The current trend towards increased data rates in mobile services has direct implications for the power amplifiers operating in these systems, which now need to be not only efficient but also wideband and, above all, linear. To operate efficiently, an amplifying device needs to see its optimum terminations at the input and output. Although basic amplifier theory suggests that an almost infinite number of harmonics need to be controlled to establish a particular device class of operation (such as (inv.) Class B, E, F), in practice, control of the fundamental, second and third harmonic is most of the time sufficient to achieve high performance. Furthermore, in order to operate linearly, the active device requires—in addition to a proper control at its fundamental and harmonic impedances—the correct baseband or envelope impedance, which, in the circuit implementation, is mostly determined by the biasing networks at the device input and output.

Since active devices operate in a large-signal/nonlinear mode when reaching high efficiency operation, the properties of the device under test (DUT) cannot simply be extrapolated from small-signal measurements. Consequently, active devices intended for large-signal operation need to be characterized under the exact same conditions as in their final application. To address this need, harmonic load-pull measurement systems have become a widely established tool. In these systems, the (harmonic) reflection coefficients offered to the device under test are “physically” varied to find the optimum (harmonic) loading conditions at the device input and output. With this optimum set of loading conditions the final matching circuit for the active device can be designed. Load-pull systems can be divided into two main categories: passive and active.

**Passive Tuner-Based Systems**

A typical passive load-pull configuration, shown in *Figure 1a*, makes use of sliding mechanical probes to tune the reflection coefficient. The bias tee placed after the tuner reduces the impact of its losses, but yields larger memory effects due to the baseband inductance. These kinds of systems are currently still the preferred industry large-signal test bench choice, due to their simplicity and high power handling capabilities. Unfortunately, passive load-pull setups are constrained by losses in the tuners, connecting cables and wafer probes, which limit the magnitude of the reflection coefficients that these systems can provide to the DUT.

Mauro Marchetti
Anteverta Microwave B.V.
Delft, The Netherlands
ACTIVE LOAD-PULL SYSTEMS

Active systems can, thanks to the use of amplifiers, overcome the reflection magnitude restrictions due to losses. Two main types of active systems can be identified, namely:

- Closed-loop active systems achieve their desired reflection coefficient by coupling out of the main path (using directional couplers) a portion of the wave generated by the DUT, adjusting it in amplitude and phase and then injecting it back as an incident wave to the active device (see Figure 1b). Here, the bias tee is placed as close as possible to the DUT to reduce unwanted memory effects. Since the injected signal is a direct function of the coupled wave, the loop gain and phase shift determine the reflection coefficient independently of power and spectral content of the signal. This makes the closed-loop concept suitable for fast device characterization, but also prone to oscillations since the loop gain cannot be very selectively controlled over frequency. To avoid the risk of oscillation, most closed-loop systems include sophisticated filtering.

- Open-loop active systems instead do not reuse the waves generated by the DUT, but directly inject a signal into the DUT, which is phase coherent with the input signal to the device, to compose a desired reflection coefficient (see Figure 1c). The desired reflection coefficient is created by adjusting the amplitude and phase of the injected wave. Here also, the bias tee is placed as close as possible to the DUT, to reduce the memory effects caused by bias path inductance. Since the injected signals are no longer a modified version of the waves generated by the DUT, the realized reflection coefficients are power and signal dependent. Therefore, iterations are needed during the measurements to find the optimal injection signals to offer the desired reflection coefficients to the DUT.

ELECTRICAL DELAY IN LOAD-PULL SYSTEMS

When working with modulated signals, for a well controlled linearity behavior of the DUT, the reflection coefficients offered to the DUT should ideally be constant (not vary versus frequency) within the modulation bandwidth at the fundamental, as well as in all related frequency bands at baseband and harmonic frequencies. This situation is approximated in real circuit implementations, where the matching networks are placed directly at the reference planes of the active device.

In conventional load-pull setups, however, the actual physical impedance is always located at some distance from the DUT (see Figure 2), which is much larger than for any practical matching network. This distance, as well as any physical length within the tuning element itself (such as the length of the active feedback loop, or the position of the probe in mechanical tuners), yields very large electrical delays causing rapid phase changes of the reflection coefficients versus frequency. Typical values for these phase fluctuations can vary from approximately 3°/MHz for a passive mechanical tuner to 30°/MHz in a conventional active closed-loop system (see Figure 3). Note that based on these values for a W-CDMA signal with adjacent and alternate channels (total bandwidth of 25 MHz), the reflection coefficient offered by a passive tuner would vary 75° in phase over this bandwidth, while in a conventional active closed-loop the phase change of the offered reflection coefficient would amount to 750°.

It is clear that these large phase deviations represent nonrealistic circuit conditions and will cause measurement errors such as IM3 asymmetry, spectral re-growth and EVM degradation. In general, maintaining the reflection coefficients constant over frequency is getting more and more difficult with the increase in modulation bandwidth of communication signals, not only in practical circuits, but definitely in load-pull measurement setups.

MIXED-SIGNAL LOAD PULL FOR WIDEBAND IMPEDANCE CONTROL

To overcome the aforementioned problems of losses and electrical delay

Fig. 1 Typical load-pull configurations.

Fig. 2 Phase delay caused by electrical lengths of cables/adaptors/probe.

Fig. 3 Reflection phase rotation at the DUT reference plane.
in conventional load-pull characterization systems, while being able to work with realistic wideband communication signals, some major measurement system innovations are required. The principle of the proposed wideband open-loop active load-pull approach is shown in Figure 4. When the non-linear DUT is excited with a user-defined modulated signal $a_x$, it generates signals in the baseband, fundamental and higher harmonic frequency bands. By measuring the device reflection coefficient at every frequency, the waves to be injected are estimated at every iteration. When the required reflection coefficient versus frequency (at every controlled band) is achieved, the iteration has converged and the large-signal parameters (that is power added efficiency, output power, intermodulation distortion, etc.) are measured. To address these needs a novel open-loop system was developed\(^6\) employing wideband signal generation and signal acquisition. A simplified block diagram of such a system is given in Figure 5.

**SIGNAL GENERATION**

When working with complex modulated signals, a good place to start is the modulation test standard.\(^7,8\) According to the test standard, a test signal is created, which consists of a finite sequence of IQ data samples specified in the time domain. In conventional lab instrumentation such as vector-signal generators, this sequence for a given standard (such as W-CDMA) is typically embedded in the instrument. In testing operation, these signals are uploaded in internal arbitrary waveform generators (AWG) and up-converted with IQ mixers yielding the modulated RF signal. In practice, these test records are sequentially repeated, yielding in the frequency domain a large but finite number of discrete spectral components. More precisely, the number of samples, in combination with the sampling speed...
When testing a nonlinear device with modulated signals, the measured waveforms will include not only the fundamental frequency band, but also the baseband, higher-order harmonics and intermodulation distortion products (see Figure 4). Since the reflection coefficient is defined as the ratio of two travelling waves, the effective control of the impedances over the whole modulation bandwidth offered to the DUT can be enforced by establishing the desired linear ratios of the incident and device generated waves on the DUT versus frequency. Or, in other words, a signal must be injected, at all the frequency components of interest, according to the linear relation,

$$
\Gamma_{x,n} = \frac{a_{x,n}(f_n)}{b_{x,n}(f_n) \cdot \Gamma_{x,n}(f_n)}
$$

which allows one to analyze modulated communication signals like “classical” multi-tone signals, but now with a very large number of frequency tones.

**SIGNAL ACQUISITION**

A fraction of the waves incident and reflected by the DUT is taken by the directional couplers and fed to high linearity mixers that down-convert the signals for digitization after which an FFT is applied to obtain their spectral content. Also here the sampling speed of the A/D converter and the time span used for the data-acquisition set an effective frequency bin size, which must be compatible with the original applied test signal, as described by the following expression

$$
\Delta f_{A/D} = \frac{f_{s,A/D}}{N_{A/D}} = \frac{\Delta f_{AWC}}{k} = \frac{1}{kT_{MOD}}
$$

where $\Delta f_{A/D}$ is the resulting frequency bin size of the acquired signals, $f_{s,A/D}$ and $N_{A/D}$ are respectively the sampling frequency and the number of samples used by the A/D converters, and $k$ is an integer.

**CONTROLLING THE REFLECTION COEFFICIENTS**

When testing a nonlinear device with modulated signals, the measured waveforms will include not only the fundamental frequency band, but also the baseband, higher-order harmonics and intermodulation distortion products (see Figure 4). Since the reflection coefficient is defined as the ratio of two travelling waves, the effective control of the impedances over the whole modulation bandwidth offered to the DUT can be enforced by establishing the desired linear ratios of the incident and device generated waves on the DUT versus frequency. Or, in other words, a signal must be injected, at all the frequency components of interest, according to the linear relation,

$$
\Gamma_{x,n} = \frac{a_{x,n}(f_n)}{b_{x,n}(f_n) \cdot \Gamma_{x,n}(f_n)}
$$

where $a_{x,n}$ and $b_{x,n}$ are the incident and reflected waves at port $x$ and harmonic index $n$, while $\Gamma_{x,n}$ represents the user defined reflection coefficient versus frequency for port $x$ and harmonic index $n$. As in the classical open-loop approach, the new b-wave, containing all the frequency components of the signal of interest (that is more than 23000 frequency tones for a W-CDMA signal in a 35 MHz band), can be created from scratch, and any desired reflection coefficient behavior versus frequency can be created by adjusting the amplitude and phase of the injected waveform independently at each frequency component of interest.

Although conceptually simple, this method requires very fast highly linear data acquisition with high dynamic range to measure the reflection coefficients at every spectral component of the complex modulated signals with its related distortion products. Secondly, the a-waves need to be generated arbitrarily with a high dynamic range. Third, the signal injection at the various ports needs to be phase coherent at both the RF frequencies as well as at the baseband.

**PRACTICAL IMPLEMENTATION**

The block diagram of the system implementation is shown in Figure 5. The wideband reflection coefficients at the DUT reference planes are measured by coupling the traditional mixer-based super-heterodyne down-conversion used in conventional VNAs with wide-
band A/D converters (100 MHz sampling frequency). This hardware configuration enables the measurement of the device reflection coefficients over a wide bandwidth in a single data acquisition interval. Wider bandwidths, up to 120 MHz, can be measured by stepping the frequency of the local oscillator used for down-conversion.

The fundamental and harmonic frequency loads are synthesized by injecting fully coherent RF signals that are generated by IQ up-converted baseband signals, which are provided by (200 MS/s) arbitrary-waveform-generators (AWG). Custom bias tees with low inductance are placed directly at the DUT reference planes in order to minimize the electrical delay of the baseband (BB) impedance, implemented here as a passive impedance switch bank. On the baseband board also, the low-frequency couplers for the baseband impedance measurement are implemented. Note that in a more extended version, the baseband impedance can also be made active and controlled by an additional AWG.

**WIDEBAND MEASUREMENT RESULTS AND COMPARISON**

To demonstrate the unique capabilities of the setup, a NXP Gen 6 LDMOS device with a gate width of 1.8 mm was measured. In the experiments, the drain current and voltage are set to 13 mA and 28 V, respectively. In this experiment, the optimum fundamental load and source matching conditions are found to be $\Gamma_{L,1} = 0.6, 45^\circ$ and $\Gamma_{S,1} = 0.5, 90^\circ$, while the input and output baseband impedances are set to a short condition. In addition, the input and output 2nd harmonics are set to circuit-like open conditions ($\Gamma_{L,2} = \Gamma_{S,2} = 0.95$) to optimize the efficiency.

To highlight the excellent controllable bandwidth and the electrical delay-free operation of the new measurement setup, a comparison is made with a previously developed state-of-the-art active harmonic load-pull system, which was especially optimized for minimum electrical delay. For this purpose, a two-channel W-CDMA signal is used (centered at 2.135 and 2.145 GHz) and the input and output reflection coefficients in the newly developed setup are set to the following two cases:

- without electrical delay
- with an electrical delay of 4.85°
MHz for the fundamental source and load and 4.6%/MHz for the 2nd harmonic source and load.

Figure 6 illustrates the source and load matching conditions in the fundamental (2.1225 to 2.1575 GHz) and harmonic (4.245 to 4.315 GHz) frequency range provided to the active device under test, with electrical delay (open symbols) and without electrical delay (filled symbols). Note that the filled markers represent the source and loading conditions for the two-carrier W-CDMA signal without any electrical delay, yielding completely overlapping points in the Smith chart. As shown in Figure 6, for the case with electrical delay, the fundamental load trajectory has been shifted, such that the optimum matching condition is now centered at 2.135 GHz. This was required to avoid the unstable region of the active device.

It is important to note that this is a comparison to the “best possible case” of a classical closed-loop active load-pull system, since practical closed-loops will be subject to amplitude variations within the control frequency bands. Moreover, phase change can be even worse if high-Q resonators are used to control the harmonic terminations.

The measurement results are summarized in Table 1. There is significant performance degradation for the measured active device when the electrical delay is present in the reflection coefficients. This is also evident from Figure 7, which shows the power spectral density at the device output reference plane for the fundamental and 2nd harmonic frequency bands. Note that a 5 dB output power drop and close to 8 percent degradation of the power added efficiency (PAE) can be observed, when compared to the case with no electrical delay. It is important to stress that this performance degradation is a measurement artifact, the result of an incorrect measurement, which would mislead the user in judging the device performance.

ACTIVE LOAD INJECTION POWER AND LINEARITY

To provide the DUT with a specific $\Gamma_L$, a certain injection power is needed, which not only depends on the output power of the DUT and the desired $\Gamma_L$, but also on the output impedance of the device. When considering high-power devices, with output impedances in the order of few Ohms, the required injection power to cover the desired Smith chart area can be extremely high (2 to 10 times higher than the maximum output power of the DUT). To overcome this issue, typically, a pre-matching is used, which converts the 50 Ω impedance of the system to a value that is much closer to the output impedance of the DUT, thus reducing the power requirement of the load injection amplifier.

To give an example, a DUT with an output impedance of 2 Ω and an available output power of 200 W requires, when the system impedance is pre-matched to 10 Ω, an injection power of 360 W to synthesize a load impedance of 1 Ω. Reducing the system pre-matched impedance to 5 Ω lowers the required injection power for the same load condition to 142.2 W.

When considering multi-tone or modulated signals, the situation becomes more complicated as the linearity of the injection amplifier needs to be taken into account. To study the linearity constrains on the injection amplifier, consider a two-tone test signal, for which the power injected by the load amplifier at the IM3 frequencies of the two-tone test signal is given by

$$P_{a_i,IM3} = 3P_{a_i, fund} - 2P_{a_3,a_2} = 3P_{b_3, fund} \left(1 - \left|\frac{\Gamma_{DUT}}{\Gamma_{SYS}}\right|^2 \right) \left|\frac{Z_{DUT} + Z_0}{Z_{SYS} + Z_0}\right|^2 \left|\frac{Z_L - Z_{SYS}}{Z_{DUT} + Z_L}\right|^2 - 2P_{b_3,a_2}$$

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Without electrical delay</th>
<th>With electrical delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{\text{out}}$, Ch. 1</td>
<td>20.3 dBm</td>
<td>20.5 dBm</td>
</tr>
<tr>
<td>$P_{\text{out}}$, Ch. 2</td>
<td>20.6 dBm</td>
<td>15.4 dBm</td>
</tr>
<tr>
<td>ACLR1, Ch. 1</td>
<td>-43.9 dBc</td>
<td>-43.0 dBc</td>
</tr>
<tr>
<td>ACLR2, Ch. 1</td>
<td>-42.2 dBc</td>
<td>-41.6 dBc</td>
</tr>
<tr>
<td>ACLR1, Ch. 2</td>
<td>-42.1 dBc</td>
<td>-41.8 dBc</td>
</tr>
<tr>
<td>ACLR2, Ch. 2</td>
<td>-39.6 dBc</td>
<td>-39.2 dBc</td>
</tr>
</tbody>
</table>

Figure 7 Measured output power spectral density vs. frequency of a NXP GEN6 LDMOS device: (a) at fundamental and (b) at second harmonic.
where \( Z_{DUT} \) and \( P_{av,sfund} \) are the output impedance and the available power coming out of the DUT (see Figure 8). The load impedance offered to the DUT at the reference plane is varied by adjusting the equivalent voltage source \( E_{SYS} \), which is the passive load impedance at the DUT reference plane; \( P_{av,sfund} \) and \( IP_{3a} \) are the power injected by the load amplifier and its output third-order intercept point, respectively.

**Figure 8** shows the results of a harmonic balance simulation, where the apparent IM3 of the DUT versus decreasing output IP3 of the injection amplifier is shown for different pre-matching conditions of the system impedance. The dotted line is the actual IM3 level as would be achieved with passive matching techniques. The dot-dashed line represents the IM# level due to the \( P_{av,IM3} \) as approximated by Equation 1. A polynomial model was used for the amplifier linearity. In this experiment, the same DUT is used as for the single-tone considerations; \( P_{av,s} = 200 \text{ W} \), output impedance = 2 \( \Omega \), which is set in the simulation to have an output IP3 of 63 dBm. For this device the output power is set equal to 50 W per tone, to have the same peak voltage as in the single-tone case. These conditions yield an actual IM3 of the DUT of -30.35 dBc. From Figure 9, it can be seen that this level (a correct measurement) is only achieved for sufficiently high IP3 of the injection amplifier. When the injection amplifier is less linear, it will introduce significant IM3 products, which can be approximated by Equation 4, and are also plotted in Figure 9. Note that IM3 cancelation effects can also occur. Consequently, to have reliable linearity measurements in a conventional active load-pull setup, even when pre-matching is used, the injection amplifier linearity (and thus its peak power) needs to be at least 10 times higher than that of the DUT.

It is obvious that at high power, these amplifiers, if available, will be extremely expensive. For this reason, active load-pull systems that can offer communication standard compliant device testing, for example W-CDMA, at base station power levels (100 W and above) have not yet been demonstrated.

In a mixed-signal load-pull system, an iteration procedure is performed to obtain a user-defined reflection coefficient at each individual frequency component of a complex modulated signal, in and out of band. Due to these iterations, the injection amplifier is basically pre-distorted for its own nonlinearities, which allows the use of injection amplifiers with a much lower linearity than what is typically required in conventional active load-pull systems. As an example, consider Figure 10, which shows the ACPR and average PAE for a single-channel W-CDMA signal at 2.14 GHz, with a peak to average ratio of 9.5 dB. In these experiments the injection amplifier used has a maximum saturated power rating of only 200 W with an associated 60 dBm output IP3, while the reflection coefficient is controlled over a 35 MHz frequency span.

**CONCLUSION**

Mixed-signal techniques have been applied to extend the capabilities of traditional active load-pull setups. The realized system provides an unprecedented measurement speed, high dynamic range and is currently the only system that can handle communication standard compliant signals that are truly wideband such as multicarrier W-CDMA. The ability to eliminate losses and electrical delay, while being completely free in defining the source and load reflection coefficients versus frequency, allows perfect mimicking of in-circuit situations, making the system a tool of fundamental importance for the RF power amplifier developer.

**References**