that is available for continuous nonlinear systems excited by arbitrarily large multiharmonic signals. Let us, for simplicity, consider a single-input/single-output (SISO) system (see Figure S1). As shown in Figure S2, such a SISO system could be used for a unilateral amplifier to model either its multiharmonic transmitted or reflected waves in terms of the multiharmonic incident waves. A general formalism, the Volterra series is available in the time domain to model such continuous time-invariant nonlinear systems for arbitrary excitations. However, extracting the Volterra kernels for arbitrary excitations is quite a costly task.

Let us assume, instead, that we excite the SISO system with continuous wave (CW) RF excitations, including a fundamental and a few harmonics. It can then be verified that when a nonlinear system of infinite order is excited by such a finite number of CW harmonic tones, the output signal can be expanded in a finite number of leading nonlinear RF terms weighted by nonlinear baseband harmonics Volterra functions [12].

For example, consider the simplest case, in which the nonlinear system is excited at port 1 by the fundamental \(a_1(\omega)\) and the second harmonic \(a_1(2\omega)\) incident waves, both of arbitrary amplitude and phase. The transmitted waves at the fundamental, second, and third harmonics can then be written quite generally as [12]

\[
b_2(\omega) = V_{1,1}(x_1) \cdot a_1(\omega) + V_{1,2}(x_1) \cdot a_1^*(\omega) \cdot a_1(2\omega)
\]

\[
b_3(2\omega) = V_{2,1}(x_2) \cdot a_1(2\omega) + V_{2,2}(x_2) \cdot a_1^*(\omega)
\]

\[
b_3(3\omega) = V_{3,1}(x_3) \cdot a_1^*(\omega) + V_{3,2}(x_3) \cdot a_1(\omega) \cdot a_1(2\omega)
\]

with \(x_2 = [V_{DC}, |\sigma_1(\omega)|, |\sigma_1(2\omega)|, \angle\{\sigma_1(2\omega) \cdot \sigma_1^*(\omega)\}].\)

The harmonic Volterra functions \(V_{ij}(x_i)\) are all analytic functions of \(x_2\). The term \(x_2\) defines the dc/baseband large-signal operating point. These terms are indeed all dc terms under CW operation, and their spectra extend over the baseband under modulation. For example, the term \(|\sigma_1(\omega)|^2 = a_1(\omega) \cdot a_1^*(\omega)\) is associated with the frequency \(\omega - \omega = 0\), and the term \(a_1(\omega) \cdot a_1^*(2\omega)\) is associated with the frequency \(2\omega - 2\omega = 0\). In these expressions, the * used for complex conjugate brings about a negative frequency, because the conjugate is applied not only to the complex harmonic amplitude \(a_1(\rho\omega)\) but also to its phasor \(\exp(j\rho\omega)\). An alternate, more compact, mathematical representation with real variables can also be used:

Using the same rule, it easily verified that \(a_1(\omega) \cdot a_1^*(2\omega)\) is of frequency \(4\omega - \omega = 3\omega\) like the term \(b_2(3\omega)\), to which it contributes. This harmonic Volterra function representation holds for continuous systems of infinite orders, such as devices exhibiting an exponential dependence on their inputs [12].

For weak nonlinear excitations, when the harmonics amplitude is small compared to the fundamental \(|\sigma_1(\rho\omega)| \ll |\sigma_1(\omega)|\), a first-order Taylor series expansion of the harmonic Volterra functions can be performed relative to the amplitude of the harmonics \(|\sigma_1(\rho\omega)|\) [12], yielding (after setting to zero the phase of the incident wave at the fundamental frequency) \(\angle\{a_1(\omega)\} = 0\):

\[
b_2(\rho\omega) = \sum_{q=1}^{n_0} S_{2,1,p,q}(x_2) \cdot \sigma_1(\rho\omega) + T_{2,1,p,q}(x_2) \cdot \sigma_1^*(\rho\omega)
\]

with \(x_2 = [V_{DC}, |\sigma_1(\omega)|, \angle\{\sigma_1(2\omega) \cdot \sigma_1^*(\omega)\}].\)

This is the polyharmonic distortion approximation [22], which is also known as an \(X\)-parameter expansion. It is very helpful for modeling the behavioral response of power amplifiers, as they are usually excited by signals with weak harmonic components. However, this approximation is not applicable to the multiband modeling under study here, because, in this case, \(a_1(\omega)\) and \(a_1(2\omega)\) both have large amplitudes of comparable values. Thus, the more general harmonic Volterra function representation presented for two input harmonics must be used. See [12] for the Volterra functions for a higher number of input harmonic tones.
Volterra theory in the frequency domain [12] (see “Harmonic Volterra Theory” for a brief review). This large-signal formalism in the envelope domain is derived from the time-domain Volterra series, which provides a general nonlinear framework to study nonlinear circuits for continuous nonlinear devices. As described in “Harmonic Volterra Theory,” when a continuous nonlinear system of infinite order is excited by a finite number of CW harmonic tones, the output signal at each harmonic can be expanded into a finite number of leading nonlinear RF terms weighted by nonlinear baseband harmonic Volterra functions $V_{b,p}(x_2)$. For modulated excitations, one must also account for the memory effects of the PA, and the harmonic Volterra model given in “Harmonic Volterra Theory” can, in most cases, be accurately generalized by accounting for the dominant $M$ multipath delays:

$$
y_1(n) = \sum_{n=0}^{M-1} V_{11}^n (n-m) \cdot x_1(n-m) + V_{12}^n (n-m) \cdot x_1(n-m) \cdot x_2(n-m)
$$

$$
y_2(n) = \sum_{n=0}^{M-1} V_{21}^n (n-m) \cdot x_2(n-m) + V_{22}^n (n-m) \cdot x_1(n-m)
$$

with $V_{b,p}^m(n) = V_{b,p}^m(|x_1(n)|, |x_2(n)|, \angle \{x_1^*(n), x_2^*(n)\})$.

Using this harmonic model (denoted H2-D-MP) with the test bed described in the following paragraph, we can see in Figure 3 that the introduction of the second-order Volterra RF terms $x_1^2(n)$ and $x_1^n(n) \cdot x_2(n)$ weighted by their associated harmonic Volterra functions $V_{b,p}(x_2)$ leads to a reduction of the model error by 35 and 40 dB for the lower and upper bands, respectively, compared with the conventional 2-D memory polynomial case (2-D-MP) [8].

The test bed setup used for the measurements and linearization of the PA is shown in Figure 4. It features two transmit channels and one receive channel (DISO). A photo of the physical test bed is shown in Figure 5. A broadband PA from Mini Circuit (the ZX60-14012L+ PA) is used as a proof of concept. A 10-MHz bandwidth LTE signal is used as the fundamental signal transmitted in the lower band through the PA channel, and a 20-MHz bandwidth WCDMA signal is transmitted in the upper band. Digital predistortion (DPD) of the upper band cancelling signal is performed within MATLAB. The predistorter is extracted using the method of indirect learning [8] with the same model. The predistorted data are then sent to an Altera Arria V-GT FPGA at a sampling rate of 307.2 MHz, which is used to pass the digital data to two 16-b digital-to-analog converters at a sampling rate of 1,228.8 MHz (interpolation factor of 4), incorporated in a dual-band Texas Instruments TSW30SH84 Tx board. Then, the fundamental signal in the lower band and the cancelling signal (or identifying signal) in the upper band are transmitted simultaneously by the dual-band Tx.

The center frequencies are 978.4 MHz and 1,956.8 MHz for lower-band and upper-band signals, respectively. The output signal of the combiner is received by the TSW1266 Rx board. Within the Rx, the received data

Figure 3. The modeling results for (a) the lower band and (b) the upper band. The augmented harmonic model H2-D-MP (red circles) better predicts the data (blue) than the conventional memory polynomial (2-D-MP) model (green square); this is also indicated by the error terms (purple x’s), which are quite large for the memory polynomial (black pluses).
are downconverted, sampled, and digitized by the analog-to-digital converters at a sampling rate of 614.4 MHz and stored in the FPGA before being passed on to MATLAB for analysis.

A predistorter can now be implemented using inverse learning by exchanging inputs and outputs, while employing the same behavioral model introduced for the PA [8]. This approach works as long as both bands are being broadcast concurrently. As shown in Figure 6, using this predistortion scheme obtains a performance improvement of up to 22 dB in terms of normalized mean-square error (NMSE) and up to 20 dB in terms of adjacent-channel power ratio (ACPR), compared to the conventional dual-band memory polynomial (not including the harmonics terms). Note that a passive filter cannot be used in this situation to remove the harmonic products, because the second band overlaps with the harmonic products of the first band.

Second- and Third-Harmonic Cancellation by a Feedforward Technique

In our second example of self-test and self-linearization, we consider a single-band signal applied to a frequency-agile SDR Tx intended for operation over a broad frequency range. To suppress the unwanted harmonics generated by the PA, using bulky and lossy
switchable filter banks are commonly used in broadband radio. Note that dedicated (nonswitchable) filtering cannot be used because the SDR Tx needs to operate at any carrier frequency in the SDR band. An alternative approach based on active filtering realized using feedforward harmonics cancellation is presented in this section [10], [11].

As depicted in Figure 7, active harmonic cancellation uses an upper-band channel to send a cancelling signal that is in opposite phase (180° phase shift) to the modulated RF harmonics generated by the broadband nonlinear PA from the lower-band signal. The system thus requires careful characterization of the modulated harmonics generated by the lower-band channel [10]. Furthermore, the upper-band channel must be linearized to ensure that the desired harmonic cancellation is achieved and works for broadband modulation. Finally, the PA in the lower-band channel must also be simultaneously linearized to realize fully linear and harmonic-free broadcasting without using switchable passive filters.

Figure 8 shows the test bed for the feedforward harmonic cancellation scheme. A single feedback path is used to acquire, alternatively, the lower- or upper-band signals by selecting the appropriate local oscillator. Note that the fundamental and harmonic RF sources are phase locked. Also, the modulations of the lower band and upper band are fully synchronized. To achieve effective cancellation of the harmonics, two key features are required: 1) the accurate modeling of the harmonics in (RF_out) and 2) the accurate modeling and compensation of the distortion, dispersion, and delays in the cancelling signal (RF_Harm).

To conduct the system identification, two sets of independent signals, \( x_1 \) and \( x_2 \), are transmitted simultaneously through the two channels. As shown in Figure 8, the blue signal RF\(_{\text{fund}}\) sent through the PA channel is for PA channel identification or broadcasting and the red signal RF\(_{\text{Harm}}\) sent through the cancelling channel at the frequency band of the \( p \)th harmonic is used for either the cancelling channel identification or harmonic cancellation.
There are three requirements for the selection of the signal \( x_2 \) to be used for the system identification of the cancelling channel (RF_Harm in Figure 8). First, it must be uncorrelated to the broadcast signal \( x_1 \) used for the main PA channel, because their spectra overlap. Second, because the \( p \)th harmonic distortion extends over \( p \)th times the bandwidth of the lower band, the identification signal \( x_2 \) should have a bandwidth that covers the entire frequency band of the \( p \)th harmonic to be canceled. Third, because the harmonic distortion exhibits an amplitude varying within a specific dynamic range, the cancelling channel should be characterized using a signal \( x_2 \) exhibiting a peak amplitude that is equal to or larger than that of the \( p \)th harmonic generated by the PA for the broadcast signal \( x_1 \).

At the \( p \)th harmonic frequency, the final Tx output signal \( y_2 \) is obtained from the sum of the two channel outputs, each represented by the following Volterra model:

\[
y_2(n) = y_{2,1}(n) + y_{2,p}(n) @ p\omega \quad \text{with}
\]

\[
y_{2,1}(n) = \sum_{m=0}^{M-1} G_{2,m}([x_2(n-m)]^p) \cdot x_2(n-m) \quad \text{for the cancelling signal}
\]

\[
y_{1,p}(n) = \sum_{m=0}^{M-1} V_{1,p}^{m}([x_1(n-m)]^p) \cdot x_1(n-m) \quad \text{for the } p\text{th harmonics,}
\]

where the component \( y_{1,p} \) is the \( p \)th harmonic of the PA and the component \( y_{2,1} \) is the output of the cancelling channel. \( V_{1,p}^{m}([x_1]^p) \) is the harmonic Volterra function weighting the \( p \)th harmonic, and \( G_{2,m}([x_2]^p) \) is the nonlinear gain function of the cancelling channel.

As mentioned previously, the distortion and delay \( G_{2,m} \) of the cancelling channel should be compensated for. DPD can be used to fulfill this purpose. Again, the predistorter is first extracted using the method of indirect learning \cite{9}. Then, the inverse gain \( G_{2,m}^{(i)} \) of the cancelling channel is extracted by exchanging the input and output as shown in the following:

\[
x_2 = \sum_{m=0}^{M-1} G_{2,m}^{(i)} [y_{2,1}(n-m)]^p \cdot y_{2,1}(n-m).
\]

The predistorted signal \( z_2 \) required to produce the harmonic cancellation signal \((-y_{1,p})\) at the PA output is then simply given by

\[
z_2 = \sum_{m=0}^{M-1} G_{2,m}^{(i)} [y_{1,p}(n-m)]^p \cdot (-y_{1,p}(n-m)).
\]

The cancellation of the second and third harmonics generated by a Mini Circuit ZX60-14012L+ PA is presented here as a proof of concept. The test bed used is the same as that described in the previous section. A set of 10-MHz-bandwidth LTE signals is used as the fundamental signal transmitted in the lower band through the PA channel, and a set of 20- or 30-MHz-bandwidth WCDMA signals is transmitted in the upper band through the cancelling channel for the system identification procedure. The carrier frequencies are 888.4 MHz and 1,776.8/2,665.2 MHz for the lower-band and upper-band signals, respectively.

Figures 9(b) and 10(b) compare the original second/third harmonic signal (black line) and the cancellation...
results (red line). The second and third harmonics are suppressed by 30 and 31 dB, respectively. Because the upper band does not impact the lower band, the lower-band signal can still be linearized separately by DPD, as shown in Figures 9(a) and 10(a) (green line).

Table 1 summarizes the measurement results. The experimental verification with a broadband PA yields a reduction of 30/31 dB of the second/third harmonics to 58/59 dBc below the main channel. Simultaneously, the linearization provides −40/−41-dB NMSE and 49.3/50.3-dBc ACPR at the fundamental frequency.

The digitally controlled harmonic cancellation presented in this section may provide an attractive solution in SDR systems because of the frequency-programming flexibility this filterless solution brings to multiband transmission systems. More generally, it provides an example of new types of automatic self-calibration, harmonic cancellation, and linearization that can be implemented with SDR MIMO systems.

**Self-Testing and Mitigation of Impairments in MIMO Systems**

In the previous sections, we reviewed several examples of harmonic cancellation and linearization in SDR Tx systems intended for multiband operation. As mentioned in the opening, besides carrier aggregation and multiband operation, researchers working on 5G are also actively exploring the use of spatial diversity with the development of FD-MIMO systems.

In the literature, the effect of coupling between antennas is presented as crosstalk effect before and after the PA in a Tx chain. Figure 11 depicts a linear and nonlinear crosstalk component in a $2 \times 2$ MIMO case, which adds to the distortion in the output signal. For an example case of $2 \times 2$ MIMO, the MIMO system output can be given as

$$[\bar{y}_1 \bar{y}_2] = [A_{11} A_{12}][h_{1,1} \ h_{2,1} \ h_{1,2} \ h_{2,2}]$$

where $A_{11}, A_{12}$ denotes the observation matrix of inputs and $h_{1,1}, h_{1,2}, h_{2,1}, h_{2,2}$ denote the polynomial model coefficients, depicting the relation between the input and output baseband signals:

$$[h_{1,1} \ h_{2,1} \ h_{1,2} \ h_{2,2}] = \text{pinv}([A_{11} A_{12}])\{\bar{y}_1 \ \bar{y}_2\}.$$  

Most of the literature investigating such models for DPD of MIMO transmission [17] deals with the cross-coupling between PAs, where the coupling effect is

<table>
<thead>
<tr>
<th>Evaluation Criterion</th>
<th>Measured Value 2 $\omega$</th>
<th>Measured Value 3 $\omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harmonic model NMSE</td>
<td>−48.7 dB</td>
<td>−49.5 dB</td>
</tr>
<tr>
<td>Harmonic before cancelling</td>
<td>−28 dBc</td>
<td>−28 dBc</td>
</tr>
<tr>
<td>Harmonic after cancelling</td>
<td>−58 dBc</td>
<td>−59 dBc</td>
</tr>
<tr>
<td>Harmonic cancellation</td>
<td>30 dB</td>
<td>31 dB</td>
</tr>
<tr>
<td>Main channel NMSE after DPD</td>
<td>−41.1 dB</td>
<td>−40.1 dB</td>
</tr>
<tr>
<td>ACPR after DPD (lower and upper sidebands)</td>
<td>−49.3, −49.5 dB</td>
<td>−49.5, −50.3 dB</td>
</tr>
</tbody>
</table>
created artificially using a coupler and two uncorrelated input signals. However, to improve the link quality in a typical MIMO system, diversity techniques are employed. With such diversity techniques, multiple copies of the same signal are transmitted so that the probability of each signal component fading concurrently is reduced. The diversity can be applied at the Tx as well as the Rx [13], [14].

For transmit diversity, we can use space-frequency block codes (SFBCs) or space-time block codes (STBCs) to achieve diversity gain or spatial multiplexing, leading to enhanced robustness and increased channel capacity. One popular STBC technique is known as the Alamouti coding technique [15].

Alamouti code is shown in Figure 12(a); here, a pair of modulated symbols \((m_1, -m_2)\) are mapped on the first antenna in successive time samples. At the same time, the symbol \((m_2, m_1)\) is mapped for the second antenna in successive time samples [14]. In contrast, for the SFBC technique, two successive modulated symbols \((m_1, m_2)\) are mapped on the first antenna in successive time samples.

Maximal ratio combining (MRC) computes the overall estimate as the weighted sum of each estimate across \(N\) receiving antennas with \(G_n\) as gain factor [13]:

\[
\hat{m} = \sum_{n=1}^{N} G_n \hat{m}_n
\]

where

\[
G_n = \frac{E_n}{\sum_{k=1}^{N} E_k}
\]

From (8)–(10), the MRC expression for the best overall estimate of the transmitted symbols is given as [13]

\[
\hat{m} = \sum_{n=1}^{N} H_n \hat{y}_n
\]

\[
\frac{\sum_{k=1}^{N} E_k}{\sum_{k=1}^{N} E_k}
\]

Figure 13 shows the NMSE of the received signal with respect to the signal-to-noise ratio (SNR) of the channel. The impact on NMSE is studied when diversity techniques are introduced for different channels with and without a PA. We can observe that, without the diversity technique, the received signal quality is very poor and leads to a high NMSE. In practical scenarios, there can be multipath effects at the Rx; hence, the performance in the presence of Rayleigh and Rician channels is also depicted. To evaluate the Rayleigh channel, we have considered a mobile Rx; hence, Jakes’ Doppler spectrum applies [16]. It is evident from Table 1 that, due to moving (motion) and multipath, the NMSE increases significantly. Attenuation due to multipath fading can be closely approximated by the Rician model if the LOS component is present. For simulation, Jakes’ Doppler spectrum and a path gain of 1.0645–0.1025 are considered. As a result of the
presence of the LOS component, the NMSE is reduced considerably in this case. There is a requirement for an efficient STBC implementation [13], but SFBC leverages the requirement of efficient implementations due to the actual signal being transmitted through one antenna and conjugation being applied on other. Because of application simplicity, SFBC is preferred as a transmit diversity technique in LTE standards [16].

Once the sequence of the symbol at each antenna is decided according to the diversity technique, the degradation in the signal quality due to PA nonlinearity can be minimized using DPD techniques. Figure 14 illustrates constellation diagrams of the received signal with and without DPD for a practical 2 × 2 MIMO transmission. Table 2 shows the transmission NMSE (baseband data) and error vector magnitude (EVM) of the received data with and without DPD application. The EVM is −30 dBm, which qualifies the mask criterion for a 64-quadrature amplitude modulated (QAM) signal.

Such a simple case of single-user MIMO can further be extended for MU-MIMO. A typical architecture for a generic MU-MIMO is shown in Figure 15. Each antenna is driven by its own pair of transceivers [8]. MU-MIMO operation is controlled at the baseband processing level. Once again, the impact of hardware impairments on the MU-MIMO should be considered. Nonlinear distortions in the Tx path can not only lead to spectral regrowth polluting the adjacent bands but also degrade the in-band (EVM) and symbol-error rate (SER). A recent study reported using a realistic PA behavioral model on the degradation of the SER with output power backoff (OBO) [18]. The SER results obtained for a 2 × 4 MIMO system are shown in Figure 16. For an OBO of 6 dB, the SER clamps to 10^{-4} for an SNR above 15 dB. Under such conditions, the quality of service (QoS) is then limited by the in-band nonlinearity distortion. Thus, because of 1) the high cost and low power efficiency of using 6-dB OBO and 2) the negative impact on the QoS, self-adaptive DPD will certainly remain a remediation technique to consider, even for the relatively low-power (a few watts) base stations expected at mmW.

An additional impairment may also occur in MIMO systems due to the coupling between antennas [19]. Up to now, coupling between the Txs in MIMO systems has been prevented using isolators (see Figure 15). However, the successful development of low-cost mmW MIMO will require that MIMO transceivers be integrated. Such a case would likely preclude the use of bulky isolators. Antenna coupling then becomes an issue that will need to be addressed. Essentially, due to antenna coupling, the various amplifiers will be load-pulling one another, which will induce spectral regrowth while degrading the bit-error rate (BER).

A recent Automatic Radio Frequency Techniques Group paper presented a preliminary in-depth study of this coupling effect [20]. When the coupling between antennas reaches −10 dB, the ACPR was found to decrease by

![Figure 13. The NMSE of the received signal with respect to the SNR of the channel. AWGN: additive white Gaussian noise.](image)

![Figure 14. Constellation diagrams of the received signal after employing the diversity technique when using a PA (a) without DPD and (b) with DPD.](image)

| TABLE 2. A comparison of transmission NMSE and EVM. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Transmission Case | First Branch EVM | Second Branch EVM | EVM After MIMO Operation |
|                  | Without DPD     | DPD             | Without DPD     | DPD             |
| SISO case        | −8.95           | −16.26          | −8.95           | −16.26          | −17.14          | −23.89          |
| MIMO case        | −8.47           | −20.95          | −21.10          | −28.48          | −20.23          | −30.94          |
4 dB and the EVM to increase by 10%. Even more severe, the BER was found to degrade from $10^{-5}$ to $10^{-1}$ when the coupling varied from $-17$ to $-10$ dB, as shown in Figure 17. The experiments were conducted for a 64-QAM, 60-Mb/s signal with 20-MHz bandwidth. A good agreement was obtained between experiments and simulations. Note that the two amplifiers were directly coupled.

Measurements with real antenna arrays should be pursued for further studying this effect in FD-MIMO systems involving multiple amplifiers. Methods to reduce coupling between antennas, which are typically a half-wavelength apart, have already been reported using different polarization [21]. Nonetheless, this preliminary work supports the notion that the impact of antenna coupling and associated PA load-pulling should be considered in future studies in the context of MMIC integration of a complete MU-MIMO Tx operation.

Conclusions
As we emphasize in this review, researchers in the field of future wireless systems such as 5G are exploring multiple directions to increase data rates, including combinations of spatial diversity via FD-MIMO and multiband operation with the identification of new spectra both below 6 GHz and at mmW frequencies. This also includes the cooperative use of unlicensed bands (wireless local area networks) for reducing the cellular...
traffic in high-user-density areas. Spectrum flexibility will, therefore, certainly be a key component in future wireless systems and may eventually lead to the integration of some of SDR concepts, particularly for operation below 6 GHz.

In this review, we argue that—whether for multi-band operation, carrier aggregation, or MU-MIMO systems—built-in self-testing is expected to be used to calibrate, linearize, and possibly perform tasks such as active filtering/cancellation of unwanted modulated harmonics. In this regard, it is interesting to note that the use of self-testing in SDR MIMO systems makes them operate similarly to a generalized nonlinear vector network analyzer (NVNA).

Let us briefly consider the capabilities of modern NVNAs. These are test instruments that acquire the amplitude and phase of the fundamental and harmonics of the incident and reflected waves to reconstruct their full voltage and current waveforms. Mixer-based NVNAs, such as the PNA-X, perform this task while exhibiting extremely high dynamic ranges (>100 dB). This high dynamic range is achieved by using strictly periodic (CW or modulated) waveforms and narrowband Rxs for low noise acquisition. Gigahertz modulation bandwidth can still be efficiently acquired by using frequency stitching, owing to the signal periodicity and fast frequency scanning. It can be concluded that the PNA-X can operate as a vector signal analyzer (VSA) not only for the fundamental but also for the harmonics.

Although frequency-agile SDR MIMO systems can also acquire the amplitude and phase of harmonics for modulated signals, their dynamic range is usually reduced. Indeed, they rely on Rxs with larger bandwidth that are intended to perform single-shot acquisition for nonperiodic waveforms. Today, for calibration purposes, periodically modulated waveforms could also be used in SDR systems, permitting them to achieve higher dynamic ranges via signal averaging. Clearly, we are reaching the point where the self-testing that can be performed by frequency-agile SDR systems becomes akin to the operation of both NVNAs and VSAs. Armed with these self-testing attributes alone, wireless MIMO systems are able to calibrate themselves and broadcast via various predistortion or feedforward schemes waveforms, which approach the ideal targeted ones despite the SDR hardware impairments.

Given current trends at low frequencies and the ongoing movement in 5G research, it can be expected that built-in self-testing, linearization, and adaptation will continue to play an important role for maintaining QoS, as well as to provide new power-efficient solutions in support of a wirelessly interconnected society that is more eco-friendly.

References