Clarke, R.G., Li, C. and Ridler, N.M. (2018) An Intra-laboratory Investigation of On-wafer Measurement Reproducibility at Millimeter-wave Frequencies. In: 2017 90th ARFTC Microwave Measurement Symposium (ARFTG), Boulder, CO, USA, 28 Nov - 01 Dec 2017, ISBN 9781538643563.

An Intra-laboratory Investigation of On-wafer Measurement Reproducibility at Millimeter-wave Frequencies

R G Clarke[1], C Li[2] and N M Ridler[3]

[1]University of Leeds, UK, r.g.clarke@leeds.ac.uk; [2]University of Glasgow, UK, chong.li@glasgow.ac.uk;
[3]National Physical Laboratory, UK, nick.ridler@npl.co.uk

***Abstract*—Understanding the relative contribution of contact repeatability and overall reproducibility for on-wafer measurements provides useful insight into the significance of measurement comparisons. We report on an intra-laboratory investigation into contact repeatability and the variation that may be anticipated when measurements are reproduced in different laboratories using different equipment. We pay particular attention to the dispersion in measurement results arising from the use of on-wafer and off-wafer calibration. Experimental results are reported for measurements in the frequency range 140 GHz to 220 GHz, together with preliminary estimates of the repeatability limits for this type of measurement.**

***Index Terms*—Measurement repeatability, Measurement reproducibility, On-wafer measurements, Millimeter-wave measurements, Measurement uncertainty**

I. INTRODUCTION

Advances in applications of THz and sub-THz frequencies have led to an increased demand for semiconductor devices and integrated circuits operating at millimeter-wave and submillimeter-wave frequencies. Efficient development calls for an 'on-wafer' measurement capability in order to determine electrical characteristics without dicing the semiconductor wafer and packaging the devices. Component manufacturers and system designers must understand the uncertainty associated with the measurement data and the expected reproducibility when the same device is measured in different laboratories.

Conventional on-wafer measurements are achieved by physical contact of micro-scale probe-tips onto metalized contact 'pads' on a wafer substrate, often connecting the device-under-test (DUT) via co-planar waveguide (CPW) transmission lines. Since both the probe-tips and the contact pads are fragile structures, the measurement is necessarily invasive to the extent that some physical damage is expected with each probe contact. For this reason, acceptable measurement repeatability is usually limited to a relatively small number of contacts (often < 10). The fragility of the probes means that small deformities in the probe tips can also have a noticeable impact on the measurement results due to differing parasitic elements produced by the tip geometry and by differing substrate modes excited by the probe-pad interface [1].

The quantities of interest are usually the complex reflection and transmission coefficients (i.e. S-parameters), measured with a Vector Network Analyzer (VNA). A calibration procedure is therefore necessary to enable error-corrected measurements. In order to acquire the correct reference impedance for the measurement, it is often preferable to realize a set of calibration standards on the same substrate as the DUT ('on-wafer standards') using transmission line structures identical to those used to interconnect the DUT and probe contact pads. On-wafer standards are particularly appropriate when applying one of the so-called 'self-calibration' techniques [2]-[4], in which acquisition of the impedance properties of the CPW lines realized on the DUT substrate is integral to the calibration performance. These calibration techniques are generally derived from the through-reflect-line (TRL) technique [5].

However, implicit in the use of a commercial calibration impedance standard substrate (ISS) is the assumption that measurements will be made of a DUT on a different substrate. It is assumed that the calibration procedure will establish a reference plane and appropriate reference impedance at the probe-tips and that this will enable meaningful measurements to be made directly on the target DUT substrate.

The interaction of specific probe-tip designs and the ISS contact pads will produce parasitic elements that may not be the same as those on the DUT wafer [6]. In some cases, a second-tier calibration or de-embedding procedure (for example, [7]) may be necessary to separate the DUT measurement from the residual errors that arise from using a different substrate for the calibration standards, but this is not necessarily routine practice. If we assume that the calibration substrate may, or may not, be different from the measurement (DUT) substrate, then we can investigate the consequences for the reproducibility of measurements. We can also investigate, and attempt to quantify, the relative contributions from contact repeatability and the effect of varying other physical aspects of the measurement, such as the use of different probes.

In this work, we experimentally investigate the contact repeatability and reproducibility of on-wafer measurements at frequencies from 140 GHz to 220 GHz. We conduct experiments using both on-wafer and off-wafer calibrations, physically different probes and a number of repeat contacts, in order to ascertain the relative contributions to the overall reproducibility. We use the terms 'repeatability' and 'reproducibility' based on definitions given in [8], and we follow methods given in [9] for their determination.

II. METHOD

A. Experimental Set-up

We limit our investigation to the repeatability and reproducibility obtained using manually-operated probe manipulators since this is typically the type of probe station found in laboratories engaged in millimeter-wave on-wafer metrology.

The measurement system used in the experiment consists of a Keysight Technologies PNA-X with Virginia Diodes Inc. millimeter-wave extender heads, operating from 140 GHz to 220 GHz. To reduce the measurement noise, an IF bandwidth of 100 Hz was used, although no further trace averaging was applied. We measured data at 801 equally-spaced points across this frequency range.

The VNA system is connected to a manual probe station. For the experiments, 75 μm pitch ground-signal-ground (GSG) probes were used (GGB Industries 'Picoprobe' Model 220-BT-M). The probes had both been used in prior experiments but were confirmed to be in good condition before and after the measurements obtained in this work. We note that for operation at 220 GHz, the use of 75 μm probes is at the upper limit of recommended use [10], as the probe pitch is becoming an appreciable fraction of the signal wavelength.

A single-tier, two-port SOLT calibration was performed twice, using different commercially available ISS devices. This enabled the error-correction acquired from each of the two calibrations to be applied separately to each DUT measurement. We elected to use SOLT because, although it is generally considered inferior for on-wafer calibration (compared with LRM, TRL, etc), the approach is readily implemented on any VNA. We recognize that the SOLT calibration method relies primarily on defined impedance standards rather than the propagation characteristics of CPW lines and it is therefore sensitive to probe placement errors and probe-pad parasitics.

The two ISS devices used were (a) GGB Industries CS-15 [11] and (b) Cascade Microtech 138-356 [12]. Each ISS provides open, short, load and thru standards. Substrate open-circuit standards were used for both calibrations. For both the calibration and measurements, a dielectric spacer (Cascade absorber, 116-344) was used between each alumina ISS and the metallic probe-station chuck.

For each SOLT calibration, a 'defined thru' was used, according to the manufacturer's specifications for the electrical delay of the thru standard, adjusted for the probe over-travel used in the experiment. For the open-circuit, short-circuit and load standards, we applied the nominal values of capacitive and inductive corrections typically recommended for use with 75 μm probes.

For the DUTs, we select two standards from each ISS, so that 'on-wafer' and 'off-wafer' calibrations are observed for each ISS. Some of the DUTs were nominally the same as the devices used in the calibrations (i.e. open, short and load), although they were physically different examples to those used in the calibration step. The DUTs were an open-circuit, a short-circuit, a matched load and a 900 μm length of CPW transmission line. Table I lists the DUTs used in the experiment:

TABLE I
SELECTED DEVICES FOR USE AS DUT CANDIDATES

ISS	One-Port DUTs:	Two-Port DUT:
Cascade	Open	900 μm CPW Line
GGB	Short	
GGB	Matched Load	

It is well known that, for SOLT calibration, a measurement of a device with similar electrical characteristics to one of the calibration standards will generally show good 'agreement'. In this investigation, we are not concerned with a critical analysis of the calibration accuracy; instead, we are seeking to explore the typical dispersion of measurement results obtained through realistic calibration scenarios, and compare this to the achievable contact repeatability. Nevertheless, it is helpful to consider the expected behavior of the two-port DUT which was not similar to the devices used in the calibration and may therefore serve as a verification device. For the CPW line, the nominal transmission phase, φ, can be calculated as:

$$\varphi = \frac{-360 \times l}{\lambda} \text{ (degrees)}, \tag{1}$$

where λ is the wavelength determined using the appropriate velocity factor, and l is the effective length of the line after correcting for probe over-travel. For the Cascade ISS, a velocity factor of 0.432 may be inferred from the manufacturer's data. We adjust the line length to 875 μm to account for probe over-travel and compute the expected transmission phase using (1) for three frequency points. These are shown in Table II.

TABLE II
NOMINAL TRANSMISSION PHASE FOR THE CPW LINE

Frequency (GHz)	Wavelength, λ (μm)	Phase, φ
140	925	20°
180	720	78°
220	589	– 175°

The above values assume that the CPW lines are 'well-behaved' and there are no unwanted substrate modes. We have also ignored the 'stub' effect due to the probe position being slightly offset from the true ends of the line.

The manufacturer does not specify a value for the expected loss in the CPW verification lines, but an estimate obtained from a transmission-line simulator (based on quasi-static models for CPW line structures) would suggest that values between ≈ 0.3 dB/mm (at 140 GHz) to ≈ 0.4 dB/mm (at 220 GHz) should be expected. For the 900 μm line, this equates to a linear transmission coefficient of ≈ 0.97 at 140 GHz and ≈ 0.95 at 220 GHz.

B. Measurement sequence

For the one-port devices, a single measurement of the raw (i.e. uncorrected) measurement data was acquired. Each SOLT calibration was then applied in turn to the raw measurement data in order to obtain two sets of corrected measurement data. The probes were then lifted and the DUT was physically rotated, such that the Port Two probe was now aligned with the device previously measured at Port One. For a one-port DUT with relatively low probe-to-probe transmission and cross-talk, the SOLT algorithm effectively reverts to two quasi-independent one-port SOL calibrations. Thus, by measuring the same device at Port Two, we are conducting measurements of the same DUT with a physically different probe and, essentially, an independent set of one-port error-correction coefficients.

We applied the two SOLT calibrations to this second measurement before repeating the sequence by physically rotating the DUT back to its original position and repeating the first measurement. Four 'cycles' of this process were conducted to provide contact-repeatability data interspersed with measurements using different probes. By performing the alternate orientations we avoid the possibility that a progressive change in the electrical characteristics of the DUT (due to repeat contacts) may mask the effect of using two different probes. It also ensures that the contact repeatability is conducted under the more rigorous requirement of manually realigning the probes to the DUT for each new contact. Effectively, this combines contact-repeatability with positioning-repeatability, which provides for a more realistic scenario for overall repeatability.

In this way, each one-port device was subject to eight probe contacts (four from the probe at Port 1, four from the probe at Port 2). Thus, we have four groups of measurements (two calibrations, two physical probes) and for each different probe we have four repeat contacts. These measurements now provide information about the effect of:

- Repeat probe contacts;
- Using physically different probes;
- Using different calibrations ('on-wafer', 'off-wafer').

For the two-port device, the same procedure was followed except that we did not physically rotate the device and a total of four repeat contacts were made, providing information about the effect of:

- Repeat probe contacts;
- Using different calibrations ('on-wafer', 'off-wafer').

III. RESULTS AND DISCUSSION

A. One-port DUTs

In order to compare the relative contributions of contact repeatability, the use of physically different probes and the use of different calibration substrates, we first present the complete measurement data for each DUT. Figures 1 to 3 show the measured reflection coefficient (linear magnitude) for all three one-port DUTs and Figures 4 and 5 show the measured reflection phase for the open-circuit and short-circuit DUTs. Four repeat contacts are shown.

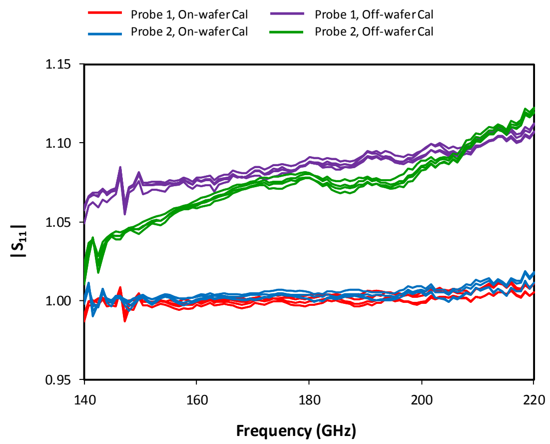

Fig. 1. Open-Circuit: Reflection Coefficient Magnitude.

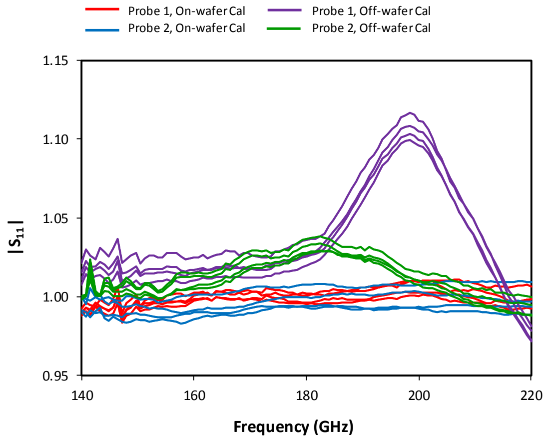

Fig. 2. Short-Circuit: Reflection Coefficient Magnitude.

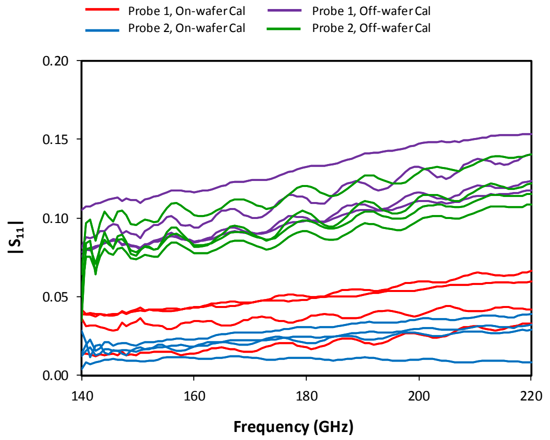

Fig. 3. Matched Load: Reflection Coefficient Magnitude.

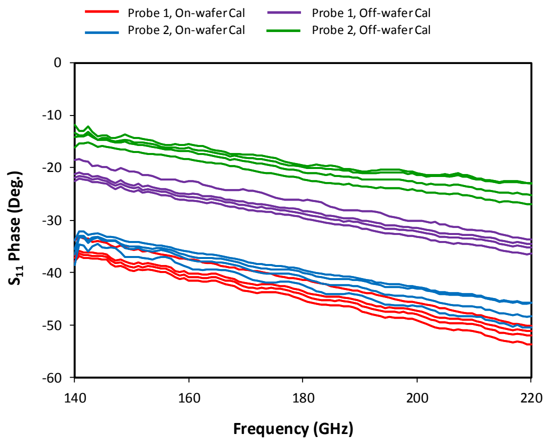

Fig. 4. Open-Circuit: Reflection Coefficient Phase.

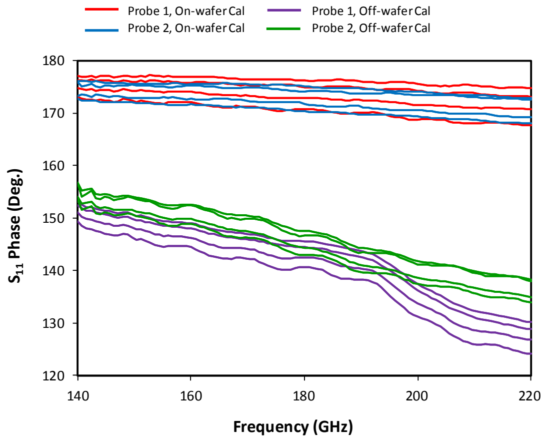

Fig. 5. Short-Circuit: Reflection Coefficient Phase.

It is evident from all the measurement results that the use of physically different probes produced much smaller variations than the use of different calibration substrates. For all three one-port DUTs, the ‘on-wafer’ calibration shows approximate agreement with the nominal reflection magnitudes for these devices. The results for ‘off-wafer’ calibration are further from the nominal magnitude values. This is true regardless of which ISS takes the role of the ‘on-wafer’ calibration. We suspect that this is a consequence of imperfect knowledge of the probe-pad parasitics used in the calibration and the extent to which the DUT characteristics may be considered to be independent of the probe-DUT interaction.

We expect the measured phase values to depart from ‘ideal’ values and be indicative of the parasitic capacitance and inductance present in these structures. The reflection phases also show considerable differences between the on-wafer and off-wafer calibrations.

To quantify the contribution from contact repeatability we begin by summarizing the maximum standard deviations, across all frequencies in the measurement band, for the repeat contacts within each group. These are shown in Table III (reflection magnitude for all three DUTs) and Table IV (reflection phase for the open-circuit and short-circuit DUTs).

TABLE III
MAXIMUM STANDARD DEVIATIONS IN REFLECTION MAGNITUDE FOR CONTACT-REPEATABILITY

	Maximum Standard Deviation (Linear Magnitude)			
DUT	Probe 1, On-wafer Calibration	Probe 2, On-wafer Calibration	Probe 1, Off-wafer Calibration	Probe 2, Off-wafer Calibration
Open	0.004	0.006	0.005	0.009
Short	0.006	0.009	0.009	0.007
Load	0.016	0.013	0.018	0.014

TABLE IV
MAXIMUM STANDARD DEVIATIONS IN REFLECTION PHASE FOR CONTACT-REPEATABILITY

	Maximum Standard Deviation (Phase)			
DUT	Probe 1, On-wafer Calibration	Probe 2, On-wafer Calibration	Probe 1, Off-wafer Calibration	Probe 2, Off-wafer Calibration
Open	1.9°	2.3°	1.8°	1.9°
Short	3.3°	2.5°	2.9°	2.2°

To a reasonable approximation, the repeat contacts within each group produce similar standard deviations. The standard deviations for the low-reflecting DUT are slightly higher than those obtained for the high-reflecting DUTs.

B. Two-port DUT

The two-port DUT provides a more meaningful verification of the performance of each calibration since a transmission line did not form one of the calibration standards. Figures 6 and 7 show the measured reflection and transmission coefficient for the two-port DUT (CPW line). Figure 8 shows the measured transmission phase.

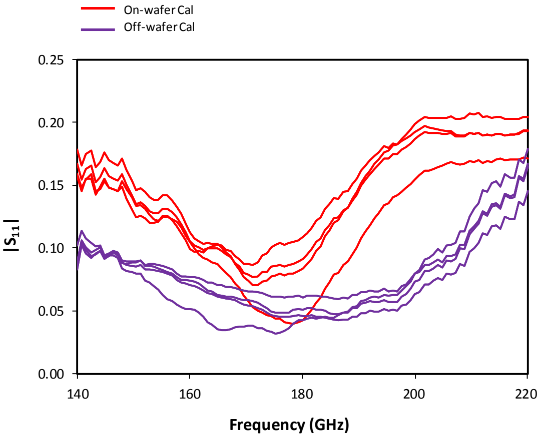

Fig. 6. 900 µm CPW Line: Reflection Coefficient Magnitude.

The measured reflection coefficient indicates some residual mismatch between the reference impedance established by the calibrations and the line impedance for both on-wafer and off-wafer calibrations.

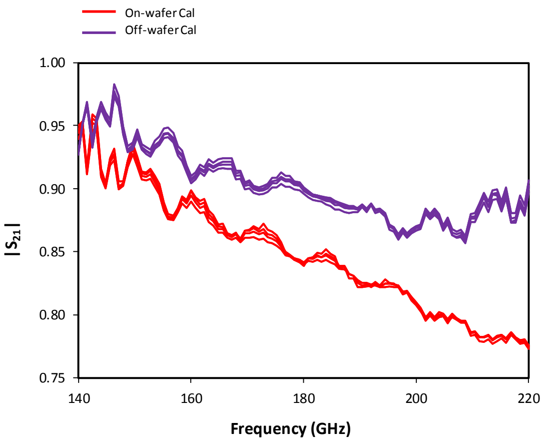

Fig. 7. 900 µm CPW Line: Transmission Coefficient Magnitude.

The measured transmission coefficient suggests greater loss than predicted (for example, the predicted loss was 0.95 at 220 GHz) and it is also different for the two calibrations.

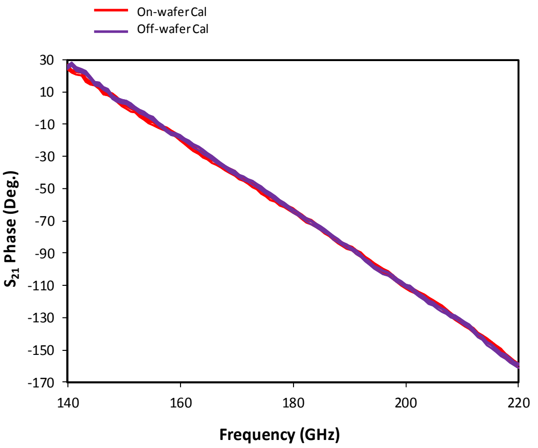

Fig. 8. 900 µm CPW Line: Transmission Phase.

The measured transmission phase deviates from the expected value (as shown in Table II) by up to 15°. There was no clear systematic difference in the transmission phase result when the on-wafer and off-wafer calibrations are compared.

As with the preceding one-port DUTs, we summarize the effect of contact repeatability in terms of the maximum standard deviations, for each calibration. Tables V and VI show this for the linear reflection and transmission magnitude and the transmission phase.

TABLE V
MAXIMUM STANDARD DEVIATIONS IN REFLECTION AND TRANSMISSION MAGNITUDE FOR CONTACT-REPEATABILITY

DUT	Maximum Standard Deviation (Linear Magnitude)			
	S_{11} On-wafer Calibration	S_{11} Off-wafer Calibration	S_{21} On-wafer Calibration	S_{21} Off-wafer Calibration
CPW Line	0.028	0.016	0.005	0.005

TABLE VI
MAXIMUM STANDARD DEVIATIONS IN TRANSMISSION PHASE FOR CONTACT-REPEATABILITY

DUT	Maximum Standard Deviation (Phase)	
	S_{21} On-wafer Calibration	S_{21} Off-wafer Calibration
CPW Line	1.2°	1.2°

IV. ANALYSIS

A. *One-port DUTs*

To quantify the effect of using the on-wafer and off-wafer calibrations and from using physically different probes, we consider the means from each measurement group to represent samples from a population of measurements with various systematic errors. We may then compare the standard deviation of these measurement-group means with the standard deviations obtained from the repeat contacts, in order to compare the likely contributions to overall reproducibility.

To avoid exaggerating the contribution from these ‘randomized’ systematic errors, we compare the average standard deviation (from across the measurement frequency band) to the worst-case standard deviation for contact repeatability. This is shown for both the reflection magnitude and phase standard deviations (SD) in Tables VII and VIII.

TABLE VII
AVERAGE STANDARD DEVIATIONS IN THE MEAN LINEAR MAGNITUDE FROM EACH GROUP COMPARED WITH STANDARD DEVIATIONS OBTAINED FROM CONTACT-REPEATABILITY

	Group Means Average SD $\lvert S_{11}\rvert$	Contact Repeatability Worst-Case SD $\lvert S_{11}\rvert$
Open	0.045	0.009
Short	0.020	0.009
Load	0.044	0.018

TABLE VIII
AVERAGE STANDARD DEVIATIONS IN THE MEAN REFLECTION PHASE FROM EACH GROUP COMPARED WITH STANDARD DEVIATIONS OBTAINED FROM CONTACT-REPEATABILITY

	Group Means Average SD S_{11} (phase)	Contact Repeatability Worst-Case SD S_{11} (phase)
Open	11.1°	2.3°
Short	17.4°	3.3°

Tables VII and VIII suggest that the contribution to measurement variability from contact repeatability is typically much smaller than that from various systematic effects such as the use of on-wafer/off-wafer calibrations.

B. *Two-port DUT*

For the two-port DUT, we have only two groups of means, (obtained from the four repeat measurements for each of the two calibrations). It is therefore preferable to compare the average difference in these means with the standard deviations obtained from the repeat contacts. Table IX shows this for both the reflection and transmission magnitudes.

TABLE IX
AVERAGE DIFFERENCES IN THE MEAN LINEAR MAGNITUDE FROM EACH CALIBRATION COMPARED WITH STANDARD DEVIATIONS OBTAINED FROM CONTACT-REPEATABILITY

	Differences in Mean & Maximum Standard Deviations			
DUT	Group Means Average Difference $\lvert S_{11}\rvert$	Contact Repeatability Worst-case SD $\lvert S_{11}\rvert$	Group Means Average Difference $\lvert S_{21}\rvert$	Contact Repeatability Worst-case SD $\lvert S_{21}\rvert$
CPW Line	0.060	0.028	0.055	0.005

For the reflection magnitude, the average difference in the means is slightly larger than the worst-case standard deviation in contact repeatability. For the transmission magnitude, the average difference is greater by a factor of ten. Again, this suggests that the contact repeatability does not contribute as much to measurement variability as the systematic effects.

Finally, we compare the average difference in the means for the transmission phase with the standard deviations in transmission phase from contact repeatability (Table X).

TABLE X
AVERAGE DIFFERENCES IN THE MEAN TRANSMISSION PHASE FROM EACH CALIBRATION COMPARED WITH STANDARD DEVIATIONS OBTAINED FROM CONTACT-REPEATABILITY

	Group Means Average Difference S_{21} (phase)	Contact Repeatability Worst-Case SD S_{21} (phase)
CPW Line	1.1°	1.2°

The results for the transmission phase are interesting because, although there are clearly some differences between the measured and expected values (Figure 8), there are minimal difference between the two calibrations. Further, the small differences that were observed were of a similar size to the standard deviation obtained from contact repeatability. This suggests that the primary error term responsible for transmission phase correction (the transmission-tracking error vector) was determined more consistently by the two SOLT calibrations than the other error terms.

5. CONCLUSIONS

In summary, comparisons have been made between repeat-measurements of some on-wafer devices and measurements of the same devices when the calibration and probe conditions are varied, as might be expected when different measurement laboratories are involved.

We conclude that the impact of using different calibration substrates (as might be expected in different laboratories) is typically larger than the effect of contact repeatability. This is evidently the case for the SOLT calibrations used in our experiments. Where more advanced calibration algorithms might be used, possibly leading to smaller differences in the results, we have quantified the contact repeatability with which such measurements may be compared. Additionally, for the number of repeat contacts made in our experiment (typically eight), there was no obvious trend in the results due to changes in DUT characteristics, despite the invasive nature of the measurement. We surmise that most differences observed within an inter-laboratory comparison would be difficult to attribute to the effect of contact repeatability.

An inter-laboratory comparison is currently being planned for this frequency range under the EMPIR project 14IND02, “Microwave measurements for planar circuits and components”. It is expected that the results will add further experimental insight into the reproducibility limitations for on-wafer measurements at these frequencies.

ACKNOWLEDGEMENT

The authors acknowledge support by the European Metrology Programme for Innovation and Research (EMPIR) project 14IND02 “Microwave measurements for planar circuits and components”. The EMPIR program is co-financed by the participating countries and the European Union’s Horizon 2020 research and innovation program.

REFERENCES

[1] S. A. Wartenberg, “Selected topics in RF coplanar probing,” *IEEE Trans. Microwave Theory & Tech*., vo. 51, no.4, pp.1413–1421, Apr. 2003.

[2] A. Davidson, E. Strid, and K. Jones, “Achieving greater on-wafer S-parameter accuracy with the LRM calibration technique”, *34th ARFTG Conf. Dig*., pp. 61-66, Dec. 1989.

[3] A. Davidson, K. Jones, and E. Strid, “LRM and LRRM calibrations with automatic determination of load inductance”, *36th ARFTG Conf. Dig*., pp. 5763, Dec. 1990.

[4] R. B. Marks, “A multiline method of network analyzer calibration,” *IEEE Trans. Microwave Theory & Tech*., vol. MIT-39, no. 7, pp. 1205-1215. Jul. 1991.

[5] G. F. Engen, and C. A. Hoer, “Thru-Reflect-Line: An improved technique for calibrating the dual six-port automatic network analyser”, *IEEE Trans. Microwave Theory & Tech*. vol. 27, no. 12, pp. 987-993, Dec. 1979.

[6] U. Arz, and D. Schubert, “Coplanar microwave probe characterization: caveats and pitfalls,” *67th ARFTG Conf. Dig*., pp. 214-218, Jun. 2006.

[7] H. Cho, and D. Burk, “A three step method for the de-embedding of high frequency S-parameter measurements,” *IEEE Trans. Electron Devices*, vol. 38, no. 6, pp. 1371–1375, 1991.

[8] ISO 3534-1:2006, Statistics – Vocabulary and symbols – Part 1: General statistical terms and terms used in probability.

[9] ISO 5725-2:1994, Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.

[10] “Upper Frequency Recommendations for Various RF Probes and Pitches”, available online from Cascade Microtech Inc., Beaverton, OR, (www.cascademicrotech.com).

[11] “Calibration substrate CS-15”, GGB Industries Inc., Naples, FL

[12] “Impedance Standard Substrate Map: 138-356”, Cascade Microtech Inc., Beaverton, OR.