A bilateral comparison of on-wafer $S$-parameter measurements at millimeter wavelengths

Roland G. Clarke$^1$, Jemeela Quraishi$^2$ and Nick M. Ridler$^2$
$^1$University of Leeds, UK; $^2$National Physical Laboratory, UK
Email: R.G.Clarke@leeds.ac.uk and nick.ridler@ieee.org

Abstract—This paper reports on a comparison of measurements that has taken place recently between two UK-based measurement facilities – the University of Leeds and the National Physical Laboratory. The comparison involved making complex $S$-parameter measurements of a commercially available co-planar waveguide calibration substrate. Most measurements were made at frequencies up to 65 GHz, although some measurement data was obtained up to 110 GHz. Subsidiary investigations also looked at measurement repeatability and the effects of using different VNA calibration schemes. A breakdown of the likely error processes affecting these measurements as a function of frequency is also given leading to rudimentary uncertainty estimates for such measurements.

Index terms—measurement comparisons, co-planar waveguide, on-wafer measurement, $S$-parameter measurement

I. INTRODUCTION

In recent years, many new applications have emerged that exploit large proportions of the millimetre-wave region, i.e. to 110 GHz and sometimes beyond [1]. These applications are driving demands for reliable measurement capabilities with well-understood accuracy limits. A large proportion of these applications utilise planar circuit geometries (co-planar waveguide, microstrip, etc) that require measurements to be made using probe stations attached to Vector Network Analyzers via cables connected to wafer probes. Such a system is able to make measurements (of $S$-parameters, etc) using probes of the correct size and configuration to interface with the transmission lines and components found on-wafer.

However, to date, demonstrating the accuracy and assuring the reliability of measurements in these on-wafer architectures has remained a challenge. This is partly due to the lack of a classical traceability infrastructure being in place for these measurements.$^1$

In addition, a bewildering variety of circuit conditions affecting the definitions of such measurements (e.g. the choice of transmission line; the number, sizes and spacings of the conductors; the choice of substrate material(s); frequency ranges; etc) means that it is unrealistic to expect measurement references for all such conditions.

Recognition of this situation has led researchers and measurement specialists to investigate alternative methods to achieve metrics for the accuracy and assurance for these measurements. One such method is through the use of measurement comparison programmes between users of a given measurement architecture. This paper reports on such a comparison between two measurement facilities involved in co-planar measurements to 110 GHz and beyond. It is hoped that this work will be extended in the future to include other participants.

II. COMPARISON DETAILS

The comparison consisted of performing a series of measurements on devices found on a GGB Industries co-planar waveguide (CPW) impedance standard substrate (ISS) model CS-5 [2]. The devices were short-circuits, open-circuits, ‘matched’ loads and thru connections. Measurements were made at the University of Leeds (UoL) and the National Physical Laboratory (NPL) under the following conditions:

- using different calibration schemes:
  - Short-Open-Load-Thru (SOLT)
  - Line-Reflect-Match (LRM) [3-5] (using either a short-circuit or an open-circuit as the Reflect standard)
  - Line-Reflect-Reflect-Match (LRRM) [6]
- over different bandwidths:
  - 40 MHz to 65 GHz
  - 45 MHz to 110 GHz

$^1$ Such an infrastructure relates measurements made locally with similar measurements made elsewhere, as well as establishing links to other measurement quantities and ultimately the SI base units (i.e. the metre, ampere, second, etc).
under repeatability conditions (as defined in [7]) involving replicate measurements.

Measurements were made on three sets of Open-Short-Load-Thru devices found on the ISS. These therefore were the devices under test (DUTs) for the comparison. However, this paper presents an analysis of only a small subset of the overall data gathered during the comparison exercise, as described below.

First, results are presented for selected measurements to 65 GHz using the LRRM calibration technique. This is the technique that is often used by both UoL and NPL for routine measurement investigations. A subsidiary investigation by one participant (NPL) into the observed repeatability for these measurements is also given. This provides a benchmark for the variability observed between both participants.

Some discussion of the likely error processes affecting these measurements is then given leading to rudimentary uncertainty estimates for these measurements. Finally, the effect of using different calibration schemes (i.e. LRRM, LRM and SOLT) is considering including some selected results of measurements to 110 GHz made by UoL.

Measurements were made initially by NPL followed by UoL. NPL then made a second set of measurements to complete the comparison exercise. These repeat measurements showed little difference for the first set of measurements demonstrating the continued ‘health’ of the ISS throughout this comparison exercise. This is an important cross-check to make since it is well known that the performance of on-wafer devices can deteriorate significantly due to mechanical effects caused by repeated probing.

III. MEASUREMENT DETAILS

Information on the VNAs and probes used during the comparison exercise is given in Tables 1 and 2.

<table>
<thead>
<tr>
<th>Table 1: VNA details</th>
</tr>
</thead>
<tbody>
<tr>
<td>VNA</td>
</tr>
<tr>
<td>UoL Agilent 8510XF</td>
</tr>
<tr>
<td>NPL Anritsu 37397C</td>
</tr>
</tbody>
</table>

For the LRRM calibrations, UoL used a Cascade Microtech ISS whereas NPL used a set of standards on the GGB Industries ISS that was used to provide the DUTs in the comparison exercise (albeit a different set from those used as the DUTs). Only one set of DUTs (i.e. Open-Short-Load-Thru devices) is discussed in this paper.

IV. RESULTS – LRRM CALIBRATIONS

A. S-parameter magnitude measurements

The results obtained using LRRM calibrations for the measured magnitudes of the voltage reflection coefficients (VRCs), $S_{11}$ and $S_{22}$, for the shorts, opens, loads and thru are shown in Figures 1 to 4, respectively.

---

2 There is not much published material on measurement comparison exercises of S-parameters in CPW. However, three recent related such activities have been reported in [8-10].

3 There is not much published material on measurement comparison exercises of S-parameters in CPW. However, three recent related such activities have been reported in [8-10].
The most noticeable feature with these results is that the magnitude reflection coefficient values measured by both UoL and NPL for the shorts and opens show many values that are significantly different from the nominal value of 1. This suggests the presence of significant systematic errors in these measurements, possible emanating from assumptions in the calibration algorithm that are not been fully met.

The above results are summarized in Table 3 in terms of the maximum difference between the UoL and NPL values over three bandwidths, to; 50 GHz, 60 GHz and 65 GHz.

<table>
<thead>
<tr>
<th></th>
<th>≤ 50 GHz</th>
<th>≤ 60 GHz</th>
<th>≤ 65 GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shorts</td>
<td>0.02</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Opens</td>
<td>0.02</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Loads</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Thru</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>

The results obtained using LRRM calibrations for the measured magnitudes of the voltage transmission coefficients (VTC), $S_{21}$ and $S_{12}$, for the thru are shown in Figure 5.

As before, these results are summarized in Table 4 in terms of the maximum difference between the UoL and NPL values over three bandwidths, to; 50 GHz, 60 GHz and 65 GHz.

<table>
<thead>
<tr>
<th></th>
<th>≤ 50 GHz</th>
<th>≤ 60 GHz</th>
<th>≤ 65 GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thru</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

B. S-parameter phase measurements

In general, the S-parameter phase measurements made by both participants showed good agreement with each other. For the shorts and opens, the measured reflection coefficient phase values were also compared with nominal values. For the shorts,
nominal values for the reflection coefficient phase, $\phi$, as a function of frequency, $f$, can be established using

$$\phi = 2 \tan^{-1} \left( \frac{Z_0}{2\pi L f} \right)$$

where $Z_0$ is the characteristic impedance (assumed to be 50 ohms) and $L$ is the inductance of the short. The manufacturer’s specification for the ISS [2] gives $L = 5 \text{ pH}$. So, for example, at 65 GHz, the nominal value for the phase is $\phi = +175^\circ$. Similarly, for the opens, a nominal value for the reflection coefficient phase, $\phi$, can be established using

$$\phi = -2 \tan^{-1} (2\pi Z_0 C f)$$

where $C$ is the capacitance of the open. The manufacturer’s specification for the ISS [2] gives $C = 6.5 \text{ fF}$. So, for example, at 65 GHz, the nominal value for the phase is $\phi = -15^\circ$. For both shorts and opens, the measured values were within $5^\circ$ of these nominal values at 65 GHz.

For the loads and thru, since the magnitudes of the reflection coefficients are small (i.e. close to zero), the phases should be difficult to determine. This is because the magnitude of the vector representing the reflection coefficient is close to zero and so a small change in the (spatial) position of the vector can cause a large change in the phase, making the observed phase appear unstable. This was indeed the case for the measurement of the thru. This can also be seen in Figure 6 for the load below 10 GHz (where, from Figure 3, the magnitude of the reflection coefficient is less than 0.005), where the observed phase varies erratically, often taking apparently random values anywhere between $-180^\circ$ and $+180^\circ$. However, above 10 GHz, the measured phases of both $S_{11}$ and $S_{22}$ settle down to a near-constant value of approximately $-90^\circ$, independent of frequency. It is not clear why this is happening since the phase of a (measurable) vector usually rotates clockwise about the origin of the complex reflection coefficient plane according to Foster’s Reactance Theorem [11]$.^4$ Since this is not happening on this occasion, this may indicate an error in the overall measurement process, e.g. caused by an assumption in the calibration scheme that is not being fully met. This requires further investigation.

\[\text{Fig 6. Loads reflection coefficient phases}\]

In a similar way to the shorts and the opens, it is possible to provide a nominal value for the phase, $\theta$, of the transmission coefficients for the thru. The manufacturer’s specification for the ISS [2] states that the propagation velocity is 0.442 that of light. This means that the effective electrical length, $l$, is given by:

$$l = \frac{150}{0.442} \mu\text{m} = 340 \mu\text{m}$$

So, for example, at 65 GHz, the free space wavelength, $\lambda$, is approximately 4.6 mm, and so:

$$\theta = \frac{-360 \times l}{\lambda} = -\frac{360 \times 340}{4.6 \times 10^3} \times 10^{-6} = -27^\circ$$

The measured values for the thru were within $5^\circ$ of this nominal value at 65 GHz.

V. MEASUREMENT REPEATABILITY ASSESSMENTS

Some replicate measurements were made by NPL to 65 GHz under repeatability conditions [7]. Summary worst-case standard deviations for these repeated measurements are given in tables 5 and 6.

$^4$ This theorem states that, for a general one-port reactive termination, the rate of change of reactance and susceptance with frequency will be positive.
It is interesting to compare the repeatability values in Tables 5 and 6 with the between-participant difference values in Tables 3 and 4. In all cases, the difference values are larger than the repeatability values. This suggests the presence of significant systematic errors in these measurements. However, as the bandwidth of the measurements increases to include measurements above 60 GHz, the sizes of the repeatability standard deviations begin to approach the between-participant difference values suggesting that repeatability errors are becoming more significant at these higher frequencies (i.e. above 60 GHz). This degradation in repeatability of the measurements can also been seen in Figures 1 to 4 where the measurement traces become more ‘noisy’ (indicating the onset of significant random errors affecting the measurement processes) as the frequency rises towards 65 GHz.

VI. UNCERTAINTY PREDICTIONS

We can further use the information in the above tables to construct some rudimentary estimates for the likely sizes of the uncertainties for these measurements. If we assume that by statistically analyzing measurement data from a series of different participants we are effectively randomising the systematic errors that may be present in the values of the participants, we can establish estimates for the likely overall sizes for such systematic errors. Now, in our case, we have only two participants, so that differences between their values define the sample range. We can convert a sample range to the equivalent effective standard deviation, $\sigma$, using values from statistical tables (see, for example [12]). In the case of a sample size of two, the multiplying factor to convert the range to the equivalent standard deviation is 0.89. This enables us to establish an effective standard deviation, $\sigma_e$, representing the systematic error processes in the measurements. Such a standard deviation can then be combined with the associated standard deviation, $\sigma_r$, due to the random error processes to provide an effective standard deviation, $\sigma_c$, for the combined errors, using

$$\sigma_c = \sqrt{\sigma_s^2 + \sigma_r^2}$$

i.e. the so-called Root-Sum-Squares (RSS) approach to combining error (uncertainty) contributions. Finally, an estimate of the ‘uncertainty’, $U$, can be achieved by multiplying $\sigma_c$ by a coverage factor, $k$. To obtain an uncertainty at a 95 percent level of confidence (95% cl), a coverage factor $k = 2$ is used.

The above process has been applied to the reflection coefficient values in Tables 3 and 5 for each of the three bandwidths, ≤ 50 GHz, ≤ 60 GHz and ≤ 65 GHz. In each case, the largest values of difference and standard deviation for all four devices have been used. The resulting values of $\sigma_e$, $\sigma_r$, $\sigma_c$ and predicted uncertainty are shown in Table 7.

It is recognised that the predicted ‘uncertainty’ values in Table 7 have been established based only on very limited information. In fact, in the absence of any well defined traceability route for these measurements (i.e. traceable back to, and harmonized with, SI base units), it is probably more appropriate to consider these ‘uncertainty’ intervals more in the context of a “reproducibility limit” for a standard measurement method [14] within which one would expect most other measured values to fall (i.e. values supplied by other measurement facilities in good working order). As such, these ‘uncertainty’ estimates are considered to be useful values against which other future uncertainty estimates (and other measurement quality metrics) can be compared.

A similar uncertainty evaluation process has not been applied to the transmission coefficient data since information is only available at one value of transmission (i.e. a transmission coefficient with magnitude of 1). Since the uncertainty in a transmission coefficient measurement will vary significantly as a function of the magnitude of the

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6 The “reproducibility limit” is defined (in [13]) as the value less than or equal to which the absolute difference between two measurement results obtained with the same method on identical DUTs in different laboratories with different operators using different equipment is expected to be, with a probability of 95 %.

---

**Table 5: Worst-case standard deviations for reflection coefficient magnitudes**

<table>
<thead>
<tr>
<th></th>
<th>≤ 50 GHz</th>
<th>≤ 60 GHz</th>
<th>≤ 65 GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shorts</td>
<td>0.002</td>
<td>0.004</td>
<td>0.012</td>
</tr>
<tr>
<td>Opens</td>
<td>0.002</td>
<td>0.004</td>
<td>0.016</td>
</tr>
<tr>
<td>Loads</td>
<td>0.004</td>
<td>0.006</td>
<td>0.014</td>
</tr>
<tr>
<td>Thru</td>
<td>0.001</td>
<td>0.002</td>
<td>0.012</td>
</tr>
</tbody>
</table>

**Table 6: Worst-case standard deviations for transmission coefficient magnitudes**

<table>
<thead>
<tr>
<th></th>
<th>≤ 50 GHz</th>
<th>≤ 60 GHz</th>
<th>≤ 65 GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thru</td>
<td>0.001</td>
<td>0.001</td>
<td>0.010</td>
</tr>
</tbody>
</table>

**Table 7: Uncertainty predictions for reflection coefficient measurements using LRRM calibrations**

<table>
<thead>
<tr>
<th>Freq range (GHz)</th>
<th>“1-σ” error assessments</th>
<th>Predicted Uncertainty (95% cl)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Systematic, $\sigma_s$</td>
<td>Random, $\sigma_r$</td>
</tr>
<tr>
<td>≤ 50</td>
<td>0.018</td>
<td>0.004</td>
</tr>
<tr>
<td>≤ 60</td>
<td>0.045</td>
<td>0.006</td>
</tr>
<tr>
<td>≤ 65</td>
<td>0.053</td>
<td>0.016</td>
</tr>
</tbody>
</table>
transmission coefficient, a single summarizing value of uncertainty is considered to be inappropriate.

VII. COMPARISONS BETWEEN LRRM/SOLT/LRM AND EXTENSION TO 110 GHz

Some measurements of the DUTs were made by NPL (to 65 GHz) and UoL (to 110 GHz) to compare the LRM\(^7\), LRRM and SOLT calibration techniques. As an example, Figure 7 shows the \(|S_{11}|\) measurements made by UoL of a short to 110 GHz for these three calibration techniques. The measurements made by NPL to 65 GHz showed similar trends.

As before, the measurements made using LRRM show many magnitude reflection coefficient values that are significantly different from the nominal value of 1. The same can be said for the measurements made using LRM. In fact, the measurements made using LRRM and LRM track each other very well with frequency. This suggests the presence of similar systematic errors in these measurements/calibration schemes.

However, the measurements made using SOLT generally provide a closer agreement to the nominal value of 1 at most frequencies and therefore do not track the measurements made using LRRM and LRM. This is likely to be due to a short being used as an assumed ‘known’ device (with a reflection coefficient magnitude value of 1) during the SOLT calibration scheme and so it is perhaps not surprising that this scheme returns a value close to 1 during a subsequent measurement of a short (albeit a physically different short on the same calibration substrate). However, it does suggest that there could be some benefit in using a high-reflect as a known device during an on-wafer VNA calibration scheme if this reduces systematic errors that produce measurements that defy physical constraints (i.e. the magnitude of a reflection coefficient must not be greater than 1 for a passive device).

VIII. CONCLUSIONS

A comparison of on-wafer \(S\)-parameter measurements has been undertaken by two UK-based measurement facilities (UoL and NPL). Extensive data has been collected by both facilities to 65 GHz. UoL also collected data to 110 GHz that can be used should this comparison activity be extended in the future. Subsidiary investigations looked at measurement repeatability and the effects of using different VNA calibration schemes. Most of the data presented related to the use of the LRRM calibration technique, as UoL and NPL often use this technique for performing routine measurements.

From the work presented in this paper, the following summaries can be made:

- The maximum observed difference between measurements made by UoL and NPL was of the order of 0.06 (reflection coefficient magnitude). This suggests an anticipated maximum uncertainty of measurement (at a 95 percent level of confidence) of approximately 0.11.
- At frequencies up to 60 GHz, the differences between measurements made by UoL and NPL are much larger than the repeatability standard deviations observed by NPL. This suggests that systematic errors dominated the measurements at these frequencies.
- At frequencies above approximately 60 GHz, the repeatability standard deviations observed by NPL increase significantly, as does the ‘noise’ on the plots in Figures 1 to 4 for both UoL and NPL, and begin to become comparable with the differences between the UoL and NPL measurements. This suggests that both random and systematic errors had a significant effect on the measurements at these frequencies.

\(^7\) The LRM calibration shown here used a short as the Reflect standard – hence the label “LRMS” used to identify the measurements in Figure 7. However, LRM calibrations made using an open as the Reflect standard showed very similar behaviour to the results obtained using a short as the Reflect standard.
Measurements made using LRRM contain significant systematic errors causing the magnitude of the reflection coefficients to depart significantly from 1 for the shorts and opens.

Measurements made using LRM also contained significant systematic errors similar to those observed in LRRM. However, measurements made using SOLT did not appear to contain these errors.

Phase values for both reflection and transmission coefficient measurements to 65 GHz were all within 5° of their nominal values.

ACKNOWLEDGEMENTS
The work described in this paper was part-funded by the National Measurement System Directorate of the UK government’s Department of Trade and Industry. © Crown Copyright 2007. Reproduced by permission of the Controller of HMSO.

REFERENCES