# A METHOD FOR COMPARING VECTOR NETWORK ANALYZERS

Donald C. DeGroot, Roger B. Marks, and Jeffrey A. Jargon
National Institute of Standards and Technology
Mail Code 813.06, 325 Broadway, Boulder, Colorado 80303-3328
Phone: (303) 497-7212; FAX: (303) 497-3970
E-Mail: degroot@nist.gov

## Abstract

We present a method of comparing two distinct vector network analyzer systems by taking the differences in calibrated *S*-parameters over a set of test devices. The maximum magnitude of all *S*-parameter differences in the ensemble of data provides an estimate of the upper bound on the system differences for the set of test devices measured. If the maximum ensemble difference is greater than the repeatability limits, either the residual errors in the two systems are not negligible, or they do not agree. We demonstrate our method here by making comparisons between two commercial frequency-domain network analyzer (FDNA) systems and by comparing an experimental time-domain network analyzer (TDNA) to a commercial FDNA.

## I. Introduction

In this paper we examine the issues involved in comparing two separate vector network analyzers (VNAs), develop a new approach for determining agreement or lack of agreement between distinct systems, and demonstrate our technique by comparing various network analyzers. Our intent is to provide a means of assessing the uncertainty in a new measurement system relative to a reference system, while offering a possible extension to interlaboratory comparisons of round-robin verification devices.

Both the developer of a new microwave measurement system and the user of an existing VNA require a means of comparing their instrument or method to a widely accepted and trusted system. Usually, a qualitative comparison is made by plotting calibrated data from a few example devices against calibrated data collected from the reference system. If multiple measurements are made on a single device, then statistical methods can be applied, similar to the interlaboratory comparison techniques developed at the National Institute of Standards and Technology [1, 2] and the National Physical Laboratory [3]. To date, these types of evaluations are all conducted on a device-by-device basis but do not provide an assessment over an ensemble of comparison data.

The issue we address deals with comparing two physically dissimilar network analyzers and the validity of their assumed error models over a specific set of verification devices. We propose a direct method of assessing the differences between two systems for a set of measurements by comparing the maximum of an ensemble difference to instrument repeatability limits. While existing methods [4, 5] are used to compare different calibration techniques on one vector network analyzer, this work provides a new and convenient approach to comparing a number of data sets collected on two distinct measurement systems.

## II. Network Analyzer Errors

Vector network analyzers have been described using linear calibration error models, such as the 12-term model commonly employed today. However, such models cannot account for all systematic and random errors. Since we are interested in determining whether the unaccounted systematic error in a given system is significant relative to a reference system, we must first make a distinction between unaccounted residual errors, which are systematic, and the repeatability errors, which are random.

### *A. Residual Errors*

There are various sources of residual errors. One is related to imperfections in the descriptions of calibration standards [6]. Since a physical artifact differs from even its best electrical characterization, the differences between reality and an approximation generate residual errors in the terms of the VNA error model. Under the linear system assumption, these residual errors could be determined completely with a finite set of measurements.

Differences between the physical equipment and the approximate error model are another source of residual error. A key example is system nonlinearity. All VNA error models are based on the linearity assumption of *S*-parameter measurements. Any system nonlinearity generates residual errors that cannot be accounted for to date. Further, it may be impossible to completely characterize this type of error with a finite set of ideal standards.

### *B. Repeatability Errors*

Repeatability errors, on the other hand, arise from random changes in the measurement system during the calibration and measurement sequence. Connector repeatability and system drift are examples. For a specific device, these errors can be assessed by taking multiple measurements of the same device and determining an average and a deviation from the average [1, 2]. Alternatively, the calibration comparison method [4, 5] can be used to estimate an upper bound on the maximum deviation between two measurements of any device.

In this paper, we use the calibration comparison method to estimate the repeatability uncertainty in our network analyzers. This method uses the error terms from two sequential calibrations. If the two calibrations are identical except for influences from the random errors encountered during the calibration process, then the error terms can be used to estimate the maximum difference one would encounter between the nominal value of any *S*-parameter $S_{ij}$ and subsequent measurements of that same *S*-parameter. This maximum deviation called the repeatability bound $\Delta_R$ and can be visualized as defining the loci of possible values around all measured values of $S_{ij}$ from the system under investigation (Fig. 1).

## III. VNA Comparison Method

To make a comparison between one VNA and a reference VNA, we first calibrate both systems and then collect corrected *S*-parameters from a set of verification devices using the two vector network analyzers. We take differences between the two measurements of all $S_{ij}$ for all verification devices and then find the maximum magnitude of the difference among the ensemble. This maximum difference we call the difference bound $\Delta_D$.

The difference between any two measurements of a specified *S*-parameter is a vector with length

$$\delta_{ij} = |S_{ij}^A - S_{ij}^B|,$$

where $i$ and $j$ are indices specifying a two-port $S$-parameter, and $A$ and $B$ identify two measurements made on the two different VNAs. This is illustrated in Fig. 2.

For each of the devices in our verification set, we find the maximum $\delta_{ij}$:

$$\delta_{max}^m = \max_{ij}\{\delta_{ij}^m\},$$

where the maximum is taken over all $i$ and $j$ and where $m$ identifies one of the verification devices.

Finally, to form the difference bound, we find the maximum of all $\delta_{max}^m$:

$$\Delta_D = \max_m\{\delta_{max}^m\},$$

where the maximum is taken over all verification devices in the set.

This difference bound gives an estimate of the maximum deviation between any $S$-parameter measured with the two systems. If system $A$ is the reference with negligible error, $\Delta_D$ gives an estimate of the uncertainty in measurements made with system $B$ (Fig. 2). This bound, however, is valid only for the specific set of devices used to form the ensemble of verification data.

Once the difference bound is determined, we can also make an assessment of the residual error by considering system repeatability. To accomplish this, we measure the repeatability bounds $\Delta_R$ for the two VNA systems in the comparison. If the calculated difference bound $\Delta_D$ is bigger than the sum of the repeatability bounds $\Delta_R^A + \Delta_R^B$, we say system $B$ does *not* agree with system $A$ within the repeatability limits of the two systems; in other words, the differences in residual errors are larger than the random repeatability errors (Fig. 3a). If $\Delta_D \le \Delta_R^A + \Delta_R^B$, we say the differences in residual errors are not larger than the predicted repeatability error or that the systems agree within the repeatability limits (Fig. 3b).

## IV. System Comparisons

This section provides three example VNA comparisons. These illustrate the application of our method to two similar systems where it is likely the error models should account for the systematic errors; to two similar systems when there are errors in the calibration artifact descriptions; and to an experimental time-domain network analyzer (TDNA) system where it is not obvious that a conventional VNA error model should be effective.

We compared both a commercial three-sampler FDNA and an experimental TDNA to a commercial four-sampler FDNA system using the method described above. We assumed our high-end four-sampler FDNA to be our most accurate system and used it as our reference system (system $A$). For each comparison, we determined the repeatability bound $\Delta_R$ of both systems with back-to-back calibrations as described above. We then measured calibrated $S$-parameters for a set of verification devices and determined the difference bounds $\Delta_D$ over the entire ensemble of data, and also for a subset of devices. By plotting both the repeatability bound $\Delta_R^A+\Delta_R^B$ and the difference bound $\Delta_D$ on the same graph, we determined if the residual errors exceeded the worst-case repeatability limits predicted by the calibration comparison method.

First, we compared the three-sampler FDNA to our four-sampler reference system. For this experiment, we performed an OSLT (open-short-load-thru) calibration on each system using a set of commercial 7 mm coaxial standards. The electrical behavior of the standards was defined using the manufacturer's specifications. Our set of verification devices consisted of four coaxial devices: namely a 20 dB attenuator, a 50 dB attenuator, a 10 cm air line, and a mismatch airline. These were intentionally chosen to be different than our calibration standards. Our measurements in this comparison covered a frequency span of 50 MHz to 6 GHz, which is the limit over which the two systems overlap. Figure 4 plots the sum of the repeatability bounds of the two systems along with the difference bounds determined using all four of the devices in the verification set. The difference bound for all four devices in this comparison nearly duplicates the repeatability bound estimate over the entire frequency range. We would say the two systems agree within the repeatability limits, at least for the four specific verification devices measured.

In addition, we determined the difference bound using only two of the devices (the attenuators). This curve, also plotted in Fig. 4, clearly shows that the difference bound will depend on the types of devices chosen for the verification set. If the differences between the two systems are due to different linear residual errors, it may be possible to quantify a general worst-case bound based on a finite set of verification devices. However, if the residual errors result from nonlinear processes, then the differences between any two systems may not be generalized using a finite set of measurements.

Next, we compared the same two systems, but this time we wanted to observe the effects of defining the standards incorrectly for the three-sampler FDNA. To do this we intentionally used calibration kit definitions from 3.5 mm coaxial standards even though we were actually using 7 mm coaxial standards. We then performed an OSLT calibration and measured the same set of verification devices as above. Figure 5 shows the repeatability bounds of the two systems and the difference bounds for the two-attenuator subset. The first point to note is that the repeatability bound increased somewhat over the previous experiment due to incorrect standard definitions, but this difference is probably not sufficient to use as an indicator of calibration error. Second, the difference bound for the two-attenuator subset is again well below the repeatability bound for most of the frequency range. However, when we included all four devices in $\Delta_D$, the difference bound greatly exceeded the repeatability bound (Fig. 6). This not only illustrates again that the difference bound is truly valid only for the specific set of devices measured, but it also shows that repeatability measurements alone may not reveal serious systematic errors. In this case, the two calibrated systems, although repeatable, do not agree.

Finally, we compared an experimental time-domain network analyzer [7, 8] to our four-sampler FDNA system. In this case, the architecture of the two systems is radically different. Unlike a conventional FDNA, the TDNA uses a calibrated digital sampling oscilloscope with time-domain reflection/transmission (TDR/T) capabilities. The system does not use directional couplers, but rather calibrates Fourier-transformed reflection and transmission waveforms using a conventional network analyzer error model. For this experiment, we performed an on-wafer multiline TRL [9] calibration on each system. The standards were coplanar waveguide (CPW) structures on a GaAs wafer. Our verification set consisted of three CPW transmission lines of varying length (3 mm, 6 mm, and 19 mm), a two-port CPW short circuit, and a two-port CPW resistor termination. Our measurements covered a frequency span of 45 MHz to 12 GHz for this comparison. Figure 6 plots the sum of the repeatability bounds of the two systems with $\Delta_D$ computed for two of the verification devices (the short circuit and the resistor) and with $\Delta_D$ based on all five of the verification devices.

In this case, the repeatability bound of the TDNA system is much larger than the reference system. Here, the five-device $\Delta_D$ approximates the estimated repeatability errors and, over the middle frequencies, exceeds the predicted repeatability errors. This may indicate that the error model used for the TDNA calibration may not account for all systematic errors appropriately. Certainly, if the difference bound is larger than an acceptable uncertainty, the comparison identifies the need for further study of the sources of error in the experimental system.

## V. Summary

We have developed a new method of comparing two distinct vector network analyzer systems using measurements of a specific set of verification devices. This method produces a single scalar estimate of the worst-case differences between an ensemble of measurements made with two vector network analyzers. The resulting difference bound can be used as an estimate of maximum differences between two systems. When compared to the estimated repeatability bounds, it can also be used to assess the agreement, or lack of agreement, between the two systems. This is useful not only to developers of new instrumentation but also to users of VNAs looking to verify the accuracy of their measurements.

Through the application of this method to three example system comparisons, we noted that this method is not perfectly general in that the computed difference bound cannot predict uncertainty for all devices to be measured. This is the result of the dependence of residual errors upon the characteristics of the devices.

While our paper focused on system comparisons, an application to interlaboratory comparisons may be possible. This approach is not now as statistically robust as new comparison methods being developed [2], but it offers simplicity in assessing agreement over a larger number of measurements with the use of a single scalar bound.

## References


[1] R. M. Judish, and J. G. Burns, "Measurement program compares automatic vector analyzers," *Microwaves & RF*, pp. 203-205, May, 1991.

[2] J. R. Juroshek, C. M. Wang, and G. P. McCabe, "Statistical analysis of network analyzer measurements," Submitted to *Cal Lab.*

[3] N. Ridler, and G. Jones, "Comparison assesses the quality of network measurements," *Microwaves & RF*, pp. 101-104, January, 1995.

[4] D. F. Williams, R. B. Marks, and A. Davidson, "Comparison of on-wafer calibrations," *38th ARFTG Conference Digest*, pp. 68-81, December, 1991.

[5] R. B. Marks, J. A. Jargon, and J. R. Juroshek, "Calibration comparison method for vector network analyzers," *48th ARFTG Conference Digest*, pp. 38-45, November, 1996.

[6] K. H. Wong, "Characterization of calibration standards by physical measurements," *39th ARFTG Conference Digest*, pp. 53-62, June, 1992.

[7] D. C. DeGroot, and R. B. Marks, "Optimizing time-domain network analysis," *46th ARFTG Conference Digest*, pp. 19-28, November, 1995.

[8] R. B. Marks, D. C. DeGroot, and J. A. Jargon, "High-speed interconnection characterization using time domain network analysis," *Advancing Microelectronics*, pp. 35-39 November/December, 1995.

[9] R. B. Marks, "A multiline method of network analyzer calibration," *IEEE Transactions on Microwave Theory and Techniques*, vol. 39, no. 7, pp. 1205-1215, July, 1991.

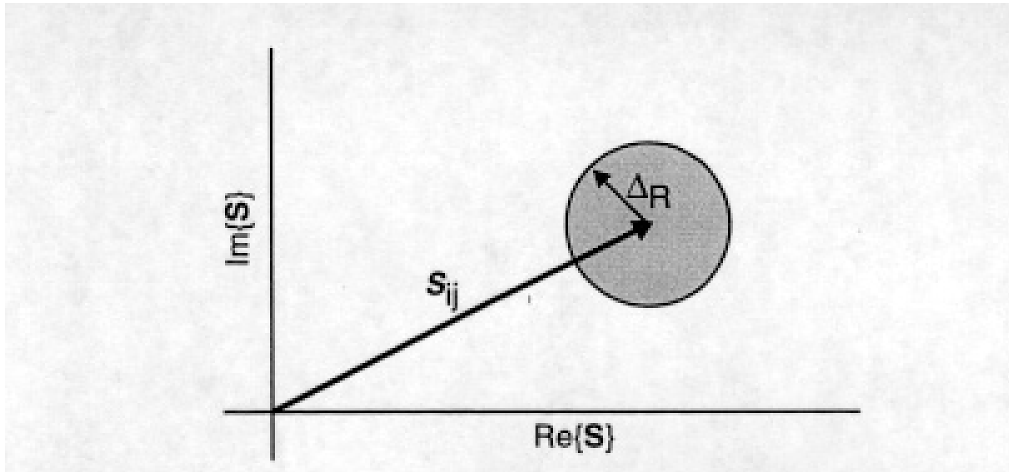


Fig. 1. Depiction of worst-case repeatability bound about an *S*-parameter.

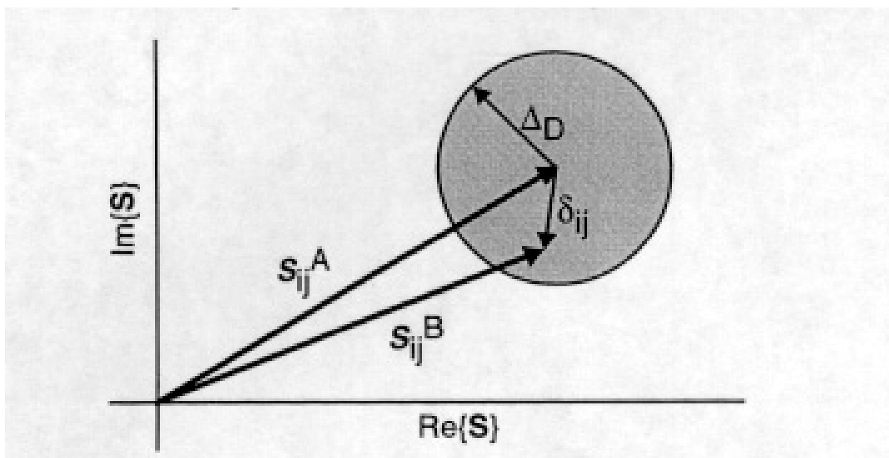


Fig. 2. Difference in two calibrated $S_{ij}$ values measured with system *A* and system *B*. The difference bound $\Delta_D$ defines the worst-case uncertainty in system *B* when system *A* is used as the reference.

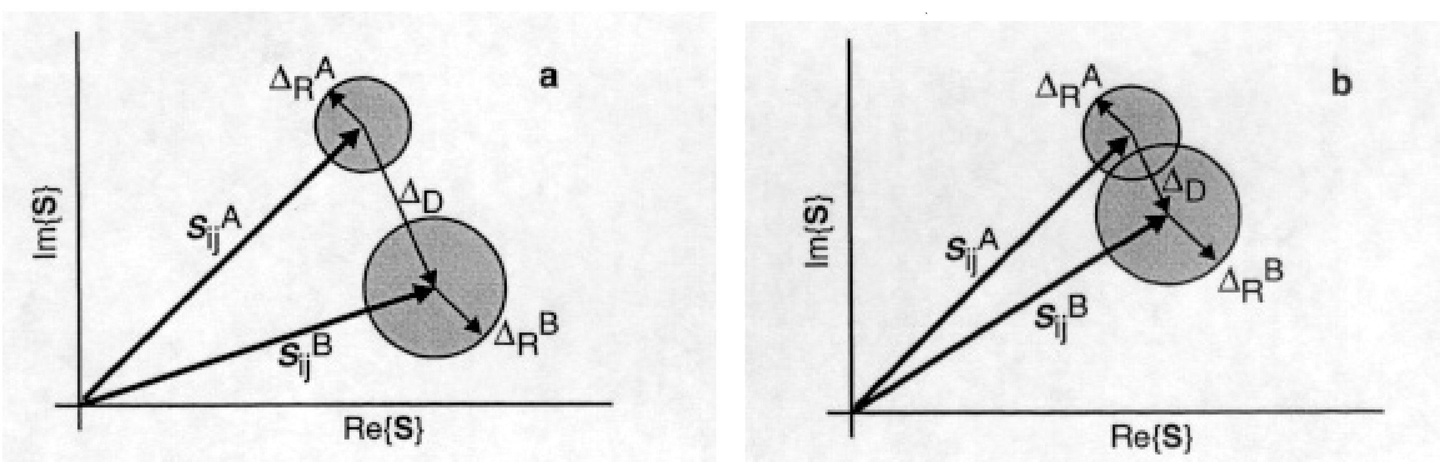


Fig. 3. a) Difference in two calibrated $S_{ij}$ values with $\Delta_D$ falling outside the estimated repeatability bounds $\Delta_R^A + \Delta_R^B$; b) Difference in two calibrated Sij values with $\Delta_D \leq \Delta_R^A + \Delta_R^B$.

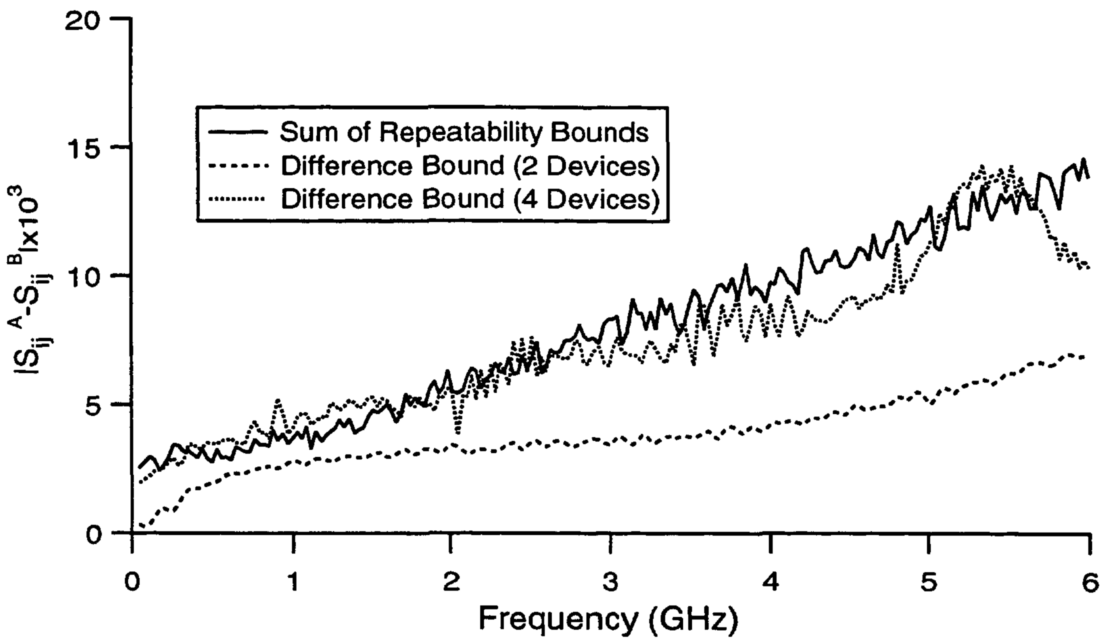


Fig. 4. Four-device and two-device comparisons of a three-sampler frequency-domain network analyzer to a four-sampler reference FDNA system.

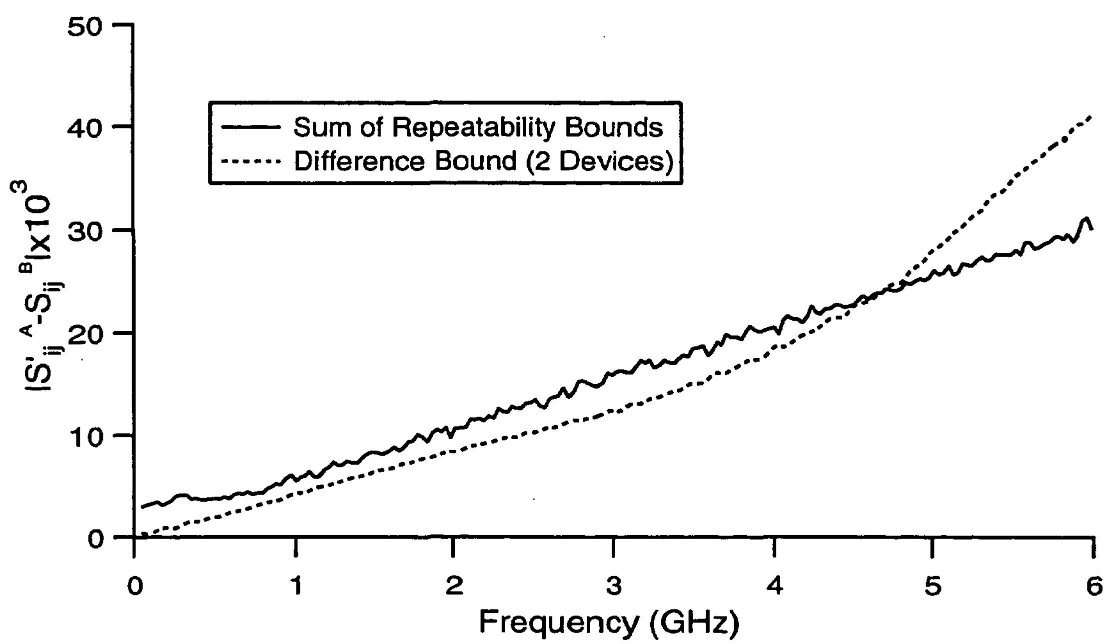


Fig. 5. Two-device comparison of a three-sampler FDNA to a four-sampler reference FDNA with known errors in the electrical descriptions of the three-sampler FDNA calibration standards.

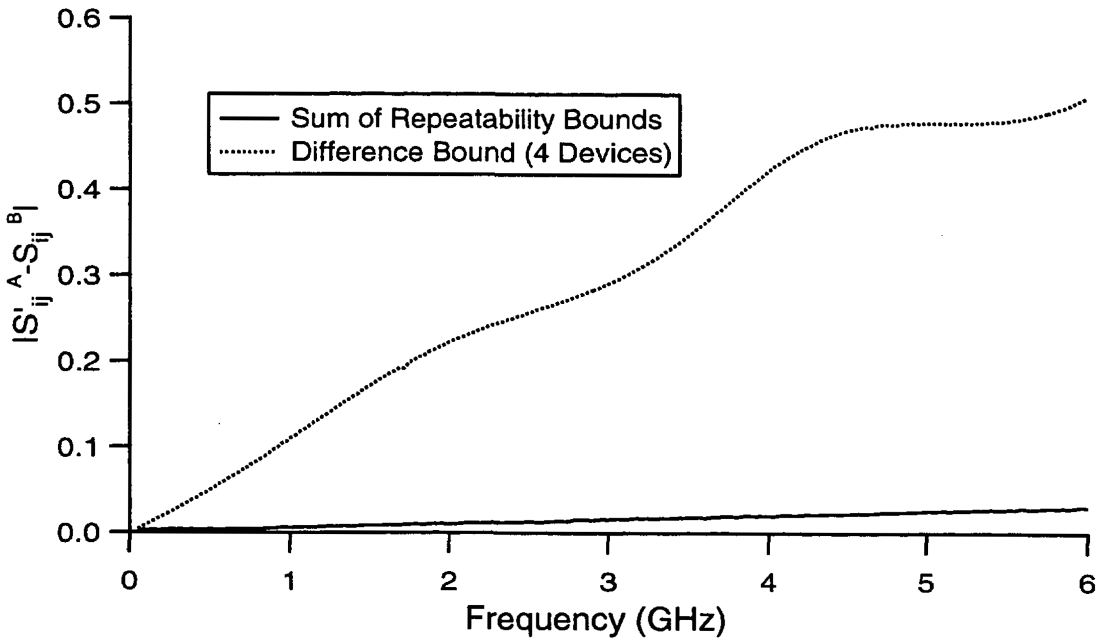


Fig. 6. Four-device comparison of three-sampler FDNA to four-sampler reference FDNA with known errors in the electrical descriptions of the three-sampler FDNA calibration standards.

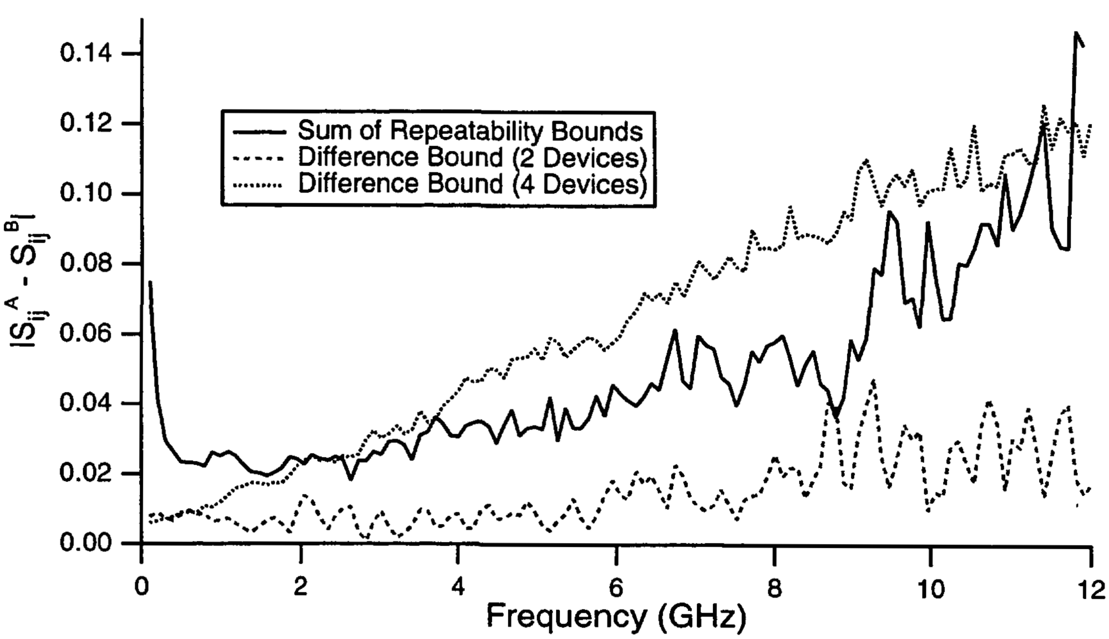


Fig. 7. Five-device and two-device comparisons of an experimental time-domain network analyzer to a reference FDNA system.