Consiglio Nazionale delle Ricerche

VNA calibration suitable for free-space one-port measurements of large targets

A. Simonetto

FP20/06 July 2020

ISTITUTO PER LA SCIENZA E TECNOLOGIA DEI PLASMI

Via R. Cozzi 53 - 20125 Milano (Italy)

Disclaimer

This report only represents the opinion of the author(s) at the time of writing.

Introduction

This report is an evolution of [1] (which implemented the calibration technique described in [2], [3]). A different technique is described here, best suited for a large or heavy Device Under Test (DUT) that cannot be moved easily. It implements the *variable short, variable load, fixed short* calibration as described in [4], and the calibration coefficients are used to remove systematic errors from individual DUT measurements. The process also computes the phase of the reflection coefficient, which is however not usually interesting for this type of measurements.

The main advantage of this method is that calibration can be made using calibration standards made with easily movable low-weight targets, using only one (frequency) sweep on the (static) DUT.

The disadvantages are the lower signal-to-noise ratio, since only one sweep is made on the DUT, and the sensitivity (i.e. maximum measureable return loss), which is better than the return loss of the *variable load* standard by one or two orders of magnitude only.

Nonetheless, this calibration technique is suitable for large heavy DUTs, normally with moderate return loss, which allows using lightweight commercial absorbers as calibration loads.

Theory

When a one-port device with reflection coefficient ρ is measured with a VNA through a two-port error matrix $\underline{S}$, i.e. a virtual scattering matrix including all the non-ideal instrument behaviour, the measured quantity ρ_M is:

$$\rho_M = S_{11} + \frac{S_{12}S_{21}}{1 - S_{22}\rho}\rho = \frac{\left(S_{12}S_{21} - S_{11}S_{22}\right)\rho + S_{11}}{1 - S_{22}\rho} \equiv \frac{a\rho + b}{1 + c\rho} \tag{1}$$

where S_{nm} are the elements of the matrix and the device is connected to port 2.

The corrected reflection coefficient of the DUT can be computed (assuming that no singularities exist) if the calibration constants a, b, c are known:

$$\rho = \frac{1}{a}\frac{\rho_M - b}{1 - \frac{c}{a}\rho_M} \tag{2}$$

Assuming that the phase of the reflection coefficient ρ of a calibration device is varied at constant amplitude by shifting it along the direction of propagation z, the measured value ρ_M describes a circle in complex space, with center X and radius R given by (see e.g. [4]):

$$X = \frac{b - ac^*\left|\rho\right|^2}{1 - \left|c\right|^2\left|\rho\right|^2} = \frac{S_{12}S_{21}S_{22}^*\left|\rho\right|^2}{1 - \left|S_{22}\right|^2\left|\rho\right|^2} + S_{11} \tag{3}$$

$$R = \frac{\left|a - bc\right|\left|\rho\right|}{1 - \left|c\right|^2\left|\rho\right|^2} = \frac{\left|S_{12}S_{21}\right|\left|\rho\right|}{1 - \left|S_{22}\right|^2\left|\rho\right|^2} \tag{4}$$

Let X_S, R_S be the center and radius of the circle described measuring the *variable short*, and X_L, R_L those of the *variable load*.
Proceeding as outlined in [4], one can solve (3) for $|\rho|^2$, substitute only in the denominator of (4) and solve for $|\rho|$, then square and substitute back in (3), thus obtaining the pair of (complex) equations (one each for the *variable load* and *short*)

$$\left(b^* - X_S^*\right)\left(1 - R_S\frac{c}{a}\right) - R_S^2\frac{c}{a} = 0 \tag{5}$$

$$\left(b^* - X_L^*\right)\left(1 - R_L\frac{c}{a}\right) - R_L^2\frac{c}{a} = 0 \tag{6}$$

The system of equations can be solved for b and c/a (fixing the typo in [4])*:*

$$b = X_L - \frac{2r^2\left(X_S - X_L\right)}{H \pm \sqrt{H^2 - 4\left|X_S - X_L\right|^2 R_L^2}} \cong X_L \tag{7}$$

$$\frac{c}{a} = \frac{b^* - X_S^*}{R_S^2 - \left|X_S\right|^2 + b^* X_S} \tag{8}$$

with the definition

$$H = R_S^2 - R_L^2 - \left|X_S - X_L\right|^2 \tag{9}$$

The plus sign in (7) is the one to choose, corresponding to the smaller magnitude of b. For loads with moderate return loss, the right hand term in (7) can be approximated as shown. The full form is considered here.
A single measurement on *fixed short* (ρ=-1) allows using (2) to determine a (and c as a consequence):

$$a = \frac{b - \rho_{FS}}{1 - \frac{c}{a}\rho_{FS}} \tag{10}$$

where ρ_{FS} is the signal measured with the fixed short.
Since the *variable short* is already as good as possible, the signal with the *fixed short* standard ρ_{FS} is simply the measurement with the *variable short* in the location chosen as reference plane.
In order to reduce sensitivity to noise, the data processing program, rather than using a single measurement, computes the *fixed short* values using the estimated X_S, R_S whereas the phase at the reference position is obtained from a linear fit of

$$Arg\left[\rho_M\left(z_i\right) - X_S\right] = m \times \left(z_i - z_0\right) + \varphi\left(z_0\right) \tag{11}$$

where z_i are the discrete positions along the propagation axis. In actual facts the program does not fit over z, but over the index i. The intercept $\varphi(z_0)$ is the same.
The phase in (11) is of course unwrapped before use.

Error bars

Using standard error propagation formulae in (2) but *neglecting covariances* gives

$$\sigma_\rho^2 = \left|\partial_a\rho\right|^2\sigma_a^2 + \left|\partial_{\rho_M}\rho\right|^2\sigma_{\rho_M}^2 + \left|\partial_b\rho\right|^2\sigma_b^2 + \left|\partial_{\frac{c}{a}}\rho\right|^2\sigma_{\frac{c}{a}}^2 \tag{12}$$

that is

$$\frac{\sigma_\rho^2}{|\rho|^2} = \frac{\sigma_a^2}{|a|^2} + \frac{\sigma_b^2}{|\rho_M - b|^2} + \frac{|\rho_M|^2}{\left|1 - \frac{c}{a}\rho_M\right|^2}\sigma_{\frac{c}{a}}^2 \tag{13}$$

Applying error propagation to (9) one can write

$$\sigma_H^2 = 4R_S^2\sigma_{R_S}^2 + 4R_L^2\sigma_{R_L}^2 + 4|X_S - X_L|^2\,\sigma_{|X_S - X_L|}^2 \tag{14}$$

and using

$$\sigma_{|X_S - X_L|}^2 = \sigma_{X_S - X_L}^2 = \sigma_{X_S}^2 + \sigma_{X_L}^2 \cong \frac{1}{N_S}\sigma_{R_S}^2 + \frac{1}{N_L}\sigma_{R_L}^2 \tag{15}$$

where N_L, N_S are the number of positions for the *variable load* and *variable short*, one can rewrite into

$$\sigma_H^2 = 4\left(R_S^2 + \frac{1}{N_S}|X_S - X_L|^2\right)\sigma_{R_S}^2 + 4\left(R_L^2 + \frac{1}{N_L}|X_S - X_L|^2\right)\sigma_{R_L}^2 \tag{16}$$

The approximation on the right hand side of (15) derives from the observation that, when phase increments are constant, the center of a circle in complex plane is the same as the average of all points, as in (26).

Hence one can write after lengthy calculations

$$\begin{aligned}\sigma_b^2 = \frac{\sigma_{R_L}^2}{N_L} + \frac{|(b - X_L)|^2}{\left|H^2 - 4|X_S - X_L|^2 R_L^2\right|} \times \left\{\left|H + \sqrt{H^2 - 4|X_S - X_L|^2 R_L^2}\right|^2 \frac{\sigma_{R_L}^2}{R_L^2} + \sigma_H^2 + \right. \\ \left. + \left|2|X_S - X_L| + \sqrt{H^2 - 4|X_S - X_L|^2 R_L^2}\right|^2 \frac{1}{|X_S - X_L|^2}\left(\frac{\sigma_{R_S}^2}{N_S} + \frac{\sigma_{R_L}^2}{N_L}\right)\right\}\end{aligned} \tag{17}$$

To compute the variance on *c/a* one needs to transform (8) into

$$\frac{c}{a} = \frac{b^* - X_S^*}{R_S^2 - |X_S|^2 + b^* X_S} = \frac{b^* - X_S^*}{R_S^2 + \left(b^* - X_S^*\right)X_S}$$

and then

$$\begin{aligned}\sigma_{\frac{c}{a}}^2 &= \left|\partial_{(b - X_S)^*}\frac{c}{a}\right|^2 \sigma_{(b - X_S)^*}^2 + \left(\partial_{R_S}\frac{c}{a}\right)^2 \sigma_{R_S}^2 + \left|\partial_{X_S}\frac{c}{a}\right|^2 \sigma_{X_S}^2 = \\ &= \frac{1}{\left|R_S^2 + \left(b^* - X_S^*\right)X_S\right|^4}\left\{R_S^4\sigma_b^2 + \left[\left|b^* - X_S^*\right|^2\left(4R_S^2 + \frac{1}{N_S}\left|b^* - X_S^*\right|^2\right) + \frac{R_S^4}{N_s}\right]\sigma_R^2\right\}\end{aligned} \tag{18}$$

Similarly from (10) one can compute

$$\sigma_b^2 = |a|^2\left\{\left|\frac{1}{b - \rho_{FS}}\right|^2 \sigma_b^2 + \left|\frac{\rho_{FS}}{1 - \frac{c}{a}\rho_{FS}}\right|^2 \sigma_{\frac{c}{a}}^2 + \left|\frac{b\frac{c}{a} - 1}{\left(b - \rho_{FS}\right)\left(1 - \frac{c}{a}\rho_{FS}\right)}\right|^2 \sigma_{R_S}^2\right\} \tag{19}$$

Software

A small FORTRAN program was written to process the data, actually a major new version of the one described in [1].

Input

The program reads five files, `short.txt`, `load.txt`, `dut.txt`, `parms.txt` and `mask.txt`.
The first two files consist each of an integer number representing the VNA gain, followed by a list of VNA data file names for the *variable short* (`short.txt`) and *variable load* (`load.txt`) at different positions along the direction of the incident beam (lists must be ordered in z coordinate, either ascending or descending).
The third file contains the same information for the data files of the device under test (`dut.txt`). Of course the list of DUT data files has no predefined ordering.
All listed data files must be in the same directory as the program.
The fourth file, `parms.txt`, contains on four separate lines the initial, final z coordinates, the step and the coordinate of the reference position.
Positions should cover approximately a half-wavelength at the minimum frequency of interest. The range of spatial scan should not be excessive, to avoid modifications to the geometry.
The last file, `mask.txt`, contains an arbitrary number of lines, each consisting of a frequency followed by a list of integers, each one representing a position in the list of *variable load* files. Data in the corresponding files is ignored when fitting *variable load* data for the specified frequency. This control was added to cope with occasional garbage data points (possibly due to EMC issues) affecting data in the frequency bands where the low power of the VNA local oscillators must be compensated with external amplification.

Circle fitting algorithm

Circles described in complex space by measured *variable short* data are fitted using the same non-iterative algorithm [5] used in [1], which is less sensitive to random errors than the one used in [4] when many equally spaced measurements are made on a calibration standard. The algorithm described in [4] is best suited when a small number of measurements (of course $\geq$3) is available. A different approach is followed with *variable load* data, though.
Complex numbers are mapped on pairs of real numbers

$$X = \left[\mathrm{Re}(X), \mathrm{Im}(X)\right] \equiv \left[x, y\right] \tag{20}$$

and a standard least-squares fit is applied to

$$f\left(x_0, y_0, R_{sq}\right) = \sum_{i=1}^{N}\left[\left(x_i - x_0\right)^2 + \left(y_i - y_0\right)^2 - R_{sq}\right]^2 \tag{21}$$

where $R_{sq}=R^2$.
As a result, the radius R of the best fitting circle is

$$R = \sqrt{\frac{1}{N}\sum\left(x_i - x_0\right)^2 + \left(y_i - y_0\right)^2} \tag{22}$$

with (x_0,y_0) minimizing the sum of deviations. The center (x_0,y_0) is the solution of equation

$$\underline{\underline{G}}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \underline{C} \tag{23}$$

where the matrix G and the vector C are given by

$$\underline{G} = 8\begin{bmatrix} \sum_{i=1}^{N} x_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)^2 & \sum_{i=1}^{N} x_i y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i & \sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2 \end{bmatrix} \tag{24}$$

$$C = 4\begin{bmatrix} -\frac{1}{N}\sum_{i=1}^{N} x_i^2 \sum_{j=1}^{N} x_j - \frac{1}{N}\sum_{j=1}^{N} y_j^2 \sum_{k=1}^{N} x_k + \sum_{j=1}^{N} y_i^2 x_i + \sum_{i=1}^{N} x_i^3 \\ -\frac{1}{N}\sum_{i=1}^{N} y_i^2 \sum_{j=1}^{N} y_j - \frac{1}{N}\sum_{j=1}^{N} x_j^2 \sum_{k=1}^{N} y_k + \sum_{j=1}^{N} x_i^2 y_i + \sum_{i=1}^{N} y_i^3 \end{bmatrix} \tag{25}$$

This algorithm is very efficient and insensitive to low-level noise, but only working for data that approximately lie on a circle, giving unexpected results otherwise. For example, if all data lie on a circle except a very far one, the best fitting circle passes exactly through the external datum and approximately through the center of the circle formed by the other data, which is indeed the best-fitting circle, but not the intended one. Anyway, this occurrence is never observed in *variable short* measurements.

The distribution of data points has a significant influence on the determination of the center, and a full half-wavelength coverage of the circle at low frequency is bound to result in extra coverage at high frequency. To prevent the estimate of the center being biased by the presence of the extra arc, the number of data required to cover uniformly a half wavelength is computed, and fits are made considering only that number of data. A running average of fit results is made considering all available data. Equivalent results could be obtained assigning weights to the data.

The dependence of the center estimate on the distribution of data was well known and already present in [1], but had no effect on results, since only the circle radii were used there.

Variable load data are intrinsically more critical than *variable short* ones because of the less favourable balance between wanted signal and stray reflections. Moreover, the return loss of the *variable load* can be moderate or high, depending on the required measurement precision. Therefore, *variable load* data are fitted both using the algorithm above and the one described below. The fit giving a lower fractional error is chosen at each frequency.

The alternate fitting algorithm is really simple: if data are uniformly distributed on a circle, which is assumed with this calibration technique, its center is the barycenter of the data

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \frac{1}{N}\sum_{i=1}^{N} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{26}$$

and its radius is the standard deviation of the data.

To prevent any garbage data to affect the output, this fit is made twice, discarding in the second step all the data not satisfying Chauvenet's criterion (see e.g. [6]).

Output

The program writes frequency-dependent calibrated data of DUT in tab-separated files named `DUT1.txt,..., DUT<N>.txt`, where <N> is of course the number of files listed in `dut.txt`.

Each file contains a two-line header and records for all frequencies.

The first header line lists the number of *variable short*, *variable load* data files used for the fits. The second header line is the title for the data columns.

Each line contains

frequency[GHz]

|ρ| [dB] modulus of calibrated DUT reflection coefficient [dB]

|ρ| [deg] phase of calibrated DUT reflection coefficient

Δ [dB] standard deviation *(σ)* of reflection coefficient [dB]

Δ [deg] standard deviation *(σ)* of reflection coefficient [deg]

|ρ (1+σ/ρ)| [dB] upper 67% (i.e. +1 standard deviation) error bar [dB]

|ρ (1−σ/ρ)| [dB] lower 67% (i.e. -1 standard deviation) error bar [dB]

Further output tab-separated files (`s###.###.txt, l###.###.txt`) are written for short and load data at all frequencies, representing the circles in the complex plane.
Each file contains a header line containing the frequency [GHz] and then 360 (or more) lines containing

i	line number
$R_{0,1}cos(i\ deg)+Re(X_{0,1})$	Best-fitting circle from fit
$R_{0,1}sin(i\ deg)+Im(X_{0,1})$	
$R_{0,1}(1+\varepsilon/R)cos(i\ deg)+Re(X_{0,1})$	Upper error bar of best-fitting circle from fit
$R_{0,1}(1+\varepsilon/R)sin(i\ deg)+Im(X_{0,1})$	
$R_{0,1}(1-\varepsilon/R)cos(i\ deg)+Re(X_{0,1})$	Lower error bar of best-fitting circle from fit
$R_{0,1}(1-\varepsilon/R)sin(i\ deg)+Im(X_{0,1})$	
x_i	i-th measurement
y_i	

A `gnuplot` script is generated in the file `gnuplotcmd` to plot all the circles from these files.
Another tab-separated file, `Pow.txt`, is written as a monitor. It contains a header line and a simple two-column list of frequency, maximum measured power (in VNA's uncalibrated dB) with the *variable short* standard. It is sometimes useful to correlate frequency ranges of visibly lower accuracy with lower available power.
Finally, another text file, `NLoadsUsed.txt`, is written to document the actual amount of data used in fits. It contains a header including the number of positions available and then one line per frequency point, containing the frequency and the number of positions used, taking into account both the exclusions in the `mask.txt` file and those due to the Chauvenet criterion.
The table below summarizes inputs and outputs.

INPUT	**DESCRIPTION**
`short.txt`	list of files for *short* measurements
<files>	text files (in VNA format) named in `short.txt`
`load.txt`	list of files for *load* measurements
<files>	text files (in VNA format) named in `load.txt`
OUTPUT	**DESCRIPTION**
`DUT1.txt`	calibrated reflection coefficient of DUT 1
...	...
`DUT<N>.txt`	calibrated reflection coefficient of DUT <N>
`s###.###.txt`	at frequency ###.###: measured short data, fitted circles in complex plane
`l###.###.txt`	same as above, for load data
`gnuplotcmd`	gnuplot script to plot s###.###.txt and l###.###.txt files into .png files
`Pow.txt`	maximum measured signal (uncalibrated) at every frequency
`NLoadsUsed.txt`	number of load datasets used at each frequency

Table 1: summary of input, output files

Results

A target with very low reflectivity was used for a test. Measurements were made with the technique described in [1] (see Figure 1). A portion of an AN73 sheet, slightly larger than the DUT, was then fixed in front of the DUT and the position scan was repeated to obtain the

variable load reference. A single measurement with frequency sweep was then made on the DUT.
The data were processed using (a new version of) the program of ref. [1], and the results are shown with the red traces of Figure 2. Then the AN73 data were used as reference load with the calibration described here, obtaining the green traces in Figure 2. Finally, the data obtained by sweeping the DUT in position, already used for computing the red traces, were used as *variable load* in this calibration technique. The result is shown with the blue traces in Figure 2.

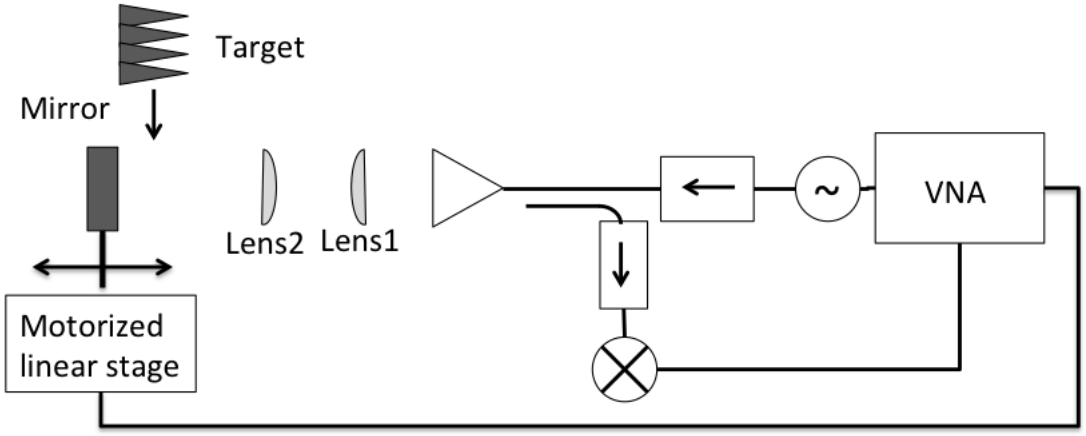

Figure 1: measurement scheme

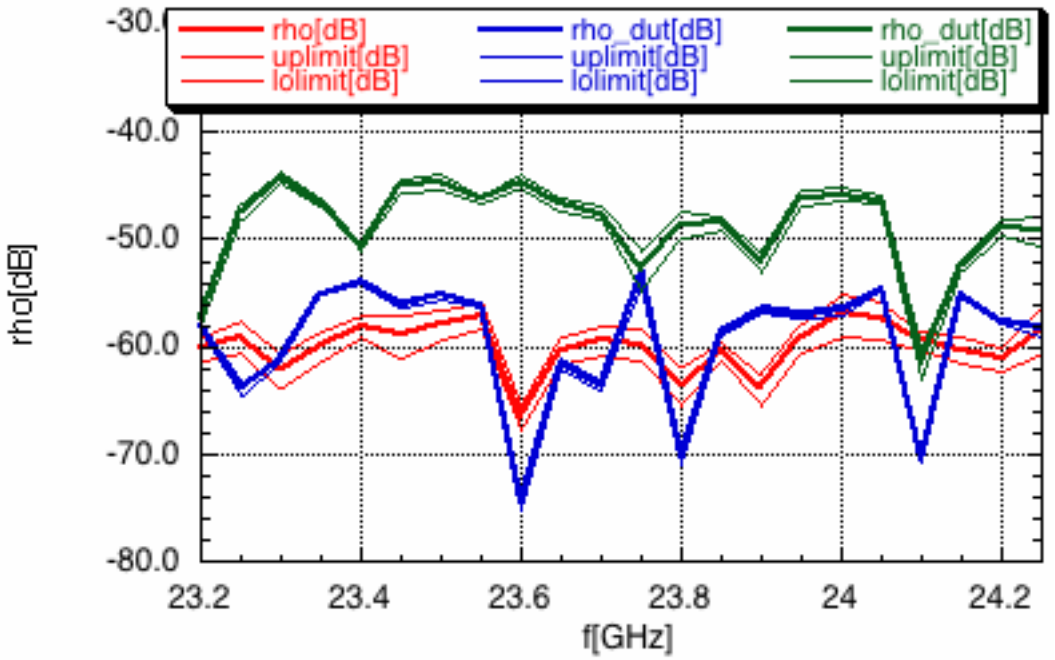

Figure 2: comparison of results using the technique of [1] (red traces) and the technique described in this report. Green traces: results obtained using AN73 as *variable load*. Blue traces: same, using the DUT as *variable load*. The thick traces are results, thin ones represent upper, lower 1 standard deviation error bars.

The DUT is a very high quality load, satisfying with ample margins the -40 dB specifications on reflectivity.
If the calibration is made using AN73 (expected reflectivity in this frequency range between 17 and 23 dB according to the datasheet [7]) the DUT is still within specifications, but the margin is nearly exhausted because of the incomplete removal of systematics. Nonetheless, the measured reflectivity is approximately 20 dB below that of the *variable load*, which means that the load used for calibration can be somewhat worse than the DUT.

When the calibration is made using a very high quality load, as the DUT itself, results are in close agreement with those obtained with the technique described in [1]. One should note that the data used here for the *variable load* are totally independent on those used as DUT, because the former belong to a set of single-frequency position scans, whereas the latter is an independent frequency scan at a single fixed position.

The DUT calibration with the technique described in [1] derives from simultaneous processing of many independent files (19 in this case) of the DUT, whereas in this case only one data file is used. As a result one should expect random variations to be more important in this case by √(19), i.e. about 6.4 dB, which is indeed similar to the discrepancy between red and blue traces in Figure 2.

The AN73 reflectivity estimated above was confirmed using the DUT as variable load and AN73 as DUT (the missing AN73 frequency sweep was simulated extracting the measurements at reference position from all the position sweeps). The result is shown in Figure 3 and confirms both the estimate above and the approx 20 dB improvement of measurable reflectivity with respect to that of the variable load.

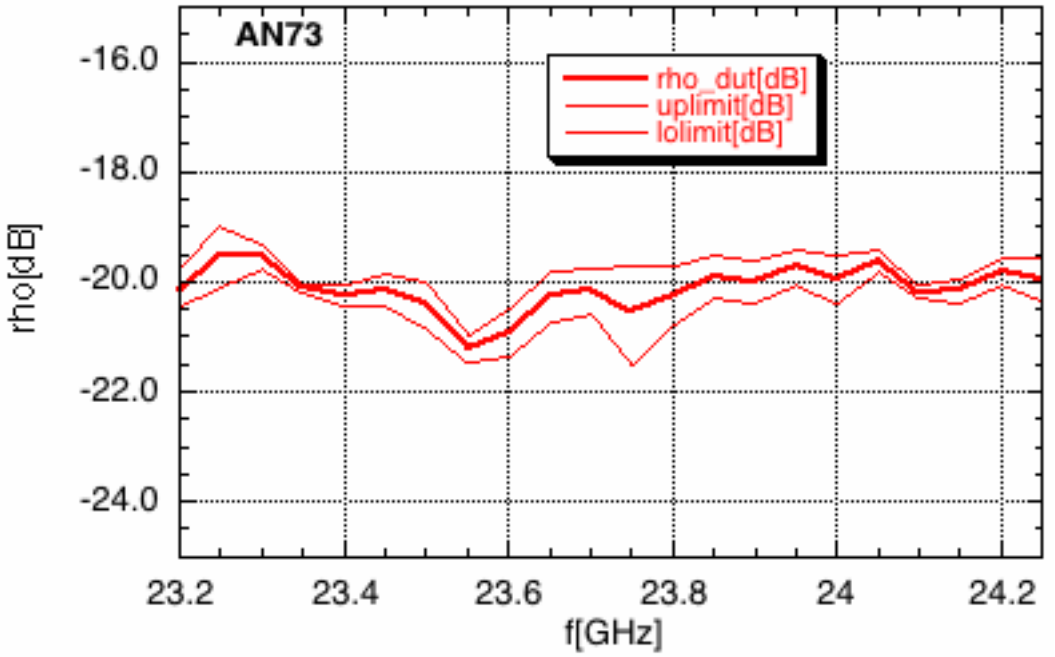

Figure 3: measured reflectivity of AN73, using the high quality load as variable load.

Conclusions

The calibration technique described here is applicable to heavy or large DUTs, for whom the technique described in [1] is difficult to apply. Quick tests using an excellent absorbing load show that results with the two techniques can be comparable, provided that the *variable load* standard is of adequate quality for the DUT.

Tests show that this calibration technique allows measuring DUTs with a return loss about 20dB higher than that of the *variable load*, which can be used as a rule of thumb for the selection of a suitable *variable load*.

As a side effect, this calibration style can provide the phase of the reflection coefficient, which can sometimes be of interest for DUTs of moderate reflectivity, allowing for example to provide estimates of the spatial location of the main contributions to the reflection coefficient.

References

[1] A. Simonetto, E. Sassolini *Free-space measurement of low-reflectivity targets* IFP Internal Report FP16/04

[2] A. Murk, A. Duric and F. Patt *Characterization of ALMA Calibration Targets,* 19th International Symposium on Space Terahertz Technology, Groningen, 28-30 April 2008. [Online] http://www.nrao.edu/meetings/isstt/papers/2008/2008530533.pdf

[3] A. Murk, A. Duric *ALMA Calibration Device Prototype Calibration Load Test Report* FEND-40.06.04.00-005-A-REP [Online] https://safe.nrao.edu/wiki/pub/ALMA/CalAmp/FEND-40.06.04.00-005-A-REP.pdf

[4] G. F. Engen *Microwave circuit theory an foundations of microwave metrology* Peter Peregrinus Ltd on behalf of IEE, ISBN 0 86341 287 4, Chapter 9

[5] L. Moura, R. Kitney, *A direct method for least-squares circle fitting,* Computer Physics Communications 64 (1991) 57—63

[6] J. R. Taylor, *An introduction to error analysis*, University Science Books, Sausalito CA (USA)

[7] Laird Performance Materials DS ECCOSORB AN datasheet