Overcoming the Challenges of On-Wafer Load Pull Measurements at Millimeter-Wave Frequencies for 5G Applications

APPLICATION NOTE / 5A-070

Authors: Steve Dudkiewicz (Vice President, Marketing & Business Development) and Richard Hilton (Applications Engineer), Maury Microwave Corporation, Ontario, California.

This reprint is based on an article that originally appeared in Microwave Journal magazine – digital edition of June, 2018.
**5G at Millimeter-Wave Frequencies**

Fifth generation mobile networks, also known as fifth generation wireless systems and more commonly referred to simply as 5G, represents the next evolution in wireless communications. With an emphasis on connectivity, 5G is expected to bring together data, voice, video, Internet-of-Things, connected cars, smart homes, smart cities, augmented reality, industrial automation and more.

5G will address this aggressive plan by deploying technologies over multiple frequency bands, from low MHz to high GHz, with much research being done at the longer-range 450 MHz to 6 GHz bands and the higher data rate 28-30 GHz and 37-39 GHz millimeter-wave bands.

While having their own unique challenges, the millimeter-wave bands promise to bring many advantages to wireless communication including larger bandwidths, faster data rates and greater capacity, increased security and privacy, and longer battery life.

A critical part of the infrastructure required to enable the benefits associated with 5G millimeter-wave technologies is the power amplifier, or PA, which must be properly designed for optimum performance. The expectation on PAs include maximizing power and efficiency while maintaining an appropriate linearity, and a PA engineer must take this into consideration during the design phase. A useful tool in PA engineer’s arsenal which can help achieve optimum performance is referred to as Load Pull.

**Load Pull Techniques**

Load pull is the process of changing the load impedance presented to a device under test (DUT), commonly a transistor, in order to measure its performance characteristics under varying large signal conditions. Impedances are systematically changed while parameters such as output power, gain and efficiency are measured or calculated. Contours representing fixed performance values (i.e. output power of X dBm or efficiency of Y%), as shown in Figure 1, are then plotted to visualize the point of maximum performance, the rate at which the performance changes, and the trade-off between various parameters.

![Figure 1. Example contours for Pout and PAE](image.jpg)
But how does load pull work? First, consider a DUT as a two-port network as shown in Figure 2.

![Figure 2. Two-port representation of DUT](image)

Next, consider that a signal $a_1$ is injected into port 1 of the DUT whereby a portion of the signal is delivered to the DUT while another portion is reflected as $b_1$ due to the mismatch between the input impedance of the DUT and the source impedance of the input network. Furthermore, consider a modified signal $b_2$ that exits port 2 of the DUT whereby a portion of the signal is delivered to the load while another portion is reflected as $a_2$ due to the mismatch between the output impedance of the DUT and the load impedance of the output network.

The magnitude of reflection, represented as $\Gamma_L$, is calculated as $\Gamma_L = \frac{a_2}{b_2}$. Load pull changes the magnitude of reflection presented to the load of the DUT by changing the reflected signal $a_2$, and the same applies to the phase of the reflection signal. In other words, any load impedance, which can be calculated as $Z = Z_0 \sqrt{\frac{1 + \Gamma_L}{1 - \Gamma_L}}$, can be presented to the DUT as long as the signal $a_2$ can be achieved.

There are two common methodologies to vary the impedance presented to a DUT: passive load pull and active load pull.

Passive load pull uses mechanical impedance tuners to change the magnitude and phase of the reflected signal $a_2$ and hence vary the impedance presented to the DUT, as shown in Figure 3.

![Figure 3. Output network of a simple passive load pull setup](image)

The magnitude and phase of the load impedance are adjusted by varying the position of a probe (or slug) in both X and Y axis along a 50Ω airline, as shown in Figure 4. The magnitude of the reflection is controlled by moving the probe vertically within the airline, while the phase is controlled by moving the probe horizontally along the airline. By moving the probe up and down, left and right, it is possible to present nearly any impedance to the DUT, as long as the magnitude of $a_2$ can remain sufficiently large so that $\Gamma_L < 1$ since $a_2$ will always be smaller due to the losses between the output of the DUT and the tuner.
Open-loop active load pull, shown in Figure 5, does not rely on a mechanical tuner to reflect part of $b_2$ back as $a_2$, rather it uses a signal generation with magnitude and phase control to create a new signal $a_2$, so that when amplified by an external amplifier, any $a_2$ and hence any $\Gamma_L$ can be achieved.

While at first glance, active load pull may seem superior to passive load pull as it has no theoretical $\Gamma_L$ limitation, the practical limitation revolves around the amount of power required to achieve the signal $a_2$ that is actually delivered to the output of the DUT.

Active tuning has a number of advantages over passive tuning including speed enhancements and increased smith chart coverage (in theory to a gamma >1). This is due to there being no mechanical moving parts and that as the $a_2$ wave is directly generated, the ratio of $\frac{a_2}{b_2}$ can equal greater than 1, the limitation being maximum output power of the amplifier.

Referring back to Figure 5, the mismatch between the 50Ω amplifier and non-50Ω DUT will cause a portion of the signal to be reflected back towards the amplifier, and the larger the mismatch, the larger the portion of the signal that is reflected. Under extremely mismatched conditions, it’s feasible that only 10% of the signal available to the output of the DUT will be delivered to the output of the DUT, and therefore a much larger amplifier may be needed.

Hybrid-active load pull overcomes this challenge by pre-matching the DUT impedance from highly-mismatched to moderately mismatched, and therefore lowers the power required to deliver the same signal $a_3$ to the output of the DUT.

When performing load pull, it is preferable to be able to close measurement contours to ensure the DUT’s maximum performance has been achieved. Without closing contours, it is possible that a superior performance could be missed, and a wrong conclusion formed.
Millimeter-wave Load Pull

Modern Gallium Nitride (GaN) transistors have an output impedance around 1-2 Ω, which can be represented as around $\Gamma=0.96$ and 0.92 respectively. Therefore to ensure closed contours, it’s important to have the ability to present magnitudes of reflection higher than the DUT’s output impedance.

With a passive load pull system, the net magnitude of reflection achievable at the DUT reference plane can be calculated as

$$RL_{tuner} + RL_{coupler+probe} = RL_{dut}$$

$$RL_{tuner} = -20\log\left(\frac{VSWR_{tuner} - 1}{VSWR_{tuner} + 1}\right)$$

$$RL_{coupler+probe} = 2(IL_{coupler+probe})$$

$$\Gamma_{dut} = 10^{\left(-\frac{RL_{dut}}{20}\right)}$$

Assuming a typical tuner VSWR and coupler, cable and probe losses at 30 GHz, $VSWR_{tuner}=20:1$, $IL_{coupler+probe}=2.5$ dB, the maximum achievable magnitude of reflection is reduced from $\Gamma=0.9$ at the tuner reference plane to $\Gamma=0.5$ at the DUT reference plane.

Figure 6 shows actual passive load pull measurement data of a GaN transistor on-wafer at 30 GHz with a maximum output power of 30.66 dBm. Notice how the contours do not close, and it is uncertain to what level the transistor would perform if further tuning could be performed.

As described above, hybrid-active load pull overcomes the limitations in magnitude of reflection of passive load pull by combining an active injection signal to increase $a_e$ and therefore increase $\Gamma$, as shown in block diagram Figure 7 and pictured in a commercially available hybrid-active load pull system in Figure 8.
Figure 7. Block diagram of hybrid-active load pull system.

Figure 8. Commercially available millimeter-wave on-wafer hybrid-active load pull system.
The formula which governs the relationship between the transistor, the system impedance, the injection power and the tuning range can be described as

\[
Z_L = \frac{Z_{Sys} + K Z_{DUT} - \sqrt{(Z_{Sys} + K Z_{DUT})^2 - (1 - K)(Z_{Sys}^2 - K Z_{DUT}^2)}}{1 - K}
\]

Where

- \(Z_L\) is the impedance presented to the DUT
- \(Z_{Sys}\) is the system impedance
- \(Z_{DUT}\) is the DUT’s output impedance
- \(K\) is defined as

\[
K = \frac{P_{a2}}{P_{b2}} \cdot \frac{1 - |\Gamma_{Sys}|^2}{1 - |\Gamma_{DUT}|^2} \cdot \frac{|Z_{Sys} + Z_0|^2}{|Z_{DUT} + Z_0|^2}
\]

- \(P_{a2}\) is the active tuning power injected into the output of the DUT at the DUT reference plane
- \(P_{b2}\) is the DUT’s output power
- \(Z_0 = 50\) ohm

And the net magnitude of reflection achievable at the DUT reference plane can be calculated as

\[
\Gamma_L = \frac{Z - 50}{Z + 50}
\]
With a driver amplifier of 40 dBm, and using the same passive impedance tuner to transform the system impedance from 50Ω to 23.17+28.12J Ω, it is possible to achieve $\Gamma=0.85$ and successfully close the output power contours. The contours shown in Figure 9 clearly demonstrate that a maximum output power of 31.12 dBm could be achieved by the same GaN transistor, 0.46 dB, or 12% more power than initially determined through passive load pull and incomplete contours.

Figure 9. Hybrid-active on-wafer load pull measurements at 30 GHz

---

**Conclusion**

As companies accelerate their development of 5G technologies, and compete on the market for best-in-class solutions, optimizing power, efficiency and linearity will become more and more prevalent. Small advantages of a few dB in power or a few percentage in efficiency may mean the difference between best-in-class and never-was.

Hybrid-active load pull helps overcome the challenges with millimeter-wave PA design by removing the uncertainty of unclosed contours and guaranteeing ideal matching, and gives those that adopt the methodology an edge in the marketplace.