Optimizing the Whole Test System to Achieve Optimal Yields and Lowest Test Costs

Jeff Sherry and Dave Haupt
Johnstech International Corporation and Agilent Technologies
Minneapolis, MN and Santa Rosa , CA
jcsherry@johnstech.com and dave_haupt@agilent.com

Abstract

As semiconductor device bandwidth, speed, functionality and resolution continue to climb, it becomes more obvious that the whole test system must be optimized to achieve the highest yields at the lowest cost per device and insertion.

The overall system performance depends not only on the automated test equipment (ATE) but equally on the interfacing between the ATE, the device under test (DUT) and the IC handling equipment. Everything must interface properly and predictably, both electrically and mechanically. If any single part of the system produces non-repeatable results, then the measurements taken on the device will not be repeatable. Because of the high visibility of capital, many organizations and their test engineers underestimate the contributions from the relatively inexpensive pieces to the test puzzle, specifically the contactor, the load board, and the mechanical interfacing between the ATE and handler (docking). This paper will provide a framework for understanding the relationships between the elements, demonstrating the importance of the DUT interface, and will present an engineering approach to maximizing yield through optimizing the whole system. Detailed examples will follow, showing how this approach is applied to the critical interface parts of the system.

The significant benefits to be gained by using modeling and simulation will be discussed. Numerous examples of the successful correlation of modeled and measured data will be presented. The combined efforts of modeling and measurement will show how details such as device package tolerances, lead or pad plating, or how long a device sits in production before testing can impact test yields and costs. Modeling will also be shown to help reduce costly board redesigns, obtain true device performance, calibrate the system, and ensure repeatability of measurements taken in production.

This paper will conclude by showing how RF bandwidth, signal integrity, and grounding can be improved by choosing the correct materials for items such as contactors and load boards. In addition, data will be shown that optimizing the load board and contactor can allow devices to be retested more quickly and more accurately, thus improving first pass yields and reducing the need for retesting failed devices in production.

Test System Setup and Challenges

Performance gains sometimes measured as overall equipment efficiencies (OEE) begin with the achievement of true device performance measurements with a reduced need for guard banding and other software patches applied to compensate for a poor interface to the DUT. Guard banding and other test software patches are additional overhead that takes time to execute for each device tested. Other bottom line performance gains that contribute to a lower cost of test include ensured repeatability of measurements for an increase in test-system throughput as measured in first pass yield. Another measure is a reduction or elimination of costly load board redesigns.

As can be seen in Figure 1, the most visible parts of a whole test system are the ATE (including its test head) and the IC handler. When the ATE is docked with the handler, the critical DUT-to-ATE interface cannot be seen, but can have a dramatic impact on yields. Its lack of visibility often causes people to underestimate its true importance in electrical testing.

Figure 1: Whole Test System - Tester and Handler Docked Together

Contributors to Yield

The test engineering department's job doesn't begin with test code or hardware, but in fact, system engineering. The relationships between ATE, DUT, load board, and contactor must be investigated carefully. From this investigation, test engineers can determine which elements need to be highly optimized and which elements can be safely neglected. Expending too much effort in areas that won't benefit yield, results in a delay in time to market (TTM) and a waste of engineering resources. On the other hand, failure to perform adequate engineering in a critical area often causes a complete redesign of hardware and software. The most significant knowledge that the test engineer brings to the engineering team is an awareness of which device parameters are critical and are likely to be an issue during testing.

Figure 2 illustrates this point. A careful evaluation of all of the elements of the test system will produce numerical values for the fixed error sources, measurement uncertainty and non-repeatability. The difference between these three

items is crucial for obtaining the desired OEE levels. The most significant contributors to guard bands are in the DUT-to-ATE interface.

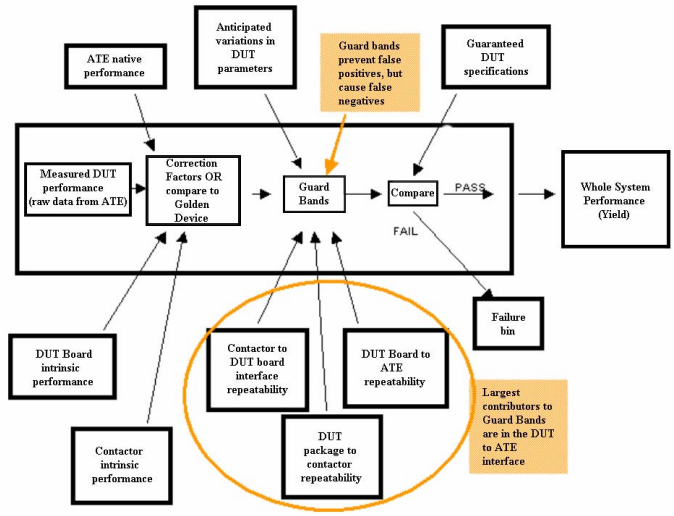


Figure 2: Contributors that Affect Guard Bands and Device Yields

Error Sources and Their Compensation

Fixed error sources is the collection of fixed losses, reflections, and other tester and fixturing error sources that are consistent from one test to another. Proper test system calibration can completely remove these error sources from the measurements. The results of the calibration become a part of the correction factors shown in Figure 2. If there were no uncertainty or non-repeatability issues in the system, the exact device performance could be derived from the corrected data.

Measurement uncertainty arises primarily from electrical noise. However, it can also arise from the normal variations from one DUT to another that cause slight changes in the ATE performance. Most commonly, a DUT will have input and output impedances specified not as absolutes, but in ranges. When two DUTs differ in their input impedances, there is a resultant variation in the apparent gain of the DUT, even if the two DUTs exhibit the same gain. This uncertainty can be eliminated by using 12-term error correction in a RF measurement system. Such error correction is not available in digital systems, and often requires substantial processing time in a RF system. If error correction is used, then these factors are no longer an uncertainty and can be completely compensated for. The remaining uncertainty becomes part of the guard bands.

Non-repeatability cannot be compensated for with any calibration or error correction technique. It resembles noise-induced uncertainty only in that it cannot be compensated. However, noise-induced uncertainty tends to produce a gaussian distribution to measured results, whereas non-repeatability often produces "rail-to-rail" measurement errors. To minimize the required guard bands, non-repeatability must be reduced as much as possible, at least below the RMS level of any noise-induced uncertainty in the system. As a general rule, the noise levels in modern ATE are low compared to the DUT-to-ATE interface. As a result, the guard bands grow larger, and a significant number of good DUTs are labeled as failed, reducing yield.

Yields

The first priority of the guard bands is to guarantee that no defective devices will be falsely labeled as passing. The second priority is to minimize the number of false failures. The optimum width of the guard bands is directly dependent on how much the non-repeatable test elements can be reduced and tolerated. Therefore, the design priorities of the test system are to first optimize repeatability and then reduce other error terms.

Improvements in yield will result from the achievement of true device performance measurements, which result in reduced guard banding. As the discussion above indicates, guard banding is usually applied to compensate for a poor interface to the DUT. Guard banding and other test software patches also add additional overhead that consumes execution time for every device tested.

Importance of Repeatability

As shown in Figure 3, many factors affect the final distribution of the devices being tested. The key is to optimize or reduce each factors' repeatability so the final distribution closely resembles the devices being tested. For each part of the test system, there are key contributors that affect the overall standard deviation.

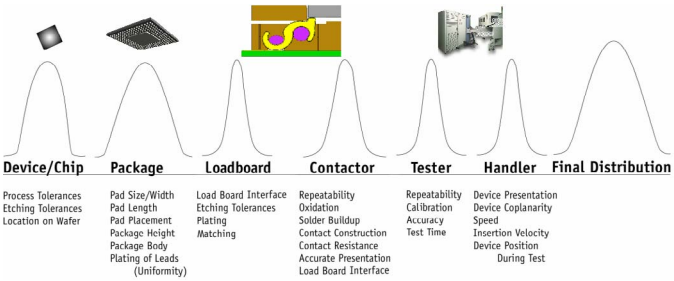


Figure 3: Factors Affecting Final Distribution

As discussed earlier, inadequate repeatability forces the widening of guard bands in the test software. Figure 4 provides one method of graphically displaying the actual distribution of performance contributed by each element within the test system. Clearly, the widest distribution should come directly from the DUT, which is composed of the "Device/Chip" and "Package" elements in Figure 3.

To drive home the importance of repeatability, Figure 4 compares a repeated simple output power measurement from an amplifier, tested hundreds of times by inserting it into a test contactor and making a measurement of output power, against a highly repeatable insertion loss measurement. The insertion loss measurement was a simple cable assembly connected to an Agilent 8722ET vector network analyzer, with the measurements taken 150 times without modifying the system. The result was scaled to make it comparable to the IC power amplifier measurement. The IC system setup included a tester, handler, 0.5mm height rigid contactor technology,

and load board and cables connected to the tester. Both sets of plotted results represent non-repeatability that cannot be compensated for during testing.

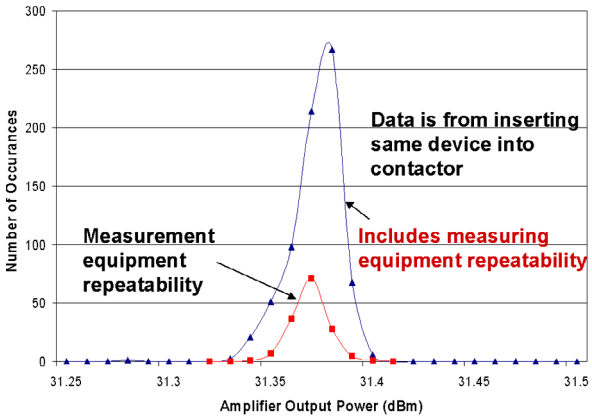


Figure 4: Amplifier Output Power Distribution and Repeatability

Modeling and Simulation

The contributors to the measured results of the device being tested in production are shown in Figure 5. The device is usually plunged into the test contactor and held in place during testing by a leadbacker. Only the top layer of the load board is shown for simplicity. For signal integrity reasons, a device with a small number of high speed or high frequency signals will place those signals on the perimeter of the package, permitting those signals to be routed exclusively on the topmost layer of the DUT board for the best results. The ground plane in such a structure will be the second layer.

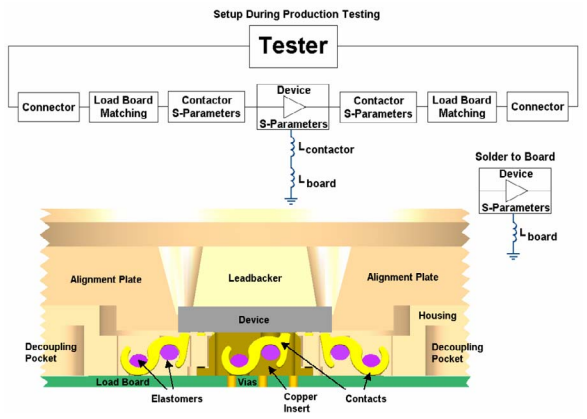


Figure 5: Modeling Block Diagram and Device in Test Contactor

Each part of the system has a set of data that can be provided to software such as ADS or SPICE to help decipher the expected system performance, or to calibrate out repeatable non-device parts of the system. It is customary for RF elements to be described in S-parameters, and digital items to be described in terms of a SPICE model file. It is becoming more common to describe all passive elements using RF-style S-parameters, whether the intended usage is RF, analog, or digital. The reason for the change is that two-port S-parameters are test-equipment independent, while the usefulness and accuracy of SPICE and IBIS models is still driven heavily by the measurement technique used to derive them.

Because the load board and contactor interact, the load board matching or pad and contactor can be modeled or measured together to improve accuracy and show the effects of the interactions. The key is to measure in production the same results that would be attained if the device were soldered to a customer circuit board.

In the test scheme, the device S-parameters are a result of testing or modeling a design or device. The contactor S-parameters are developed based on the type of technology used. The S-parameters can be modified slightly by choosing different housing materials, adding cutouts above traces or components, changing the contact thickness, modifying the proximity to ground of the contacts, or by adding an insert to improve grounding, thermal dissipation or flexibility in design. The contactor is treated as an unalterable part of the system because it is purchased from a third party. The device model is treated as unalterable and usually comes from the design team. In production, the test engineer strives to measure accurately the device parameters, which should conform closely to the design model.

The load board becomes our primary opportunity to match the rest of the system to optimize the performance at one frequency or over the device's bandwidth. The goal in matching the load board is to design it so that the non-device parts (load board and contactor) have minimal and repeatable effects. Impedance match and low-loss trace structures are the usual parameters to optimize. Another key is to derive the load board parameters through modeling or calibration so the test plan can incorporate them as correction factors, yielding equivalent at-the-device measurements. The better the impedance match (or equivalently, return loss), the lower the insertion loss, and the better the repeatability of the contactor and load board.

Modeling is used to predict results, but is only as accurate as the data used in the model. If the data used is exactly the model of what the contactor looks like after assembly, or how the load board is built, the results should correlate well with measured results. To ensure models are as accurate as possible it is important to import the mechanical 3D drawings into Ansoft HFSS in the exact configuration of the contactor during testing. Then the expected device pad and load board layout is added to determine interface effects between the contactor, the device, and the load board. In some instances, measured data for the contactors are obtained using a probing technique at the device and load board interface to get data on only the contact. This data can be combined in a software package, such as Agilent ADS, with device S-parameters and load board response to get a fairly accurate model that represents real life. This model does not take into account signal loss at the contactor interface points. If a portion of the load board trace is extended beyond the load board contactor interface, some errors will occur, even if you model the extended trace as a stub. This can be seen in the E-fields around the contact interface to the device and the load board. The brighter colored E-fields represent radiating energy and is

most severe when the contact points between the load board and contactor, or contactor and device, are at right angles, as electrically energy doesn't like to traverse corners and right angles, especially at higher frequencies.

This approach of modeling the whole load board to device interface works when the configuration of the S-parameter data on the contact matches the load board and device configurations. S-parameter data taken in a Ground-Signal-Ground configuration for the contactor will yield inaccuracies if the device pinout is Ground-Signal-Signal-Ground or is different than the model.

Predicting system performance is often hard to do because pieces of the system might be missing. Modeling the missing pieces helps provide the missing data and investigate trends to determine problem causes and solutions. Modeling is always an inexpensive alternative to building hardware because you can determine: potential problems before building hardware, expected performance, trends, effects of tolerances, and interaction between components in the system (device, contactor, handler, etc.). For instance, modeling solder buildup on contacts will show higher contact resistances and degraded RF performance. Knowing the contact resistance variations between maintenance cycles can actually improve yields if the tester is allowed to make adjustment for these differences.

Calibration

There are many aspects to calibrating a testing system. In a digital system, the ATE should be used to measure the length of the load board traces, so that edge timing can be aligned. For analog and RF testing, one of the simplest calibration methods is to use a golden device. A golden device is a device whose characteristics have been precisely measured using lab equipment. Calibration factors for the test program are simply the difference between the device's actual performance and how it measures on the test system.

Frequently, the calibration data and specification limits will be merged into a single value in the test program. For instance, if the device has an insertion loss of 1.5 dB and the specification is less than 2.0 dB, then all devices tested should have less than 0.5 dB insertion loss than the golden device had in the test setup. This method is fast and very reliable if the test system is repeatable and consistent over time, and if the parameter under study is known to be independent of other parameters which vary. To ensure that the test system continues to hold its calibration, the golden device must be periodically re-measured, and the results compared with the first time it was measured. A deviation between the original measurement and the new measurement may indicate a need for cleaning or maintenance. It is advisable to cycle in new golden devices periodically as golden devices tend to wear out due to use.

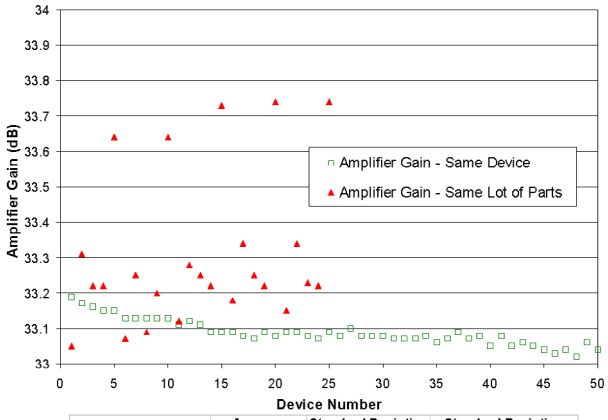
Another form of RF calibration uses a set of devices that have been designed to provide a through, open, short, and perfect 50 ohm load. This is the Short-Open-Load-Through (S-O-L-T) method. These calibrated devices are inserted into the contactor in sequence. An ATE system that provides for 12-term error correction can then calibrate the entire system from the ATE to the device being tested. When implemented properly, the tester will provide measurement data for the DUT with all effects from the contactor, load board and ATE removed.

A simple measurement of losses in a similar system is sometimes used. This calibration, known as "normalization" can be adequate if the devices have known and consistent input and output impedances, and if you are not attempting to measure very low throughput gain or loss with high accuracy.

Testing For Repeatability

Since repeatability is such a key item, it is crucial to test for it. If the test system has been developed correctly, the variation in system performance is substantially smaller than the variation from device to device. This can be easily verified by conducting two test runs. One test run repeatedly plunges the same device into the tester, measuring it over and over. Any variations from test to test represent the repeatability range of the system. The second test run sequences production devices through the system. The variations from test to test in those results are a combination of the tester's repeatability range and device production variations.

Figure 6 shows the gain test data for an amplifier. The first data set is from 25 different units in the same lot. The second data set is from testing data testing the same device 50 times to determine the variability of everything except the device. Based on the data, the device has five times as much variability, or a higher standard deviation, than that of the system. These tests were done after the system had already tested 5000 parts. This simple test determines the whole system performance and shows if a part of the test system needs to be improved to increase OEE. The system was a production test system, complete with tester, handler, cables, load board, contactor and devices. The goal is to have the smallest repeatability range to achieve the highest confidence level in the test results. Since the handler and contactor interface are the same for both tests, it does not determine the affects of the presentation or alignment of the devices into the test contactor. Improving this interface could reduce the standard deviation of the whole system.



Parameter	Average Device	Standard Deviation Device	Standard Deviation Contactor/Equipment
Amplifier Gain	33.31 dB	0.213 dB	0.037 dB

Figure 6: Gain Variation Test Data from a High Gain Amplifier

When testing the gain of the same part, the gain drops over the number of tests because the contact wipes the device in the same spot. After 10 insertions the wipe function tends to push the solder plating, creating a different pad geometry. In production, the device is normally only placed in the test contactor three times if testing is done at cold, ambient, and hot. The more times a device is tested, the more the test costs per chip rise. If the contactor is repeatable, there usually is not a need to retest devices since false failure rates are usually very low. One more important conclusion can be drawn from the graph in Figure 6. The test-to-test uncertainty of the test system is dominated by the steady removal of oxides from the contactor. Therefore, we can conclude that all other optimization of the test system has reached a level where no useful benefits would be gained by further enhancements.

In addition to demonstrating that the system's repeatability has been optimized sufficiently for the devices being tested, the results of this test help determine test limits within the range of repeatable performance and also identify what areas of the system need to be fixed or modified to improve performance. If the standard deviations indicating loss of repeatable control, within a design range, are large for both tests it might be beneficial to plunge or solder some devices to the load board to determine if the contactor is the cause of the variation. By switching equipment in the tester, it is possible to determine if the test equipment is causing the variability.

Identifying Contributors to Performance

After designing and building a load board, a time-domain reflectometry (TDR) measurement could be taken as shown in Figure 7. The benefit of the TDR measurement is that it shows the individual contributions from each feature on the load board. If the overall performance was inadequate, this measurement would identify which areas of the load board needed additional engineering. (See Figure 3 for possible causes of layout problems.) The plot shown in Figure 7 is of two small load boards, interconnected with a one-piece rigid 1mm production contactor. Pictures of the test setup are shown in the bottom right corner of the figure. The center-line is 50 ohms. The TDR instrument used exhibits a 20GHz bandwidth.

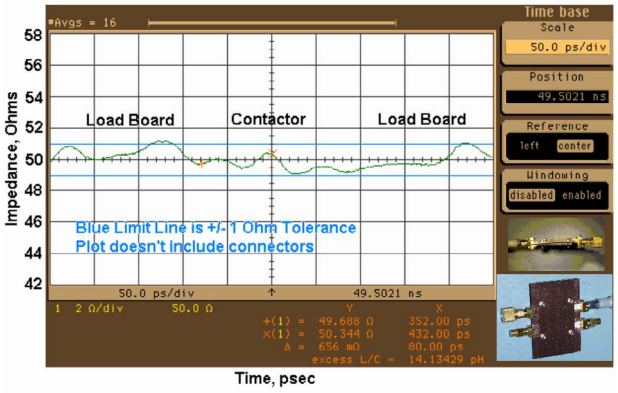


Figure 7: TDR Measurement of 1mm Contactor System

The TDR measurement demonstrates a system that is well matched to 50 ohms. This whole system was modeled in HFSS prior to construction to verify that the load board was being built to optimize the electrical performance. The display parameters were adjusted to only show the load board and contactor. The small discontinuities contributed by the end-launch SMA connectors are not shown in the photo. End-launch connectors usually provide better performance because the travelling electromagnetic waves undergo a smaller geometry transition as they exit the coaxial test cable and enter the planar structure of the PC board. The end-launch connectors used in this measurement exhibited impedances of 52 +/- 2 ohms.

Device Packaging Effects on System Performance

Testing a device in a contactor will typically leave a small witness mark where contact is made to the device. This mark is necessary to break through any potential oxides that may have formed on the pads, leads, or balls of the device. There are many different types of plating, each one providing a unique set of advantages and disadvantages. The industry is being driven toward lead-free solders, which are generally much harder substances than lead based solders, and will require greater contactor forces to break through any oxide layers formed on the devices. In these instances, contacts need to be redesigned or elastomers need to be changed. This is also sometimes true for temperature testing, as cold or hot testing might require an adjustment in forces applied to the devices inserted into the contactors. Some package leads are gold plated, which is an excellent choice for testing because gold is resistant to oxides, but its cost can rarely be justified.

Solder develops oxides over time which could have a big impact on the cost of testing devices. Oxides form in a thin film, which varies from being a poor conductor to being non-conductive. The longer the part sits in an uncontrolled environment, the more oxides tend to form. The softer the solder or higher the lead content, the more contactors will tend to stick to the solder or the faster solder debris will build up on the contactor. This results in more frequent contactor pin cleaning cycles or reduced yields.

Figure 8 demonstrates the added contact resistance caused by oxides. Each pair of vertical bars represents a specific pin on the device. The longer bar is the contact resistance upon first insertion of devices that were allowed to build up oxides. The shorter bar represents the contact resistance of devices with oxides that were plunged enough times to clean the oxides, or were from devices which had just come directly from the plating process. The plating was 85/15 tin-lead and the contactor was a 2mm rigid one-piece contactor pin. The numbers are an average from 150,000 insertion on a handler. As can be seen from the graph, the average contact resistance for all contacts was 0.010 Ohms with oxide film, and 0.0079 ohms with little or no oxide film. The variation from minimum to maximum contactor resistances for each scenario is very similar. In the case of testing done with oxide film the deviation from minimum to maximum was 0.00135 ohms. In the case of testing done with little or no solder oxide film the deviation was 0.00131 ohms.

If parts are very sensitive to contact resistance or fluctuations, such as 12 to 16 bit DACs, it will be important to test devices with no oxides present. Otherwise, false failures

might occur or inaccurate readings might be taken. One way to verify if contact resistance is a problem is to insert the same device in a wiping contactor multiple times. If the data gets better or is more repeatable after each insertion, oxides are changing the measured device performance.

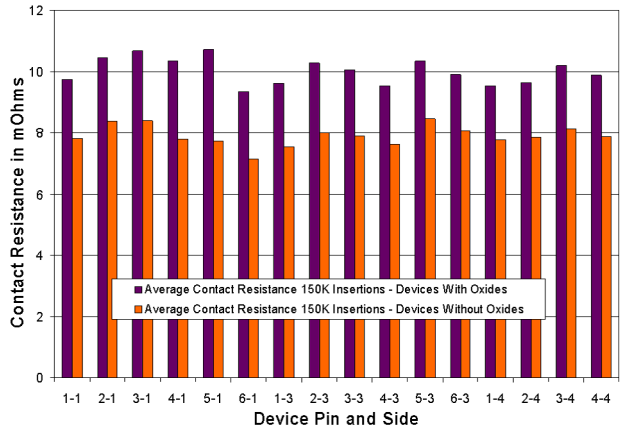


Figure 8: Differences in Device Contact Resistances with Changes in Oxide Levels

Figure 9 shows the differences between inserting a device into a contactor in a handler environment that has no or very little oxide layer versus inserting the device with oxides formed on the device pads. An increase in contact resistance and an increase in standard deviation indicate a need for cleaning. However, the data is still consistent, so for this particular contactor in this handler environment the cleaning interval goes beyond 150,000 insertions. After a period of non-use, such as a weekend, oxides will form on the solder residue left on the contact pins from previous testing. Contactors should be cleaned or actuated after a period of non-use to remove solder oxide from the contact pins. All contact pins, whether they have a wiping action or not, will tend to exhibit a higher contact resistance after a period of non-use. A contact with a wiping action, if not cleaned, will result in a period of higher contact resistance until enough insertions have been performed to clean the oxides off.

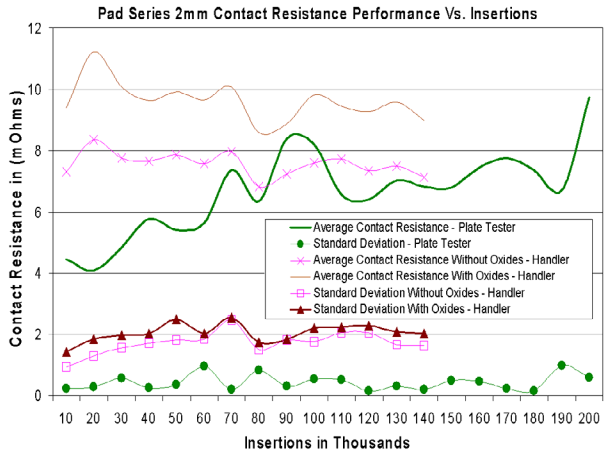


Figure 9: Contact Resistance With and Without Oxides and Effects of Device Presentation on Repeatability

Figure 9 also shows the results of just the contactor with no interface or presentation effects. This was done by testing the contact resistance using a solder coated plate inserted into the contactor with no alignment features. The plate eliminated all changes in contact resistance that might have been caused by handler interfaces with the test contactor or presentation problems.

Both sets of data used the same plating requirements for the devices being tested and neither system was cleaning during the testing. In the plate test case, there are no presentation or interface issues, and the contactor can run continuously for 200,000 insertions before any maintenance is required. Because the plate tester used just a flat copper plate with solder plating and allowed each insertion to be made on a fresh part of the plate a good and relatively inexpensive measure of contactor Mean Time Between Assists (MTBA) is determined. In addition, the differences in standard deviations between the devices tested on the handler and test results from the solder coated plate tester are a result of the handler - contactor interface. For the plate tester, the average contact resistance slowly increases up to 0.010 ohms at 200,000 insertions and all data points were below the 20 mohms contact resistance limit. Because the device was a plated plate, no package debris was removed so opens due to debris from the package were not seen during the plate testing. On the plate tester, the mounting screws were recessed so as to to not affect the solder plated tape that was used. In both tests, the same load board design and test equipment were used to eliminate any unforeseen effects. When determining the effects of the test system, it is important to vary only one parameter at a time.

In Figure 9, the data is summarized each 10,000 insertions. The contactor was run continuously for 150,000 insertions without cleaning or maintenance with the contactor on a Delta Flex Handler. From the chart we see there is a 0.0013 ohm difference between devices with and without oxides. Because the contactor has a self-cleaning or wiping action on devices being tested, the standard deviation for both tests is around 0.002 ohms and very stable. On the plate test the standard deviation averages 0.0005 ohms. The difference in standard deviations is the effect that device presentation has on the repeatability of the system.

Device package tolerances also have an impact on repeatability. Wide package tolerances could cause more open contacts on the corner pads of devices due to increased rotational errors during handler insertion. The probability that this could happen depends on device package tolerances, device pad size, and alignment of the device and contactor interface.

Figure 10 shows the slight difference in contact resistance behavior between a corner contact pin and a middle contact pin. The chart shows how the Pad Series 1mm Production Contactor behaves over 120,000 insertions with no cleaning or maintenance. After a long period of inactivity, 4 days in this case, both contact resistance averages and standard deviations increased slightly. This is because oxides tend to form on the solder of contact pins and devices, which takes a period of time to wear off. Also, around 100,000 insertions both the average contact resistance and standard deviation

start to trend upward. This is a sign that the contactor might require cleaning. The cleaning interval, in most cases, will depend on the device's sensitivity to contact resistance variation. Increasing contact pin resistance is normally the cause of a reduction in device yield.

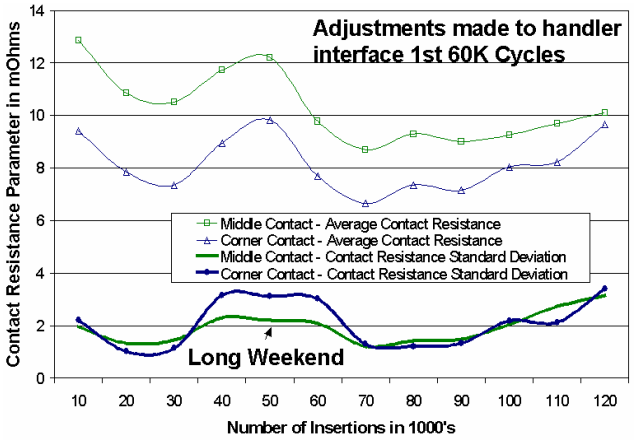


Figure 10: Contact Resistance vs. Number of Insertions for Corner and Middle Contacts

Figure 11 shows a cut away of a complete HFSS model, including the load board. The contactor does not show the elastomer and housing in order to better illustrate the signal path and packaged device characteristics, including wire bonds. Such models can include elements of package tolerancing, so that the impact of tolerancing can be predicted, and factored into repeatability calculations.

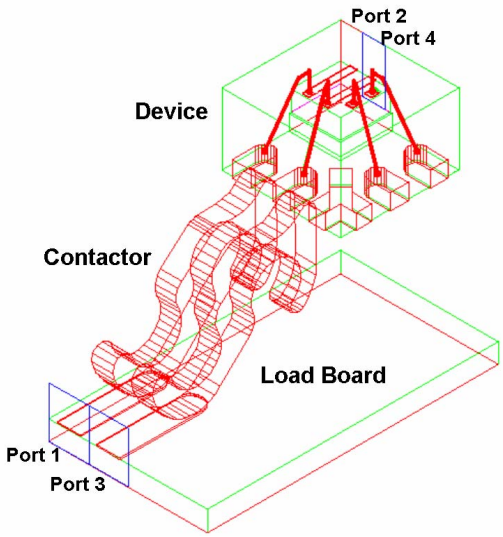


Figure 11: 3D Electromagnetic Model Used to Predict the Effects of Package Tolerances

Debris is an important factor in determining when contactors need to be cleaned. In many cases, the amount of debris will depend on the package of the device or the handler presentation. Usually, sawed packages create more debris than molded packages. If the alignment between the handler and contactor is not optimized, packages that are at the higher end of the tolerance range may either jam or rub off debris. Another question that affects the time between cleaning is how the contactor handles debris. If the contacts tend to push debris or oxides away from the contact area, due to wiping action, or the housing has holes or slots that allow debris to fall to the load board and out of the way, it will result in much longer MTBA. However, when cleaning the contactor, after removing the contactor from the load board, the load board should be cleaned of excess debris to prevent debris from getting under the contacts and creating poor connections and/or premature wear of the pad on the load board.

Debris most commonly causes contact pins to effect the measurement resulting in false failures. Contacts that provide a wiping action push the debris aside after a number of cycles. If the contact has no way to remove debris, testing will have to stop for cleaning, resulting in costly down time. Also, as the number of insertions increases, more debris from the device packaging collects around the contactor. This is especially true for sawed devices. If the debris has no place to go or can't fall through contact slots, the debris could affect contact performance by creating stuck pins or high contact pin resistance.

Load Board Effects on System Performance

In order to optimize system performance, it is important to model the contactor to load board interface together because the interface is a transition from microstrip or coplanar waveguide to contact pin. In a perfect system, the device, contactor, and load board paths would all exhibit the same characteristic impedance - typically 50 ohms for RF or high speed digital signals. In practice, a controlled impedance is never maintained at all frequencies. In most instances the load board must be matched to the contactor and the device to ensure measurements represent true device performance. Usually, the load board pad width is larger than the contact pin and usually the pad is a stub, which will radiate electrical energy. Figure 12 shows the E-field on the load board from an optimized pad layout in comparison to a layout that is not optimized. A non-optimized layout might be used to accommodate two different contacting technologies. The E-field plots are looking down on the contactor pin through the contactor housing. Notice that in the non-optimized load board footprint, the traces are wider to account for larger tolerances, possibly due to alignment and tooling hole tolerances on the load board. Also notice that the E-fields are much more intense on the non-optimized load board pad, as indicated by the orange and red areas on the plots. Non-optimized load board landing pads result in stronger fringing fields, which could result in impedance mismatch, loss and crosstalk.

The isometric view in Figure 13 shows the E-fields in the device on the left side, E-fields radiating from the contactor contact in the center, and E-fields from an optimized load board pad on the left side, which has a trace width just slightly larger than the contact width. To increase the bandwidth, the length of pad extending beyond the contactor pin to load board interface should be minimized and radiused to eliminate stub effects and sharp corners. Just like right angle connections, stubs are bad for RF signal propagation. If

the pad is extended beyond the contact pin interface point, it can be modeled as a stub, and depending on length and device frequency, could severely affect electrical performance.

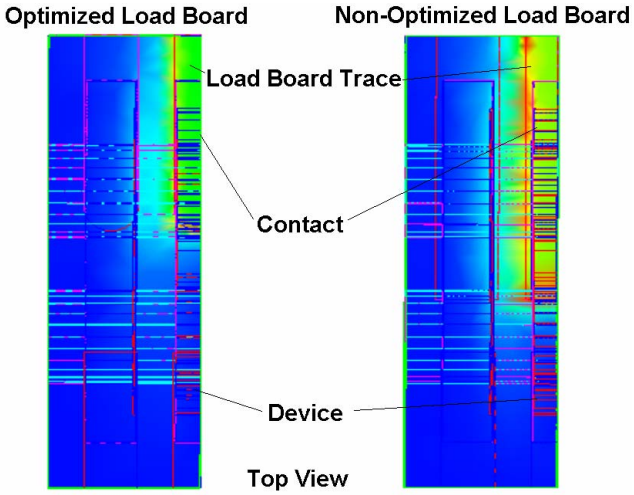


Figure 12: Top View of E-Fields Radiating from Optimized vs. Non-optimized Load Board

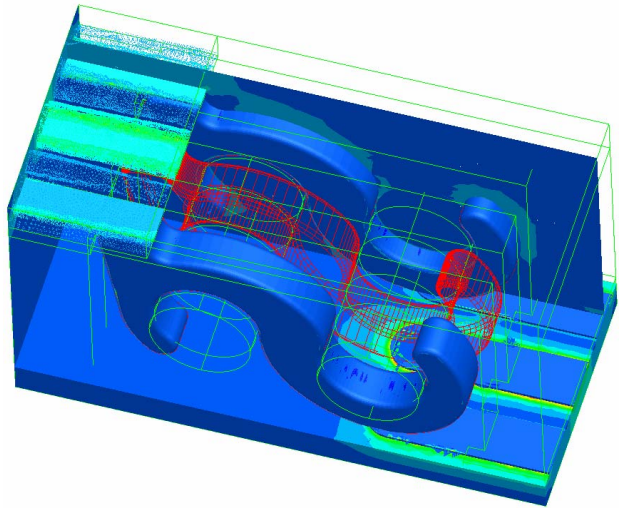


Figure 13: Isometric View of Load Board and Contactor E-Field Plots

Load Board Tolerances

If the load board has a large tolerance for the contactor mounting or alignment holes, the contactor could sit in a different position when remounted after removal for cleaning or other maintenance. This could result in contacts sitting on load board traces differently, which could affect electrical results. To accommodate loose tolerances, load board contact pads are often made larger to ensure a solid contact. However, this is beneficial only at DC. The larger pads result in undesirable stub effects. A better design would employ tight tolerancing on the contactor mounting and alignment holes.

Load Board Materials

When selecting a load board dielectric material, the dielectric constant (ε_r) and loss tangents are the two most important properties for predicting signal integrity. For the same dielectric thickness, a lower dielectric constant requires a wider trace to achieve a given characteristic impedance. This has the benefit of reducing skin effect losses, which become important with long traces. However, if the trace-to-trace spacing remains constant, a lower dielectric constant will also result in more crosstalk between adjacent traces. A highly effective technique for reducing crosstalk is to separate signal traces by ground. The ground trace must be "pinned" to a ground reference at least every 1/8 wavelength at the highest frequency expected, otherwise what was supposed to be a shield can become a resonant structure that actually increases coupling between traces. Low-density devices that primarily employ single-ended signals usually have high frequency signals separated by ground pads. Very often, clock signals are routed as differential pairs with ground pads separating the differential pair from other sensitive signals.

The substrate thickness, trace width, and substrate material are the three most significant controllable factors in determining the trace impedance. Modifying trace thickness and adding cutouts in the contactor above traces are other ways to slightly modify the trace impedance. There are many ways to construct a load board to get a 50 ohm trace. Listed below are some ways to change the trace impedance.

Microstrip and Coplanar Effects

- Substrate thickness ↓ — Impedance ↓
- Trace width ↓ — Impedance ↑
- Permittivity (ε_r) ↓ — Impedance ↑
- Trace thickness ↓ — Impedance ↑
- Adding air gap above trace (cutout in housing) — Impedance ↑

Coplanar Waveguide Effects

- Spacing (pitch) ↓ — Impedance ↓
- Adding ground plane — Impedance ↓

This list acts as a good reference when results of modeling shows the impedance is not 50 ohms. In any system, the closer the ground is to the signal trace the lower the characteristic impedance. Placing a conformal coat or a solder mask on top of traces will slightly lower trace impedance. The amount depends on the thickness and the dielectric constant of the coating.

In a production environment, the load board will need to be stiff to handle the constant pounding from the handler inserting parts into the test contactor. For low pin-count devices, very few layers will be required on the load board for trace routing and to achieve optimum electrical performance. To achieve the stiffness requirement, additional filler layers are usually added. Because there are no electrical performance requirements on the filler layers, they can be made from low cost easily obtained materials, such as FR-4 or its derivatives. Some high performance board materials can

be successfully sandwiched with FR-4. It is common to have a load board built with Rogers 4003 or 4350 stacked with FR-4.

The loss tangent or dissipation factor affects how much insertion loss will occur on the board. The bigger the load board and longer the traces, the more insertion loss and signal degradation will occur. Clock lines, which have sharp edges, are prevalent in digital designs. It is customary to route these signals differentially along the most direct path from point A to B to reduce the amount of noise generated. To make sure the clock lines have minimal degradation, vias and right angles should be avoided and a material with a low loss tangent should be chosen. If possible, routing should include a ground barrier between the differential pair and adjacent signals. In digital applications, where signal integrity is important, resistive loss of the traces should be reduced. However, sometimes there is a tradeoff between larger trace widths and matching the trace to device impedance, which most often is 50 Ohms.

In production environments, a hard substrate is required to handle the shock of repeated insertions, so Teflon based substrates should not be used. Hard materials with low dielectric constants and loss tangents are preferred and can be obtained from many different vendors. The materials come in varying thicknesses. Some standard thicknesses are 0.005", 0.008", 0.010", 0.015", 0.020", 0.032". If the board parameters vary over frequency, the best modeling solution for a broadband system would require modeling the load board with a frequency dependent materials table, which can be done in Ansoft HFSS on versions higher than 8.5. Each material's dielectric constant, loss tangent, and hardness should all be evaluated to create a load board system with the proper impedance to optimize device performance.

Optimize What's Important

Load board layout personnel and fabrication vendors spend a tremendous amount of time fretting over transmission line impedance. Typically, their measurement instrument is a time domain reflectometer, which is excellent at determining impedance, but is difficult to use for measuring actual transmission performance. For high speed and high frequency applications, trace losses and impedance discontinuities will cause more performance degradation than a trace that is not 50 ohms, as long as the trace has a constant impedance.

To illustrate this point, Figure 14 shows the eye patterns of six different characteristic impedances with the same trace length. There is no visible difference in the performance of this 3.2 Gbit/s Pseudo Random Bit Sequence (PRBS) at any of the six impedances.

However, Figures 15 and 16 show the impact made by a microstrip to stripline transition. Both test configurations include two vias and two inches of microstrip and stripline. In Figure 15, the signal via was flanked by six ground vias. In Figure 16, the ground vias were omitted. With reasonably considered grounding, via transitions from one layer to another cause negligible impact on even this 3.2Gbit/s signal. These two measurements demonstrate that "ground" can not be treated as if it exists universally. Instead, for every "signal" trace, you must consider the ground return current to be equally important, and ensure that there is a solid, uninterrupted path. Even though the traces are short, a load board with poorly conceived grounds is incapable of passing a signal properly.

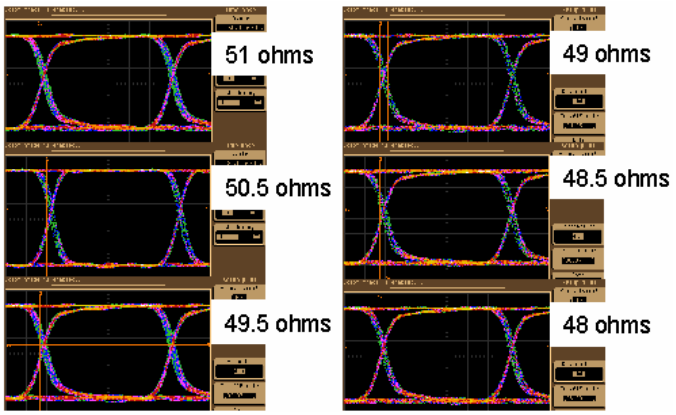


Figure 14; Effects of Trace Impedance on Signal Integrity

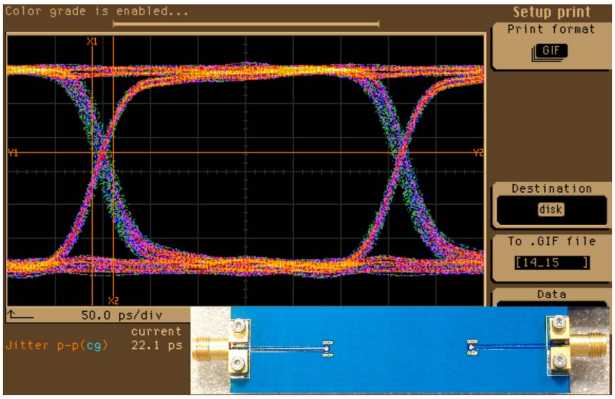


Figure 15: Eye Diagram of a Simple Load Board Structure with Vias Attaching Top and Bottom Ground Layers

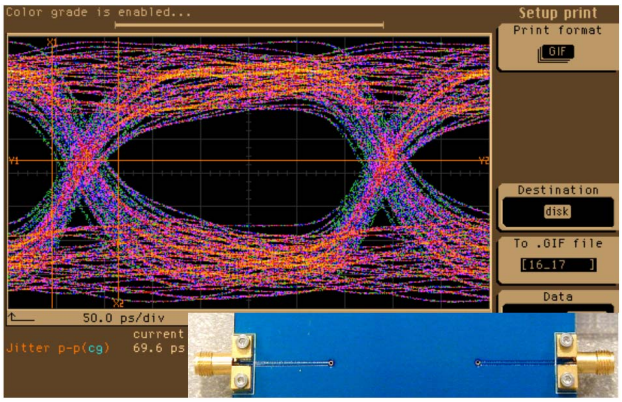


Figure 16: Eye Diagram of a Simple Load Board Structure with no Vias Attaching Top and Bottom Ground Layers

Validate the Design with Measurements

Figure 17 shows an optimized system (load board and contactor) for testing a device in a 192 ball 0.8mm pitch package at 10 Gbit/s BGA device. The connectors are edge launch, high performance extended bandwidth connectors to improve the match and allow signal harmonics to pass virtually undisturbed. Traces are routed differentially. Grounded pads were placed on either side of the differential pairs as a part of an experiment to diminish common mode signals that might be present; experiments demonstrated that they were not required.

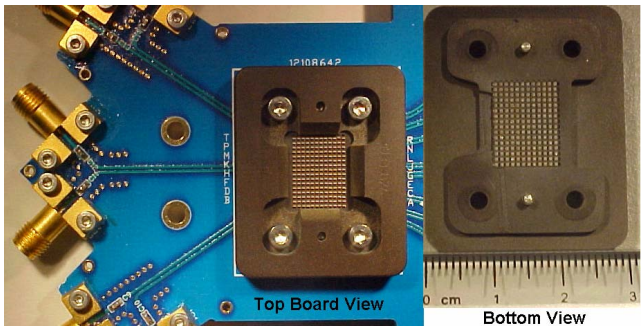


Figure 17: Optimized Load Board and Contactor

In this case, all the high frequency signals were present on the perimeter of the device, enabling direct routing on the topside of the board, with no vias in the signal paths. The trace width was chosen to match the size of the pads required by the contactor, then the substrate thickness was chosen to make that width produce 50 ohms. The presence of the contactor dielectric directly on the high-speed traces produced a four-ohm step in the line impedance. To eliminate this effect, the BGA contactor was milled to remove material that would have been in contact with the traces. Measurements demonstrated that the load board and contactor combination contributed approximately 5pS rise time to the 26pS rise time of the ATE system's pattern generator. Figure 18 shows the measurement of the 10Gbit/s board whose photograph is shown in Figure 17. The measurement shows a 35pS rise time, with no ringing, and minimal lossy line droop.

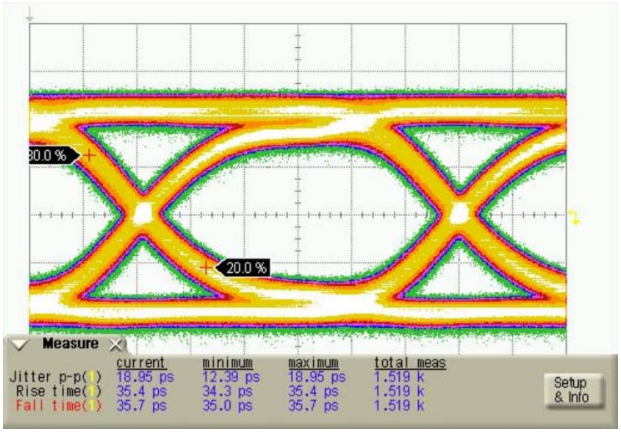


Figure 18: Measurement of Optimized Load Board, Contactor, and Surrogate Device with Through Lines

Contactor Effects on Electrical Performance

To optimize yields, the contactor must provide sufficient bandwidth for the frequencies or data rates of the devices to be tested. If the contactor bandwidth is barely adequate, then the normal variations in dimension from one packaged device to the next will severely impair system repeatability. In many high-speed digital or RF applications the contactor must not only handle the operating frequencies, but also the third and fifth harmonics. The width and pitch of the device will determine the width of the contact.

Modeling and simulation tools are useful for selecting and evaluating contactor performance. Figure 19 demonstrates the value of using modeling and simulation to compare scenarios, and then performing measurements to validate predicted results. Figure 19 demonstrates the minimal effect that a well-chosen contactor has on system performance.

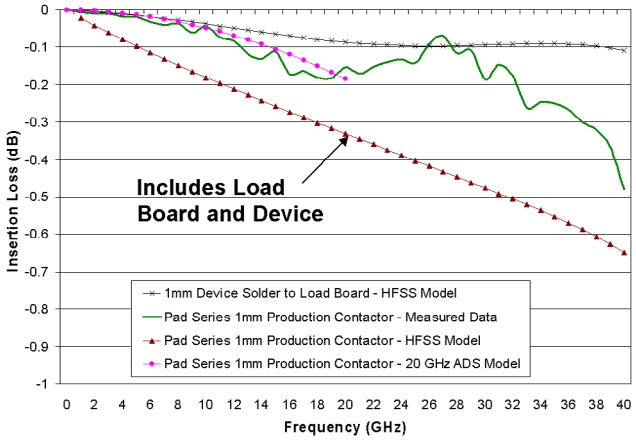


Figure 19: Comparing Modeled, Simulated and Measured Results

In Figure 19 one modeled data set shows the results of 3D electromagnetic modeling of the device pads directly soldered to the load board (simulating the performance obtained in the end-user environment). Another data set graphed shows the modeled performance with the contactor placed in between the device and the load board (simulating the performance obtained on the test floor). The difference between these two models represents the additional insertion loss introduced by the contactor. The other pair of traces shows reasonable correlation between a linear simulator's prediction of contactor performance versus actual measured contactor performance. In this graph, only the magnitudes are shown, but to make a valid correlation the phase must also be equally investigated.

Figure 20 shows how closely modeled data can match measured data when modeling is performed correctly. The load board layout and device pads were optimized to 50 ohms, which resulted in excellent correlation out to 20 GHz between the measured return loss data, the ADS simulation, and the HFSS 3D electromagnetic model. The 3D model matches the measured return loss to within 0.5 dB out to 25 GHz.

To optimize both electrical and mechanical connections, the widths of all the parts (device pad, contact pins, and load

board traces) in the system should be designed to closely match at the contact points. This reduces reflections, which improves return loss and signal integrity. In a recessed pad application, the thickness of contact pins will have to be smaller to hit the recessed pad but not the package or possibly the conformal coating on the bottom of the device. Sometimes sawed devices have a lip or some debris on the edge of the device pad, so the contactor must avoid contacting the pad edge to make a good electrical contact.

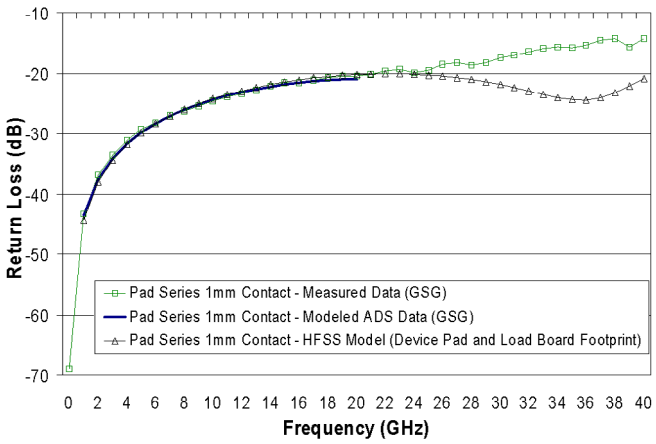


Figure 20: Accurately Modeled and Measured Data Correlate on a Pad Series 1mm Contactor

To improve the contactor's performance and increase its bandwidth, a number of effective methods can be implemented. First, the closer the 50 ohm load board trace width is to the contactor and device pad widths, the fewer mismatches will result, thereby increasing electrical performance. The tolerances for device-to-contactor and contactor-to-load board positioning will determine the width of the load board traces necessary for consistently good connections. Second, when the contactor rests on the load board, it creates a stripline effect, lowering the trace impedance. Either the load board trace can be adjusted to a smaller width, or a cutout in the housing material can be strategically placed over the high frequency lines to help preserve the impedance of the traces. Cutouts can also be used to allow matching part placement to be as close as possible to the DUT to simulate how customers will actually match and bypass the device. This will yield electrical results similar to results expected when the device is actually soldered to the customer's boards. Third, smoothing the contact and radiusing the edges will help improve the contact area and the flow of signals.

Contactor Effects on Repeatability

Figure 21 shows the effect of overall contactor yield based on device pad count with no cleaning or maintenance for three different device oxide levels and two different, yet similar, types of contacting technologies. These results are based on the yield being defined as a good connection. (i.e. a contact resistance measurement less than 0.02 ohms). Included in the data are measurements above 0.02 ohms that occurred after handler down time due to oxide formation on the contactor and test devices. Oxides will form on solder plated devices exposed to air. All measured results were taken using a Delta Flex Handler with Johnstech designed contactors, alignment plates, and other interfacing hardware, such as change kits. Also, Figure 21 is a summation of more than 100,000 insertions for all contactors with no cleaning or maintenance. Cleaning or maintenance would have removed debris, which caused some of the failures.

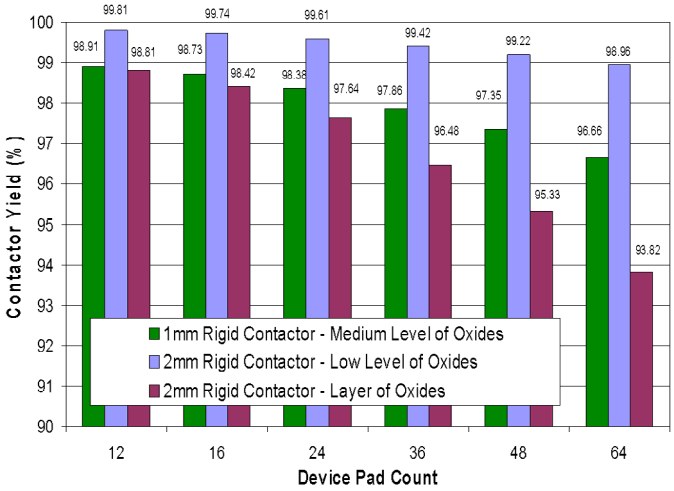


Figure 21: Effect of Contactor Yield on Device Pad Count

All devices used were 7x7mm 85/15 SnPb QFN packages with 48 pads. The data is based on a percentage of failures on both the corner pads and the middle pads of the device. Due to device coplanarity issues and alignment plate and package tolerances, the corner pads have a slightly lower probability of contact. With the contact technology used, the contact resistance standard deviations are close to 0.002 ohms. A low standard deviation means the connection between the device and contactor is extremely repeatable. Repeatable and low contact resistances are important parameters when testing analog or mixed signal devices.

If the contact resistance variation over time is known, it may be possible to attain higher yields and longer maintenance cycles for many applications without the using a two-wire Kelvin test method. This can be accomplished by programming the average contact resistance into the tester. After programming in the average contact resistance the errors are the deviation from the average contact resistance. In many cases, using the average contact resistance is better than the variability seen using a two or three piece contactor, which inherently has more contact resistance variability.

Contactor Effects on Grounding

With simulation software such as ADS, the device can be constructed as a black box of S-parameters with a reference that can be tied directly to ground to simulate soldering the device to the load board. During testing of the contactor, the reference ground is still the load board ground, but there is some additional inductance to ground from the test contactor. This inductance value will depend heavily on the housing height, number of contact pins tying device ground to the load

board ground, and type of technology used. In general, multiple ground connections add multiple inductances in parallel from the device to the load board ground. These multiple paths help reduce the effective inductance to ground. The lower the effective inductance, the closer the test data should match what will happen when the device is soldered to the load board. The longer the inductance path, the more inductance will be present. Therefore, for high frequency applications, or applications sensitive to inductance, use thin high performance substrates with low dielectric constants and loss tangents to minimize the inductance. Using thin substrates decreases the path from the top layer of the board to the ground plane layer. Adding more vias between the top ground layer and the load board ground plane will also effectively lower the inductance of the load board by creating more inductive paths to ground.

When testing devices that are very sensitive to ground inductance (i.e. amplifiers, receivers, and other RF devices), it is necessary to reduce the physical distance from the device ground to the load board ground plane. On the device, this can be accomplished by adding a body ground and more ground connections, such as bond wires or vias. On the contactor, this can be accomplished by decreasing the housing height and, if a center body ground is present, by placing an insert with contacts embedded into the insert to get Z-compliance and better repeatability. Creating multiple paths to ground will always lower the effective inductance to ground.

Figure 22 shows the effects of using the same contactor type, but adding more paths to ground to effectively lower the inductance to ground. All the modeled results were generated using the same matched load board so inductance changes could be determined. In the first plot, decreasing the inductance to ground allows the gain to improve by 0.3 to 0.75 dB. In the second plot, the Pad Series 1mm Production contactor has only 0.1 to 0.2 dB less gain in the amplifier bandwidth of 13.8 to 14.5 GHz, which is consistent with the actual measured insertion loss of the contactor.

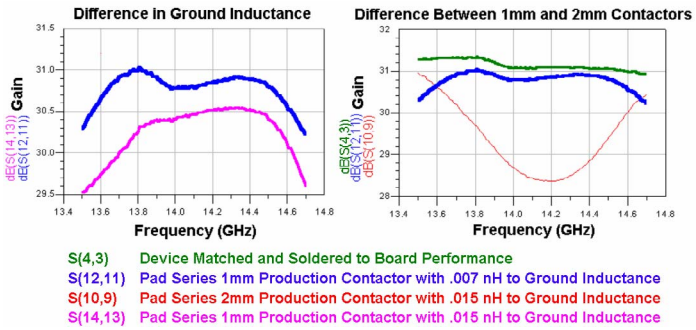


Figure 22: Effects of Ground and Peripheral Contact inductance on Amplifier Gain

The second plot shows a Pad Series 2mm contact, which has twice as much inductance to ground because its housing height is twice as high. Because the Pad Series 2mm housing height is twice as high, its peripheral contacts on the RF input and output have twice as much inductance resulting in a lower gain. The results would change slightly at different frequencies, but the trends would remain the same.

For high-speed digital devices, signal integrity and crosstalk are very important. At the device, a wider I/O pitch will result in lower crosstalk between pins or pads, but may not be feasible with package sizes shrinking. This is why Surface Acoustic Wave (SAW) filters have larger I/O pitch between inputs and outputs. To get as close to true device crosstalk as possible, a smaller contact width can be used to increase the distance between contacts. A lower housing dielectric can also be used, or a contact with a smaller surface area can be employed, to reduce both the inductive and capacitive effects the contactor will have on device performance.

Contactor Effects on Handler Interface

Handler presentation of the devices to the contactor will also have an impact on yield. In many cases, if the device presentation is not coplanar with the test contactor, or the contactor is not level (due to traces, solder bumps, conformal coating or non-uniform plating) the forces on some contactor contacts will tend to be higher. Higher forces will usually result in lower contact resistance numbers, but might have adverse effects on the load board life in production environments. If the devices and contactor are not coplanar, the contact resistance for each device pad could vary from one side of the device to the other. This variation could have some affect on DAC lines where input and output lines are on opposite ends of the device, or sensitive inputs and outputs are on multiple sides of the device.

The speed of device insertion will also affect performance and yield. Obviously, the faster companies can get parts in and out of testing the more they can ship to customers. If devices are inserted too fast, the handler might give a test ready signal to the tester before the device has stabilized in the contactor. This could result in false failures, causing the tester software to wait for the device to stabilize before accurate testing can begin. Also, higher speed insertions could translate to larger forces that wear out all portions of the system faster, including the contactor and load board. An optimum solution would be to have the handler move the device as fast as possible but then slow down the device just as it is inserted into the contactor. Many handlers have this option, or this function can be accomplished with a scheme to slow the part down mechanically. This last second slow down could allow parts to be tested as soon as possible after insertion into the contactor, and could cut milliseconds off the test time while improving repeatability and yields.

Contactor Effects on Mean Time Between Assists

Figure 23 shows over 120,000 contact resistance measurements for devices inserted into a contactor using a Delta Flex Handler. Without cleaning, the contact resistance tends to increase slightly before returning to its original state. It is also interesting to note that contact resistance will decrease slightly, then remain stable for a period of time. This is due to solder becoming deposited on the contact and slightly increasing the contact surface area. More surface area in contact with the device pad will result in lower contact resistance numbers for the same applied forces. With 85/15

tin-lead solder, the MTBA interval appears to be more than 100,000 insertions, but much depends on the device's tolerance to contact resistance variations.

Also, radiusing the edges on contacts will reduce fields radiating between contacts thereby improving electrical performance and reducing load board wear. Sharp edges on the contacts, poor handler presentation and a non-coplanar contactor could all cause premature load board wear and down time.

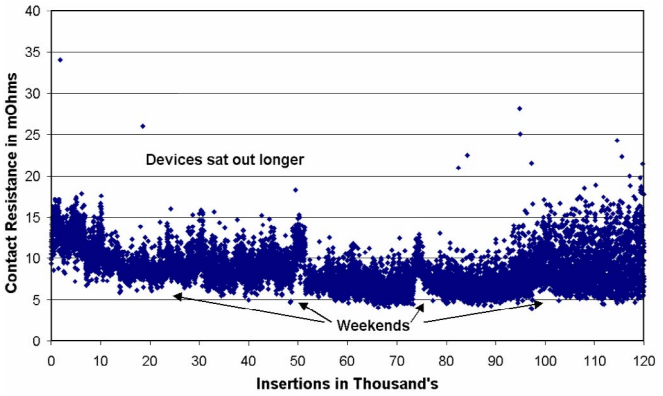


Figure 23: Contact Resistance Measurements Without Cleaning

Conclusions

Results show aspects of the interface between the tester and the device being tested help determine the cause of potential test system repeatability problems. Data from real applications showed the importance of many different aspects of test. Important features such as device package, solder plating, load board alignment and mounting hole tolerances, and inactivity can affect overall yield, even though the test engineer might not control them.

The repeatability of each segment of the overall test system will determine the final distribution of the device being tested. New contactor technologies offer good repeatability even without cleaning. Several methods of calibrating the test system are available to determine how repeatable the system is and factor out repeatable errors in the device measurements. Some of these methods might also be beneficial in determining cleaning and maintenance cycles of the contactor or the whole test system.

Modeling can help design the system right the first time, by predicting results and trends before hardware is ever assembled. Our data show the importance of choosing the correct contactor and load board materials to optimize both electrical and mechanical performance ensure testing success and decrease overall test costs.

Our data indicate that choosing the proper contactor or load board configuration could make testing closer to true device parameters. Accurate data can be obtained by using load board materials and contactor technologies that minimize insertion losses. Also, developing good grounding schemes in the load board and contactor is sometimes critical to accurately measuring device performance for devices sensitive to ground inductance.

Acknowledgments

The authors would like to thank all those at Agilent Technologies and Johnstech who collected data and provided comments for this paper.

References

1. Hall, S. H., et al, High-Speed Digital System Design: A handbook of interconnect theory and design practicies, John Wiley & Sons, Inc (New York, 2000)
2. Dalley, W. J., et al, Digital Systems Engineering, Cambridge University Press (Cambridge, 1998)
3. Bogatin, E., Signal Integrity–Simplified, Prentice Hall PTR (New York, 2003)