Confidence and Prediction Intervals for Microwave Calibrations and Measurements

Dylan Williams[1], Benjamin Jamroz[2], Jacob Rezac[3]
National Institute of Standards and Technology, Boulder, CO
[1]dylan.williams@nist.gov, [2]ben.jamroz@nist.gov, [3]jacob.rezac@nist.gov

***Abstract*— We discuss the estimation of confidence and prediction intervals for microwave calibrations and measurements from measurement residuals and highlight the difficulties of applying a conventional approach when correlations must be preserved in the uncertainty analysis. We then suggest an alternative and illustrate the procedure with probe-to-probe coupling corrections in an on-wafer measurement environment.**

***Index Terms*— Confidence intervals, coupling corrections, on-wafer measurement, prediction intervals.**

I. Introduction

Most microwave calibration models are imperfect or incomplete, making it difficult to accurately predict their performance from first principles. Examples include on-wafer calibrations in the presence of probe-to-probe coupling [1] and calibrations for over-the-air test. We present approaches for using measurement residuals to determine uncertainties in these situations.

Specifically, we present an approach for estimating confidence intervals for microwave calibration models from measurement residuals when correlations must be preserved, such as to propagate uncertainties through Fourier and other transformations critical in complex measurements of microwave communications systems, signals and metrics. We then discuss methods of determining the prediction intervals for calibrated measurements of a device under test (DUT) from measurement residuals. We illustrate the approach with on-wafer coupling corrections, which are notoriously difficult to model and correct for perfectly [1].

The terms "confidence intervals" and "prediction intervals" are used in the context of regression to differentiate between the estimated error in the curve fit to data (*i.e.* confidence intervals) and the estimated spread of future results around that curve (*i.e.* prediction intervals) [2]. We will adopt this nomenclature because the process of determining microwave calibration models is analogous to fitting a curve to *x*-*y* data, and the process of calibrating a DUT with this model is analogous to estimating future values in a regression problem. Thus, we will refer to estimates of uncertainty in microwave calibration models as confidence intervals and estimates of the uncertainty of calibrated results as prediction intervals.[2] In this context confidence intervals are akin to repeatability errors within an experiment and prediction intervals are akin to reproducibility errors between repeated experiments. As in regression problems, estimating the total uncertainty in a calibrated microwave measurand requires estimating both the confidence intervals for the calibration model and the prediction intervals for the DUT, and combining these uncertainties together to estimate the total uncertainty in a calibrated measurement of a DUT [2].

In the following, we will outline a general approach for determining the confidence intervals associated with microwave calibration models and prediction intervals associated with calibrated microwave DUTs from measurement residuals and maintaining correlations between uncertainties in the analysis. As we proceed, we will first illustrate the approach with simple regression examples, and then follow up by applying it to the microwave problem of correcting for probe-to-probe coupling using the on-wafer measurements performed in coplanar waveguide (CPW) in [1].

II. Coupling Corrections

We base our on-wafer coupling corrections on the "internal" 16-term calibration model of [1], a variant of the 16-term models introduced by Speciale and Butler, *et al.* in [3-5]. These coupling models are more appropriate in on-wafer environments than the classic 12-term calibration model developed for coaxial and rectangular-waveguide calibrations and have been applied to a variety of on-wafer measurement scenarios (*e.g.* [1, 6-9]).

In [1], only frequency-point-by-frequency-point confidence intervals were obtained from measurement residuals. Here we will develop both confidence and prediction intervals for the coupling-corrected measurements in a two-step process that separately estimates confidence intervals for the calibration model and prediction intervals for the calibrated result. The procedure also maintains correlations throughout the analysis, allowing the measurement uncertainties to be propagated through transformations such as the Fourier transform and to a variety of metrics, including error-vector magnitude (EVM) [10].

---

[1] U.S. Government work not protected by U.S. copyright.

[2] This choice of nomenclature can be somewhat misleading as we are not solely interested in developing traditional intervals that expect our results to lie between. We also determine probability distributions and maintain correlations in microwave measurement uncertainty and include these additional aspects of uncertainty when we use the terms confidence and prediction intervals in this paper.

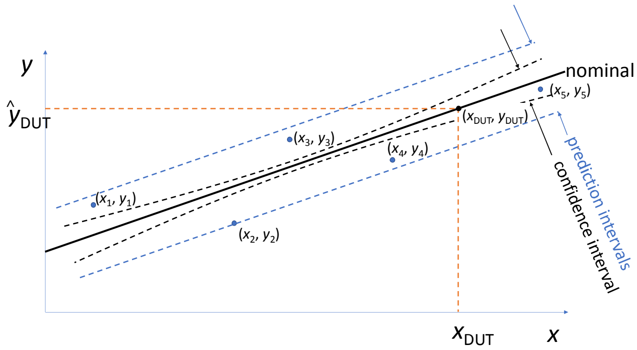

Fig. 1. A simple regression problem. The curve $y = a\,x + b$ is fit to the x-y pairs of data to estimate a and b. Then the curve $y = \hat{a}\,x + \hat{b}$ is used to estimate $\hat{y}_{\mathrm{DUT}}$, the estimate of the DUT's true value of y, from x_{DUT}.

III. Calibration Confidence Intervals

Figure 1 illustrates a classic scalar regression problem. The curve $y = ax + b$ is fit to five x-y pairs of data. If we assume that all the errors are in y, we can estimate the two coefficients a and b with linear regression. We will denote the estimates of the linear regression algorithm as $\hat{a}$ and $\hat{b}$.

When using linear regression, we can determine the uncertainty in our best estimate for the mean value $\hat{y}_{\mathrm{DUT}} = \hat{a}\,x_{\mathrm{DUT}} + \hat{b}$ analytically when measurement errors are independent and Gaussian. The result is

$$\mathrm{Var}(\hat{y}_{\mathrm{DUT}}) = \sigma^2\left(\frac{1}{n} + \frac{(x_{\mathrm{DUT}}-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right), \tag{1}$$

where σ^2 is the actual variance of the y data and n is the number of data points [2]. The first term $(1/n)\,\sigma^2$ in (1) is the uncertainty in the constant term $\hat{b}$, while the second term in (1) captures the impact of the uncertainty due to the slope $\hat{a}$. When needed, we can use the measurements residuals $\Delta y_i = y_i - (\hat{a}\,x_i + \hat{b})$ to estimate σ^2 from the sample variance of the y_i around the curve $\hat{a} + \hat{b}\,x_i$ with n-2 degrees of freedom. When using nonlinear regression, as is often the case in microwave problems, we can use a sensitivity analysis based on the measurement residuals with algorithms such as [11].

Either approach yields uncertainties on $\hat{y}_{\mathrm{DUT}}$, $\hat{a}$ and $\hat{b}$, and thus characterizes the confidence interval around the curve $y = ax + b$ illustrated as black dashed lines in Fig. 1. When sampling from a fixed distribution, confidence intervals on the curve in the figure will tend to decrease asymptotically as more data is used to estimate $\hat{a}$ and $\hat{b}$. Thus, we see that the width of the confidence intervals is a measure of the precision with which we have fit the curve $y = ax + b$, and are not a measure of how close the x-y pairs are expected to be to the curve $y = \hat{a}x + \hat{b}$.

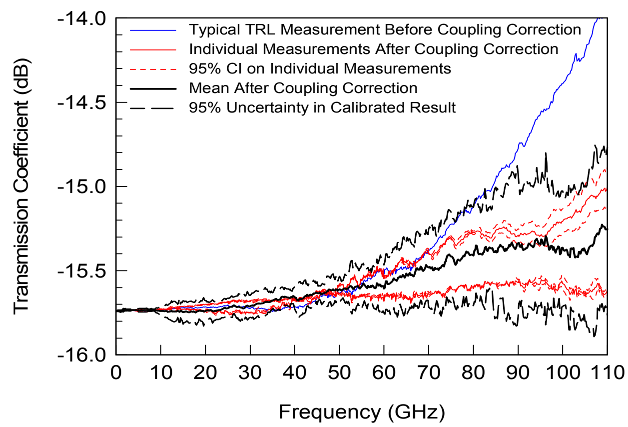

Fig. 2. Two measurements (red) of the attenuator investigated in [3] and their confidence intervals. Also shown are a measurement of the attenuator performed before coupling correction (blue), the mean of a set of measurements of the attenuator (black) and the total estimated uncertainty based on contributions from both the confidence and prediction intervals (dashed black).

A. Confidence Intervals for On-Wafer Coupling Corrections

We based our microwave experiments on the CPW data from [1]. We started by using the ALGLIB[3] Improved Levenberg-Marquardt Optimizer to find a solution for the four complex coupling terms in the internal 16-term coupling-correction model based on non-linear optimization at each frequency point. We did this by using the ALGLIB algorithm to fit the coupling terms in the model to TRL-corrected measurements of the CPW open-open, short-short, load-load, open-short, load-open and load-short crosstalk calibration artifacts used in [1]. We studied the convergence of the algorithm and verified these fits against results from [1]. Finally, we applied the coupling corrections to a small printed attenuator, which served as a DUT in the experiments. Previous simulations indicated that, in the absence of external coupling, these attenuators should have a nearly flat magnitude response and linear phase response (see Fig. 3 of [1]).

B. Applying the Calibration to a DUT

Returning to our linear-regression analogy, performing the crosstalk correction is analogous to measuring x_{DUT}, the value of x corresponding to a given DUT, and then using $\hat{a}$ and $\hat{b}$ estimated from the curve fit to find the estimated mean response $\hat{y}_{\mathrm{DUT}} = \hat{a}\,x_{\mathrm{DUT}} + \hat{b}$, as illustrated in Fig. 1. Likewise, the confidence intervals on the estimated mean response $\hat{y}_{\mathrm{DUT}}$ *only* estimates the uncertainty in this *mean* response. Thus, we anticipate that the confidence interval for the estimated mean response $\hat{y}_{\mathrm{DUT}}$ will be considerably smaller than the spread of future values of the y_i around that mean response and will underestimate the actual error in $\hat{y}_{\mathrm{DUT}}$.

[3] We use brand names only to better specify how we performed our experiments and analysis. The National Institute of Standards and Technology does not endorse commercial products. Other products may work as well or better.

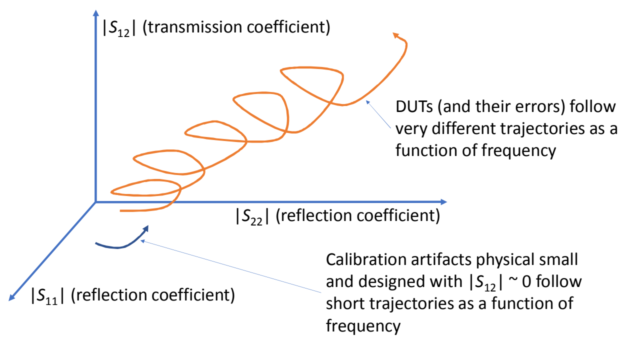

Fig. 3. An illustration of the trajectories taken by the calibration artifacts and the DUTs as a function of frequency through the scattering-parameter space.

Figure 2 compares the TRL-corrected attenuator response before and after applying the coupling correction in the two experiments. As can be seen from the figure, the coupling corrections largely flatten out the magnitude response, as we hoped.

A number of calibrations performed with the CPW supported by different chuck materials and employing different CPW access-line lengths were tested in [1]. Attenuator measurements corrected with two of those calibrations, along with their confidence intervals, are shown in thin red solid and dashed lines, respectively, in Fig. 2. As expected, the confidence intervals for the DUT magnitude response appear to greatly underestimate the deviations of the DUT's response. In fact, above 60 GHz where the coupling corrections become most significant, the 95% confidence intervals of the two measurements never overlap.

IV. Prediction Intervals

Figure 2 clearly indicates that the confidence intervals for the coupling corrections will underestimate the actual errors in the coupling-corrected DUT. The prediction intervals, on the other hand, capture the anticipated variation of the corrected measurements around the mean calibrated result (see the dashed blue lines in Fig. 1) due to lack of fit of the calibration model to measured data. To develop the total uncertainty in the calibrated result, we must combine the uncertainties around the mean indicated by the prediction intervals with the uncertainty in that mean indicated by the confidence intervals.

A. Conventional approach to estimating prediction intervals

In Fig. 1, the variance associated with the prediction interval is given by σ^2, the actual variance of the y data and the total uncertainty in y_{DUT} is found by summing σ^2 , which is related to the variation of the y_i around the curve $y = \hat{a}\,x + \hat{b}$, and the variances associated with the confidence interval given in (1) related to the variation in the curve fit. The result is (see Section 6.2.4 of [2])

$$\mathrm{Var}(y_{\mathrm{DUT}}) = \sigma^2\left(1 + \frac{1}{n} + \frac{(x_{\mathrm{DUT}}-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right). \quad (2)$$

Equation (2) is extremely convenient. It gives the uncertainty for any y_{DUT} based solely on the statistics of the *x*-*y* pairs used to fit the curve $y = ax + b$.

However, our problem is different in that each *x*-*y* pair of scalars defining a point in Fig. 1 corresponds to a complex trajectory of points that travels through a multi-dimensional space of scattering parameters as a function of frequency in our microwave problem, as illustrated in Fig. 3.

The calibration artifacts used to determine the calibration model are physically small devices with low transmission coefficients and a slowly varying frequency response. Their trajectories through the multi-dimensional scattering-parameter space as a function of frequency (indicated by blue trajectory in Fig. 3) are relatively simple and restricted to the volume for which the transmission coefficients are small.

The DUT, on the other hand, may have large or small transmission coefficients, and thus will take trajectories as a function of frequency that may travel through a large portion of the multi-dimensional scattering-parameter space (indicated by the orange trajectory in Fig. 3), many of which are never visited by the coupling-correction calibration artifacts. Furthermore, as DUTs are often physically large, they may take much more complex trajectories as a function of frequency through that larger multi-dimensional scattering-parameter space. Thus, we see that we cannot expect any of our calibration artifacts to take trajectories through the multi-dimensional microwave scattering-parameter space similar to those taken by the DUT.

Because we want to preserve correlations in the uncertainties, which also take trajectories through this same multi-dimensional scattering-parameter space that are tightly linked to the nominal response, we cannot expect to construct uncertainties from the calibration artifacts that will mimic the correlations of the errors in the estimated DUT response. We conclude that, unlike the scalar case, we cannot determine uncertainties solely from measured residuals in the calibration.

B. Alternative approach to estimating prediction intervals

Developing analytic models for underlying uncertainty mechanisms is one of the most common alternatives to the conventional method of estimating prediction intervals. However, in many cases, such as in interlaboratory comparisons [12] and common microwave calibrations, including our coupling-correction problem, understanding the source of the underlying uncertainty mechanisms can be extremely challenging, and is precisely why we are interested in using measurement residuals for this task. We borrow an idea from interlaboratory comparisons to devise experiments which allow us to estimate prediction intervals despite the difficulties discussed above; without giving details, we demonstrate that prediction intervals can be calculated from these experiments.

In the case of interlaboratory comparisons, the most common approach is to perform separate measurements of the *same* DUT in different laboratories, and base prediction intervals on the differences between those calibrated measurements [12]. This has the advantage of guaranteeing not only that the measurement errors themselves are independent, but that the corrected measurements of the DUTs are similar and share the same structure, allowing correlated uncertainties to be extracted from their differences. This is the approach we will follow here.

In the context of probe-to-probe coupling, we examined the measurements performed in [1] with different access-line lengths and different chuck materials. These changes in the experiments maintained roughly the same overall calibrated response of the DUT while varying the coupling levels and error mechanisms in the experiments. Then we calibrated the DUT in each experiment using the internal 16-term error model. We were careful to perform all calibration and DUT measurements with the same chuck material and access-line length in each experiment, and to set the reference planes for the measurements in the same place relative to the DUT. Thus, we expect each of the calibrated results to follow the same trajectory in the multi-dimensional scattering-parameter space as a function of frequency, and we can use the variations between those trajectories to estimate prediction intervals and their correlations for that DUT.

Finally, based on the assumption that the corrected DUT response in each experiment is equally valid and that the uncertainties related to the confidence intervals in each of the experiments were highly correlated between experiments, which we verified with a separate analysis,[4] we estimated prediction intervals and combined them with the confidence intervals to predict the total uncertainty in our attenuator measurements. The total uncertainty estimated by this approach for the transmission through the attenuator is shown in the thin black dashed lines in Fig. 2. This shows a level of uncertainty consistent with the actual deviations between the different experiments we used, and do not greatly underestimate the total uncertainty, as do the confidence intervals shown in red.

V. Conclusion

We examined the common case of microwave calibrations with imperfect or incomplete calibration models. We presented an approach for estimating confidence intervals for microwave calibration models and prediction intervals for calibrated results from measurement residuals when correlations must be preserved, such as to propagate uncertainties through Fourier and other transformations critical in complex measurements of microwave communications systems, signals and metrics. We argued that, unlike in the simple scalar case, the total correlated uncertainty in calibrated microwave measurements requires auxiliary measurements of the DUT and cannot be determined from calibration data alone.

While we illustrated the approach with on-wafer coupling corrections, it is quite general, and can be applied broadly to microwave measurements based on imperfect or incomplete microwave calibration models. Other examples include vector-network-analyzer calibration algorithms based on imperfect calibration models and calibrations for over-the-air test.

Acknowledgment

We thank Michael Frey, Adam Pintar and Amanda Koepke for their help with the statistical concepts we discuss.

References

[1] D. F. Williams, F. J. Schmuckle, R. Doerner, U. Arz, and W. Heinrich, "Crosstalk Corrections for Coplanar-Waveguide Scattering-Parameter Calibrations," *IEEE Transactions on Microwave Theory and Techniques,* vol. 62, no. 8, pp. 1748-1761, 2012.

[2] D. C. R. Montgomery, G.C.;Hubele,N.F., *Engineering Statistics*, 5 ed. Hoboken, NJ: John Wiley and Sons, 2011.

[3] R. A. Speciale, "A Generalization of the TSD Network-Analyzer Calibration Procedure, Covering n-Port Scattering-Parameter Measurements, Affected by Leakage Errors," *IEEE Trans.Microw.Theory Techn.,* vol. 25, no. 12, pp. 1100-1115, 1977.

[4] J. V. Butler, D. Rytting, M. F. Iskander, R. D. Pollard, and M. Vanden Bossche, "16-term error model and calibration procedure for on wafer network analysis measurements [MMICs]," *Int.Microw.Symposium Digest,* vol. 3, pp. 1125-1127, 1991.

[5] J. V. Butler, D. Rytting, M. F. Iskander, R. D. Pollard, and M. Vanden Bossche, "16-term error model and calibration procedure for on-wafer network analysis measurements," *IEEE Trans.Microw.Theory Techn.,* vol. 39, no. 12, pp. 2211-2217, 1991.

[6] L. Hayden, "Modal calibration of GSSG probes," presented at the ARFTG Microwave Measurement Conference, 2010.

[7] A. Rumiantsev, R. Doerner, and E. M. Godshalk, "Influence of calibration substrate boundary conditions on CPW characteristics and calibration accuracy at mm-wave frequencies," *Automatic RF Techniques Conference Digest,* vol. 72, 2008.

[8] E. M. Godshalk, "Surface wave phenomenon in wafer probing environments," *Automatic RF Techniques Group Microwave Measurement Conference Digest,* vol. 40, pp. 10-19, 1992.

[9] D. F. Williams, A. C. Young, and M. Urteaga, "A Prescription for Sub-Millimeter-Wave Transistor Characterization," *IEEE Trans.THz Sci.Technol.,* pp. 433-439, 2013.

[10] K. A. Remley, D. F. Williams, P. D. Hale, C. M. Wang, J. Jargon, and Y. Park, "Millimeter-Wave Modulated-Signal and Error-Vector-Magnitude Measurement With Uncertainty," *IEEE Transactions on Microwave Theory and Techniques,* vol. 63, no. 5, pp. 1710-1720, 2015.

[11] P. T. Boggs, R. H. Byrd, J. E. Rogers, and R. B. Schnabel, "User's reference guide for ODRPACK version 2.01 software for weighted orthogonal distance regression," Boulder, CONISTIR 92-4834, 6/1992 1992.

[12] M. Thompson and S. L. R. Ellison, "Dark uncertainty," *Accreditation and Quality Assurance,* journal article vol. 16, no. 10, p. 483, 2011.

[4] The prediction intervals would have captured the impact of confidence intervals that were uncorrelated between each of the experiments.