New test methodologies for 5G wafer high-volume production

By Daniel Bock, Jeff Damm [FormFactor, Inc.]

Companies developing 5G technologies are racing to develop the first chipsets in order to set the standard of deployment and be the leader. While initial standards for 5G were set at the end of 2017 [1], and there are ideas about the applications of 5G (**Figure 1**), it is still unclear how exactly it will all come together. This uncertainty demands unprecedented levels of collaboration and partnership. This article explores the challenges and changes in test methodology of 5G devices, and showcases the results of a collaboration with Intel [2].

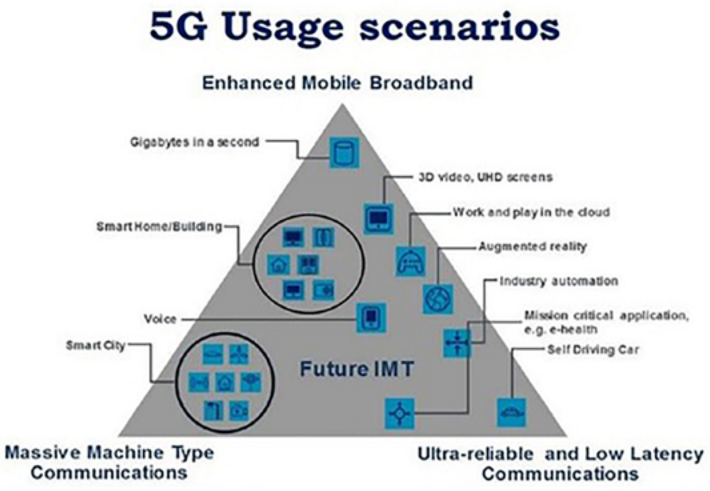

Figure 1: Usage scenarios for 5G [3].

5G brings three technical improvements/enhancements when compared to the current deployed, 4G standard:

1. Greater available bandwidth, increasing to more than 4GB per connection per month from today's less than 1.5GB per device [4].
2. Lower latency, for critical applications to be more responsive [5].
3. The ability for up to 1 million devices (such as sensors and smart devices, per square kilometer [6]) to be connected to the network.

Some of the resultant solutions to meet these requirements for 5G include:

- The opening of millimeter wave (mmW) frequencies: ~30GHz and above;
- The increase in the number of mobile sites to allow for more devices [7]; and
- The deployment of edge cloud nodes so that data doesn't always need to go back to the central node.

Because of all of these changes, the full network infrastructure will need to be upgraded. Prior to 5G, the upgrades were primarily around changing from analog to digital in the first few generations, as well as improving modulation techniques [8]. "So the scale of 5G when compared to previous upgrades to the infrastructure is much larger than seen in the past because of the technology extensions. However, as Scott Fulton points out, "5G is a capital improvement project the size of the entire planet, replacing one wireless architecture created this century with another one that aims to lower energy consumption and maintenance costs [9]."

In order to provide the chips required for this change in the landscape, there will be a large number of changing requirements in wafer test that come out of these architectural requirements. FormFactor partnered with Intel to investigate these changes, and tested one such example of a new test methodology [2].

The history of mmW testing

Historically, millimeter wave testing of wafers was relegated to labs and low-volume production for defense, aerospace and other somewhat-exotic applications. This is because of the low transmission range, high cost of generating RF signals with IC chips, and low data rates that were required. Therefore, wafer high-volume manufacturing (HVM) production floors topped out at 6GHz because mobile phones were the devices using a majority of the RF ICs.

Millimeter wave testing, however, has been moving into high-volume production because of automotive radar and high-speed digital parts that require the increased performance gained with higher frequencies. That is either with more accurate resolution of a nearby vehicle or obstacle, or more data being moved in data centers and over fiber connections. Some of the challenges identified in automotive test [10] include:

1. Power accuracy;
2. Maintaining RF calibration of the test equipment (and final RF signal path);
3. Setting appropriate test limits at these higher frequencies;
4. Millimeter "anything" is just more expensive; and
5. Test engineering is not familiar with mmW testing.

Following the path laid by automotive radar applications, 5G is going to be pushing semiconductor test developments and will expand them because of the unique challenges with higher channel counts and good signal integrity. This article provides what manufacturers—and in particular, test engineers—need to know about changes in testing for mmW wafers used in 5G. It reviews some of the changes that will be required to support wafer testing in the bands starting at 26GHz and all the way up to the 67GHz range in HVM, and will discuss the advancements in probe card technology that enable multi-site production-level testing.

Test protocols that are going to be developed for 5G will need to handle multi-site calibration and testing with minimum cross talk, as well as handle the large number of RF channels due to phased-array antennas for beam forming. Essentially, testing needs to be designed specifically for multi-site production tests with the limited number of tester channels that are available. In short, production testing of these parts has just gotten dramatically more expensive, with costs easily increasing by a factor of 2 (or more).

The new challenges for 5G production-level test

The emergence of 5G is changing the landscape for RF production compared to what we saw over the past 30-40 years. We explore the major requirements below.

Higher frequencies with more channels per device. The first big shift that is affecting the test cell is the new, higher speed channels operating up to almost 70GHz. Traditional HVM wafer testers are capable up to 6GHz, which covered all of the bands being used by 4G. In addition, these testers could be extended to higher frequency with custom extenders, adding 1-2 channels at these frequencies. The 5G devices that are being developed will include up to 64 RF channels per device under test (DUT) in the highest channel count. This means that a traditional tester is not the best solution because of the frequency limitations. Semiconductor manufacturers are requesting new testers with more channels that are more capable than current off-the-shelf testers. While this is a possible solution, it should also be noted that any tester that is developed cannot be too expensive as production test will become economically unviable. Adding higher frequencies with the large channel counts could feasibly increase tester cost more than 10x.

Increased parallelism. Increased parallelism also leads to the need to perform multi-site RF test. Currently, mobile RF system-on-chips (SoCs) are tested at x4 multi-site wafer test, with a desire to get to x8 fully parallel test when possible. This still holds true with 5G devices. However, the x8 parallelism pushes the total channel count upward to more than 256 RF channels on many of the devices being developed today. Making a tester that could connect to this number of RF channels is not feasible for two reasons: 1) cost, and 2) space limitations within the test floor. Without increasing the channel count in the tester, there are alternative methods to increase the total channels that can be tested. These methods include using baluns, power combiners, switches, and loopback test, to name a few.

Verifying signal accuracy. The last requirement for 5G parts is accurate RF test in the test cell. This is not only to validate RF performance, but to prevent packaging bad devices. With most of the current advanced packaging technologies that can support RF being quite expensive, yield of more than 95% is required from a wafer where it would be economically viable to not do wafer RF test and only do final package test. However, the high-speed signals in 5G mean that it is more sensitive to process variation, especially in the initial stages of technology development. That means that yield will be well below 95% for a period of time, therefore making it impossible to remove wafer RF test.

In order to support high signal integrity, measurement errors—such as impedance mismatch, cable loss, and RF source variation over time—that might be considered insignificant at lower frequencies can become important at millimeter wavelengths. Using a simultaneous multi-site calibration during multi-DUT testing provides the highest electrical accuracy because all the DUT RF channels are in a known and controlled state.

RF calibration is used to move the measurement reference plane from the tester to the device in order to obtain the best device measurement and to remove the effects of the test fixture. This is done by measuring RF on a calibration substrate (**Figure 2**). In order to accurately measure the effects of the text fixture, the calibration substrate should mirror the multi-site layout of the probe card.

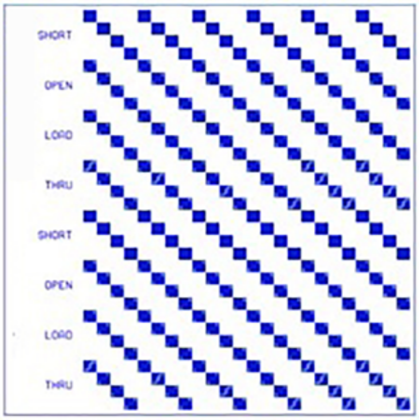

Figure 2: Example of a custom multi-DUT calibration substrate for SOLT.

For RF calibration, there are a number of options that use a combination of RF standards (**Table 1**). For lower frequencies, short-open-load-thru (SOLT) is a standard calibration technique. For highest accuracy, SOLT requires good definitions of all of the standards. It is possible to use short-open-load-reciprocal (SOLR) as an alternative as long as channel-to-channel cross talk is less than 20dB. That is because when the cross talk reaches the level of less than 20dB, the SOLR algorithm gets confused and will not be able to properly calculate the thru length. Therefore, in such cases, it is best to go back to SOLT.

Test Method	Wafer Test Cost	Functional Test Coverage	Full RF Bandwidth Test	Probe Card Complexity
No Wafer Test	🟢 / 🔴	🔴	🔴	🟢
DC Test	🟢	🔴	🔴	🟢
Full Channel	🔴	🟢	🟢	🔴
Loopback	🟢	🔴	🟢	🟡
Baluns, Switches, Combiners	🟡	🟢	🟢	🔴
Antenna Coupling	🟢	🟡	🟢	🟡

Legend: Green = Best, Yellow = Acceptable, and Red = Unacceptable.

Table 1: Comparison of six different wafer test methods.

As the frequencies get higher, an alternative model—multi-line reflect thru (mTRL)—is used. mTRL was developed by NIST, and is considered to be the gold standard in RF calibration. However, it is difficult to use with probe cards because of the fixed distance separation between the probe tips.

Another thing to consider in RF calibration is that when there is a lot of "noise" in the post-calibration verification, it could be from cross talk as well. In order to reduce this, it has proven useful to use sweep averaging and a reduced intermediate frequency (IF) bandwidth (BW) on a vector network analyzer (VNA). This reduces the effects of the cross talk signals, narrowing the acceptable signal being measured. This brings the measurement into a +/-0.1dB of 0 in a post-calibration verification, which is an acceptable level in production test.

Production test methods for 5G

In order to evaluate the deficiencies of current methods for use in a 5G production wafer test floor, we considered four primary metrics of a test method:

1. Wafer test cost;
2. The comprehensiveness of the functional test coverage for the DUT, requiring all channels to get back to the tester;
3. The ability to do full bandwidth test, where the signals do not need to get back to a tester channel; and
4. The complexity of the probe card, which then results in an increased cost for the consumable.

We then looked at six possible test methods (**Table 1**). As can be seen, no single method is fully capable today and requires some amount of compromise. All methods were evaluated using FormFactor's Pyramid Probe probe card technology.

All of the methods listed above are fairly well understood and have been used for years. The only method that is new for wafer test is antenna coupling. Antenna coupling demonstrates the ability to do power combining at a large bandwidth (much larger than most combiners and baluns) and requiring less area and no power, like a switch. This results in a lower cost of test impact when compared to the other methods, which all have at least one red column. The following sections will go over the antenna coupling method, including a technical description on how the method works, as well as test structures being evaluated, and then finally with a test that was done with Intel.

OTA testing using antenna coupling

When using antennas, it should be noted that there are different regions of what the electromagnetic field looks like in relationship to the distance that it is away from the antenna (**Figure 3**). Traditional over-the-air (OTA) testing is done in package test, where test engineers place a horn antenna into the tester at the far-field region of the antenna, and where electromagnetic radiation is the dominant type of energy transfer. However, the use of the antenna coupling method for wafer test places the antenna much closer, in the near-field reactive region at a distance of <100µm, where non-radiative energy transfer is the dominant type of energy transfer.

Although the antenna is not forming a true antenna beam pattern on account of the closeness of the antenna to the DUT, it does provide some benefits. For one thing, most antennas operate in a narrow band, with a fractional bandwidth (BW) of maybe 10-20%. Operating in the reactive region allows for the BW of performance to be much larger, closer to 80%, or more. In addition, the advantage of placing the antenna in the membrane is that it can be used also for antenna power combining with several channels, and arbitrary antenna designs can easily be designed within the membrane.

In order to evaluate the capability of the Pyramid Probe for OTA testing, a test Pyramid Probe was designed with two different types of antennas: a ring antenna and a dipole antenna (**Figure 4**). In addition, there were two different sizes made of each type, to evaluate how the performance changes with size. In order to simulate the DUT, a calibration substrate was

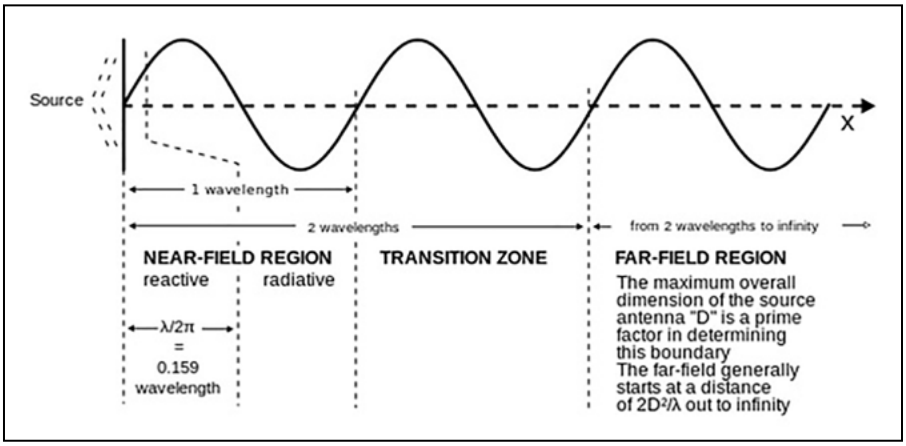

Figure 3: The relationship of distance to the type of field measurements that are done using an antenna. The wafer test antenna coupling method sits in the near-field, reactive region [11].

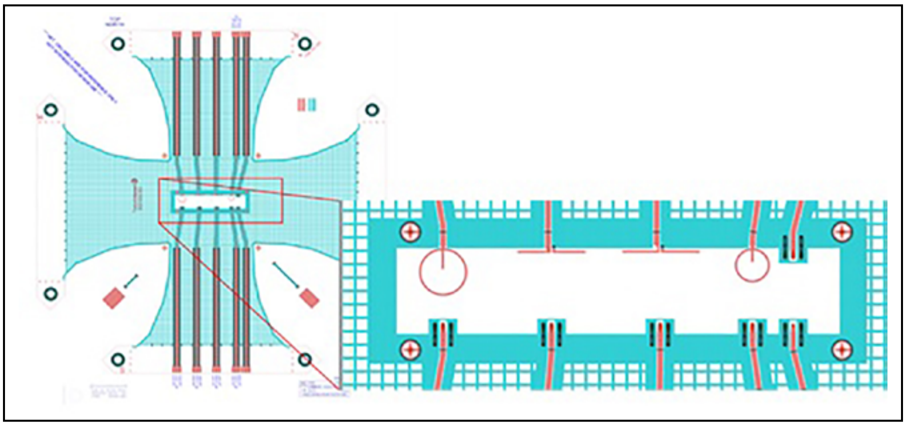

Figure 4: The test membrane designed by FormFactor for evaluation of in-membrane antennas. It included two different types of antennas (dipole and ring), with each having a large size and a small size.

designed with the same antennas on it. The results (**Figure 5** and **6**) demonstrate that the antennas have good signal-to-noise ratio, as well as operating ranges from about 5GHz to 50GHz with flat performance.

Another test was to use the ring antenna in the membrane, but place a pad on the calibration substrate and see if a pad on the DUT could provide a large enough signal to be detected (i.e, does the DUT need to have an antenna on it for measurement, or can a pad be used without any additional structures). The results (**Figure 7**) indicate that it is possible to receive a signal, but the response is not as flat as with the antennas on the calibration substrate.

Probe touchdown	Transmit Power from Same DIE (dBm)		
	1/4λ to Ring Antenna	Ring Antenna only	Fully Conducted Path
1	-63.27	-86.67	-38.593
2	-63.169	-85.95	-38.594
3	-63.8	-86.68	-38.588
4	-63.825	-86.62	-38.589
5	-63.636	-85.63	-38.59
6	-63.687	-85.51	-38.597
7	-63.793	-86.62	-38.602
8	-64.043	-86.23	-38.61
9	-64.728	-85.14	-38.616
10	-64.673	-85.98	-38.615
11	-64.955	-86.69	-38.634
12	-64.866	-85.43	-38.649
13	-65.111	-85.95	-38.648
14	-65.785	-84.65	-38.698
15	-65.826	-84.25	-38.711
16	-65.854	-84.13	-38.757
17	-65.748	-84.32	-38.762
18	-65.831	-84.61	-38.766
19	-65.696	-84.21	-38.753
20	-65.692	-84.74	-38.778

Table 2: OTA testing for 5G devices using three different methods.

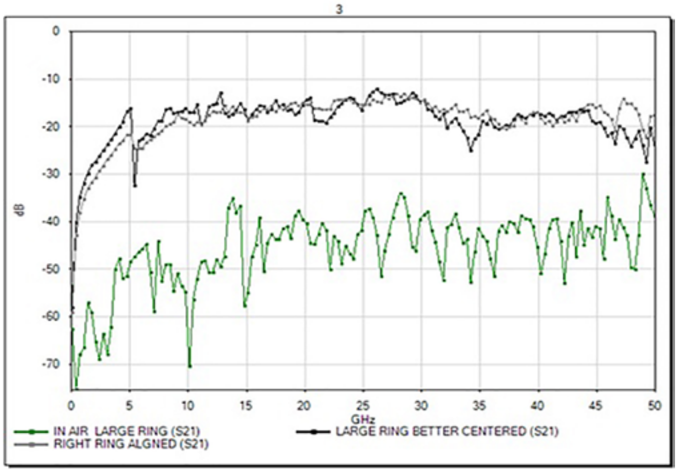

Figure 5: Measurements of the antenna coupling of the ring antenna. The green line denotes the noise floor of the test system.

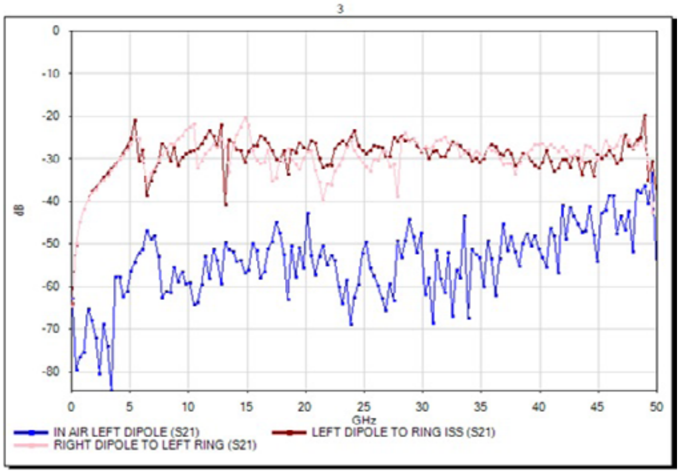

Figure 6: Measurements of the antenna coupling of the dipole antenna. The green line denotes the noise floor of the test system.

OTA testing with 5G devices

In a joint collaboration with Intel to develop a test methodology for their 5G RF-SoC devices, OTA testing was explored because it could provide simpler power combining in the dense RF-SoC device layouts [2]. A test device was selected by Intel, and a membrane was designed for this device that then included three different measurement paths:

1. Standard conduction – direct electrical connection of the DUT to the tester;
2. One-quarter wave antenna in the membrane that connected to the solder ball, which transmitted to a ring antenna in the membrane, which then took the signal to the tester; and
3. Ring antenna directly above a BGA solder ball.

In order to show repeatability, each method was done with 20 touchdowns on the same DUT. Each method showed a high amount of repeatability (**Table 2**). Although the power level is lower in the quarter wave to ring antenna method (as well as solder ball method to ring antenna) when compared to the traditional conduction method, the ability to power combine multiple channels provides some benefits in reducing the complexity of the active circuitry and cost of the test setup while still providing a good signal for known good die testing. In addition, a measurement of a single tone going through the system showed both a clean, unmodified signal, as well as high repeatability from the first to last touchdown (**Figure 8**).

The final test evaluation of the antenna test method was then to look at the signal during a 12-tone, linearity measurement to check for harmonic distortions. The signal was shown to have no cross-mode effect, with a highly linear response across the full band, as well as a signal that was exceptionally clean (**Figure 9**).

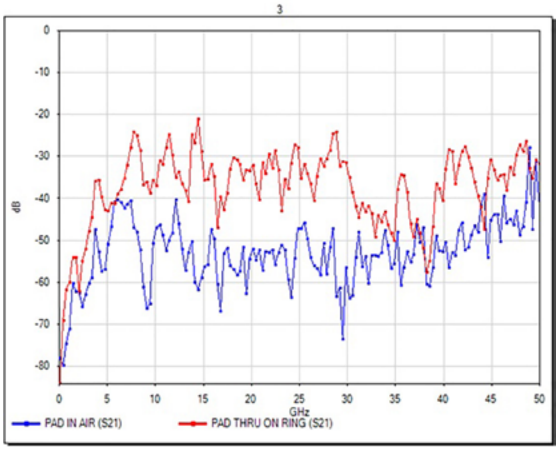

Figure 7: Measurement using a ring antenna to receive a signal from a pad on the ceramic substrate. The blue line is the noise floor.

Summary

The development of 5G is requiring a paradigm shift in test methodology with a tight relationship between the device manufacturers and test hardware manufacturers. Millimeter wave testing and OTA have increased challenges and a corresponding impact on yield. An exponential increase in RF channels is driving the collaborative development of production test methodologies in order to establish 5G deployment standards. Our collaboration with Intel to develop a new test methodology for 5G exemplifies a path that semiconductor manufacturers could pursue in full production.

References

1. https://www.pcmag.com/article/345387/what-is-5g
2. D. Bock, M. Biship, J Damm, B Singh, M. Engelhardt, M. Hemena, et al., “5G: The next disruptive technology for production test,” Semiconductor Wafer Test Workshop, June 3-6, 2018; http://www.swtest.org/swtw_library/2018proc/PDF/S01_03_Bock_SWTW2018.pdf
3. https://www.edn.com/electronics-blogs/5g-waves/4458756/The-5-best-5G-use-cases
4. https://www.ctia.org/news/global-race-to-5g-spectrum-and-infrastructure-plans-and-priorities
5. P. Schulz, et al., “Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture,” in IEEE Communications, vol. 55, no. 2, pp. 70- 78, Feb. 2017.
6. https://www.lifewire.com/5g-wireless-4155905
7. G. Fodor, N. Rajatheva, W. Zirwas, L. Thiele, M. Kurras, K. Guo, et al., “An overview of massive MIMO technology components in METIS,” IEEE Comm.; vol. 55, Issue: 6, pp. 155-161, 2017.
8. https://kb.accelerated.com/m/66701/l/708211-a-brief-history-of-mobile-networks
9. https://www.zdnet.com/article/what-is-5g-everything-you-need-to-know/
10. https://semiengineering.com/5g-lessons-learned-from-automotive-radar-test/
11. https://www.osha.gov/SLTC/radiofrequencyradiation/electromagnetic_fieldmemo/electromagnetic.html

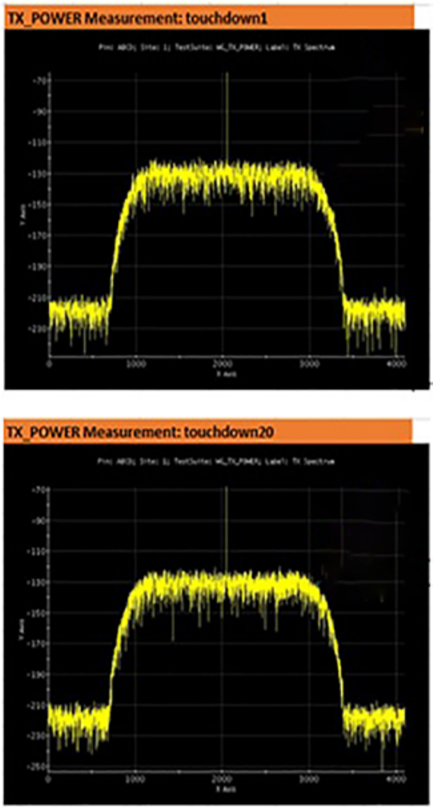

Figure 8: Single-tone measurement of the antenna method showing repeatability. The measurement on the top was on the first touchdown, while the one on the bottom is on the 20th touchdown, on the same device.

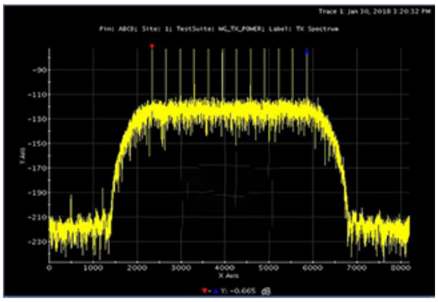

Figure 9: Evaluation of a 12-tone linearity measurement.

Biographies

Daniel Bock is an RF Applications Specialist at FormFactor, Inc. He received his PhD in Physics at Carnegie Mellon U. in 2006. He went to work in 2007 with Physical Optics Corp. in Torrance, CA, where he was awarded more than $4M in SBIR research grants from the Department of Defense, developing innovative high-power tunable filters. Since he joined FormFactor (then Cascade Microtech) in June of 2012, he has been focusing on mmW RF test solutions including RF Calibration in a HVM environment, RF design of structures, and the requirements for automotive radar and 5G; email Daniel.bock@formfactor.com

Jeff Damm is a Senior Product Applications Engineer at FormFactor working with mmWave probe technology. He began his career in research and development at Dexcel, Inc., working with GaAsFET waveguide low-noise amplifiers. He next joined Tektronix GaAs research group, and then co-founded Triquint Semiconductor focusing on GaAs/pHEMT/HBT RFIC product design and development group. Prior to joining FormFactor, Jeff worked in the development of mesh-network wireless instrumentation at Lizard Monitoring, a developer of temperature sensing network systems for retail food chains. He graduated from Oregon State U. with a BS in Electrical & Computer Engineering.