ACCURATE MEASUREMENT OF OPTICAL DETECTOR NONLINEARITY

Shao Yang*, Igor Vayshenker, Xiaoyu Li, and Thomas R. Scott
National Institute of Standards and Technology
Electromagnetic Technology Division
Boulder, Colorado 80303
*University of Colorado
Department of Electrical and Computer Engineering
Boulder, Colorado 80309

Abstract

In this paper we describe the results of our efforts to analyze and compare by computer simulation three common methods of measuring detector nonlinearity: (1) superposition method, (2) attenuation method, and (3) differential or ac-dc method. We describe a definition and expressions which we used to intercompare the data analyses of these methods. Issues that are common to these methods or specific to one individual method and have an impact on the measurement accuracy were studied. We conclude that superposition and differential methods are better choices than the attenuation method. Suggestions about the choice of polynomial order of the fitting curve with regard to the data accuracy and detector nonlinearity are also given.

Introduction

To extend the useful power range of optical fiber power measurement equipment, it is often desirable to characterize their output response nonlinearity over large power ranges. When selecting the optimum method for performing these nonlinearity measurements, we must carefully consider the major issues that affect the uncertainty of the results. These issues include: measurement procedures, characteristics of the data acquired, and the data-processing algorithms used. Unfortunately, results of different measurement methods are often reported in terms that make comparisons difficult.

A quantitative analysis is helpful in choosing a nonlinearity measurement method which will achieve the highest accuracy. Since it was impractical for us to test every option by experiment, we conducted simulations to study these issues. We compared three commonly used methods under identical conditions and studied the error propagation in each case. The results of this study are presented using a single definition of nonlinearity in order to assure a straight-forward comparison of the results.

Definition and Basic Expressions

Detector nonlinearity is defined as the relative difference between the responsivity at an arbitrary power and the responsivity at the calibration power[1]. It can equivalently be expressed as an error due to the nonlinearity in the measurement of optical power incident onto the detector:

$$NL \equiv (P_m - P) / P, \qquad (1)$$

where P is the incident optical power and P_m is the measured optical power. The nonlinearity NL, thus, varies with input power. P_m depends on the power at which the detector is calibrated. For convenience, this definition of nonlinearity is expressed in terms of the detector output V, which can be electric current, voltage, or the reading of a power meter:

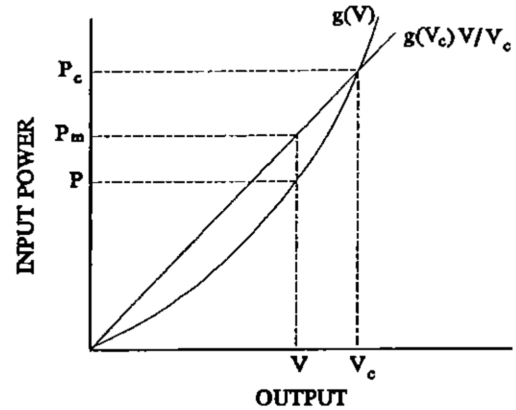

Figure 1. A conversion function.

$$NL(V;V_c) = \frac{g(V_c)V}{g(V)V_c} - 1, \qquad (2)$$

where V is the output at which nonlinearity is evaluated, V_c is the output at which the detector is calibrated, and $P=g(V)$ is called the conversion function, which relates the input power P to the output V. Once the conversion function g(V) of a detector is known, its nonlinearity can be calculated. Figure 1 illustrates the conversion function. The inverse function of the conversion function is called the response function and gives the output response per input power relationship. The response function and the conversion function represent the same curve in inverted variables. The nonlinearity can equivalently be expressed and calculated in terms of the response function and input power P.

It is often necessary to evaluate the nonlinearity of a detector before it is calibrated and sometimes even though it has been calibrated, the calibration point needs to be changed. For most commonly used detectors whose nonlinearity is no more than a few percent, we derived a simple expression that relates nonlinearity referenced to one calibration point V'_c to nonlinearity referenced to another calibration point V_c:

$$NL(V;V'_c) = NL(V;V_c) - NL(V'_c;V_c). \qquad (3)$$

This expression allows us to evaluate the nonlinearity of a detector by assuming a dummy calibration point and later translate this nonlinearity to a new value referenced to the real calibration point. For theoretical evaluation, the dummy calibration power can be 0.

When the nonlinearity is small, a polynomial is a good representation of the conversion function of a detector[2,3]:

$$g(V) = \sum_{k=1}^{n} b_k V^k. \quad (4)$$

It is assumed here that the dark output of the detector is always adjusted to 0. The nonlinearity then has the form:

$$NL(V;V_c) = -\sum_{k=2}^{n} b_k (V^{k-1} - V_c^{k-1}). \quad (5)$$

Measurement Methods

Methods for the measurement of detector nonlinearity can be divided into two categories: dependent and independent measurements. Dependent methods always involve the known values or the measurement of some physical quantities other than the output of the detector under test. The attenuation method falls in this category because the prior knowledge of the transmittance of a filter is required. Independent methods rely solely on the measurement of the output of the device under test. The superposition (addition) method and the differential method belong to this group. Independent methods are now favored over dependent methods. The superposition, differential, and attenuation methods are considered in this study. If the transmittance of the filter is treated as an unknown parameter in the curve fitting, the attenuation method without prior knowledge of the transmittance of the filter can also be used. In this way, it becomes an independent method.

Superposition method[2,4,5]

The superposition method relies on the fact that for a linear detector, the sum of the outputs corresponding to two individual beams should equal the output when the two beams are combined and incident on the detector at the same time. If the outputs of the individual and combined beams are measured at enough points, the conversion function g(V) can be derived by curve-fitting and the nonlinearity of the detector can be calculated.

Attenuation Method[4]

A filter with a density constant throughout the power range of the measurement is used as an optical power attenuator. Powers before and after the attenuation are measured at several different levels. If the detector is linear, the attenuation at different powers should be the same. If it is not linear, the conversion function can be determined from the measurement data. If the transmittance of the filter is known, the method falls in the dependent group; if it is unknown, it is an independent method. Data processing is different for these two cases.

Differential Method[3]

A small, constant ac power is superimposed on a dc power. A dc meter and an ac meter are used to measure the ac output at different dc outputs. When the ac input is sufficiently small, the ac output is approximately proportional to the reciprocal of the derivative of the conversion curve at V. The first derivative of the conversion function is thus obtained, and the conversion function is determined by integration.

Computer Simulation

We conducted computer simulations to study the effects of random error in the data, systematic error due to the truncating of polynomial, their combined effect, and their effect on different methods. The simulations were performed in the following way. We assumed a polynomial of finite order as an original response function whose nonlinearity is known. Usually, a real conversion function would need a polynomial of infinite order to represent it. Also, the inverse function of an arbitrary polynomial of finite order is generally a polynomial of infinite order. In this sense, the conversion function which is the inverse function of our assumed response function, resembles the conversion function of a real detector, and we can use it to study the error due to the truncating of higher orders. The specific forms of the response functions we assumed for our simulations are given in the Appendix.

From the assumed response polynomial, we created data of incident powers and their corresponding outputs in the same way as they are taken by different methods. Conversion polynomials of orders from 2 to 5 were determined from these data by curve fitting. The difference between the resultant curve and the original curve is the systematic error due to the truncation of the polynomial. The magnitude of this systematic error will depend on the methods used.

We then introduced random errors with chosen standard deviations into the output data (V) and processed the data to get the conversion curve. The random errors are Gaussian distributed. The difference now between the conversion curve and the original curve is caused by both the truncation of the polynomial and the random deviation of the data. This difference is the error in the measurement of the nonlinearity. We call it nonlinearity error in the presentation of the simulation results below.

This process was repeated 50 times and the average and standard deviation σ of the difference of the nonlinearities of the original and the resultant curves were calculated

$$\sigma = \left[\frac{1}{50}\sum_{j=1}^{50}(NL_j - NL_o)^2\right]^{\frac{1}{2}}, \tag{6}$$

where NL_o denotes the nonlinearity of the original curve and NL_j denotes the resultant nonlinearity of the jth measurement. This standard deviation can also be expressed as

$$\sigma = \left[(\bar{NL} - NL_o)^2 + \frac{1}{50} \sum_{j=1}^{50} (NL_j - \bar{NL})^2 \right]^{\frac{1}{2}}, \tag{7}$$

where $\bar{NL}$ is the average nonlinearity of the 50 measurements. The first term represents the systematic error due to truncation of the polynomial, and the second term is the uncertainty caused by data random error scatter. We call σ in Equation (6) total error or combined error. In the figures of the simulation results, this σ is called nonlinearity error.

There are three types of random errors representing different measurement situations. The first type is where all the data have the same absolute standard deviations independent of the power P. This is the case when the error is due to the least count of the digital meter or to the detector's dark current noise. Over one decade of power, the relative random error for this type of data varies by one order of magnitude. The second type results when the data have equal relative standard deviations. In this case, the standard deviation of the random error is proportional to the power. A typical example is noise due to source intensity fluctuations. The third type is error due to shot noise. In this case, the error is proportional to the square root of the power. Our simulation considers the first two cases only because the magnitude of the error in the third case lies somewhere between the first two cases.

The simulation results are presented in the following figures, where, when different methods are presented on the same plot, SP stands for superposition, AT-1 for attenuation method with known transmittance, AT-2 for attenuation method with unknown transmittance, and DF for differential method. The response functions we assumed in the simulations are:

3%: $V = 2P + 0.05P^2 + 0.008P^3 + 0.002P^4$
1%: $V = P - 0.03P^2 + 0.05P^3 - 0.01P^4$
0.5%: $V = P + 0.013P^2 - 0.002P^3 - 0.006P^4$
0.15%: $V = 2P + 0.005P^2 - 0.003P^3 + 0.001P^4$

where the percentages are nonlinearities calculated at P=1 with a dummy calibration power at 0. These percentages are also used as indicators of these functions in the figures. When only one response function is involved, it is always the first function listed above.
The power range considered is P=0.1 to P=1 in arbitrary units. The dummy calibration power is always at 0 unless otherwise stated. The transmittance is 0.5 for attenuation method.

Figure 2 shows the systematic error due to truncation of the polynomial for the methods considered. The same trend is shown for all the methods: the systematic error decreases with increasing order of polynomial. The superposition method and attenuation method with known transmittance have almost the same systematic error (they almost overlap in Figure 2), while differential method yields a slightly larger error and attenuation method without a known τ gives significantly larger errors. The difference is caused by the different fitting equations used for these methods. For example, instead of fitting the conversion curve directly, we are actually fitting the derivative of the conversion curve in the differential method.

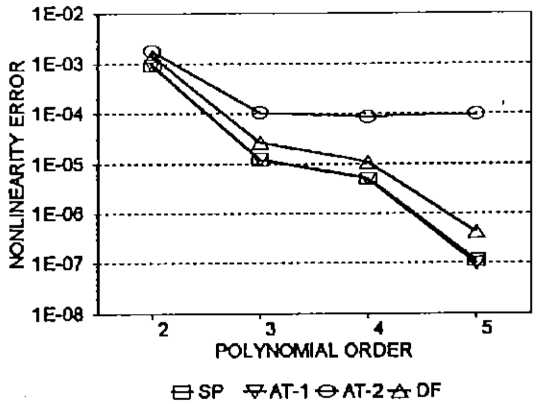

Figure 2. Systematic error due to the truncation of polynomials.

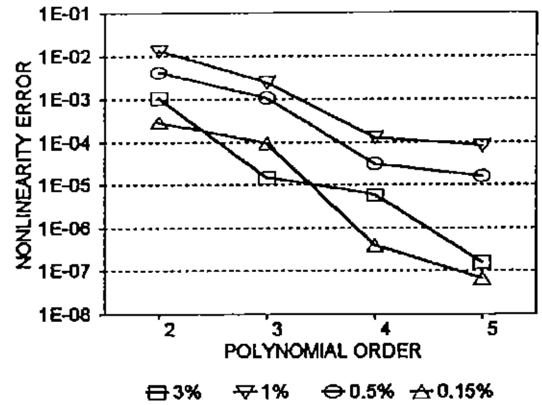

Figure 3. Systematic error of superposition method for four response functions with different nonlinearities.

The same simulation was conducted with four assumed response functions, each having a different nonlinearity. The relative significance of the terms in each polynomial was varied from one polynomial to another. The result for superposition method is shown as an example in Figure 3. Although the main trend remains the same for all the cases, the magnitude of systematic error and the slope from one order of polynomial to the next can be quite different. This should depend on the relative significance of each term of the conversion polynomial instead of the nonlinearity.

Uncertainties caused by random error in the data are demonstrated in Figure 4, where all the data have an error with same standard deviation of 0.01. In all cases, error of the result increases with increasing polynomial order. This direction of tilt is contrary to that of the systematic error. The explanation is that a lower order polynomial has less freedom of fitting. Its result represents the main shape of the conversion curve more than the details of the curve. It is therefore less influenced by any single data point. As a result, lower order polynomials are less affected by the random uncertainties of the data and thus yield results with smaller deviations. The attenuation method with unknown transmittance τ yields significantly larger resultant error than all the other methods, and the differential method gives the least random error.

When systematic error and random error are considered at the same time, we have the combined error, demonstrated in Figure 5 where all the data have an error with the same standard deviation of 0.0001. Starting from the highest order of polynomial, where the random error is higher than the systematic error, the random error decreases by lowering the polynomial order until it is lower than the systematic error. Then the systematic error dominates, and the slope reverses its direction. As a result, in this case, a third-order polynomial yields the least total error.

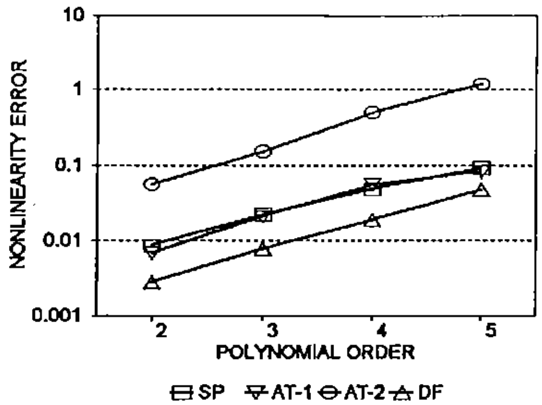

Figure 4. Uncertainty due to random data error.

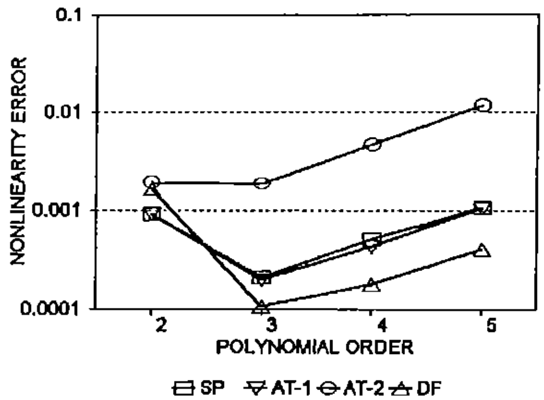

Figure 5. Combined error.

In Figure 6, we show four curves representing combined errors for four different response functions with different nonlinearity errors. These are the same four situations represented in Figure 3. The data in this case have an error with standard deviation of 0.0001. It is shown that at second order, the combined error is limited by the systematic error for all the four situations and they are all different. At the third order, two situations are systematic error limited and two are random error limited. At fourth and fifth orders, random error is dominant for all the situations and shows no difference in magnitude, which implies that the resultant error due to random uncertainty does not rely on the form of the response function. Therefore, the conclusions about the random error can be applied to any real detectors, which do not usually have similar response functions. Since combined error can never be made lower than the systematic error, we suggest using polynomials whose systematic error is well below the desired error.

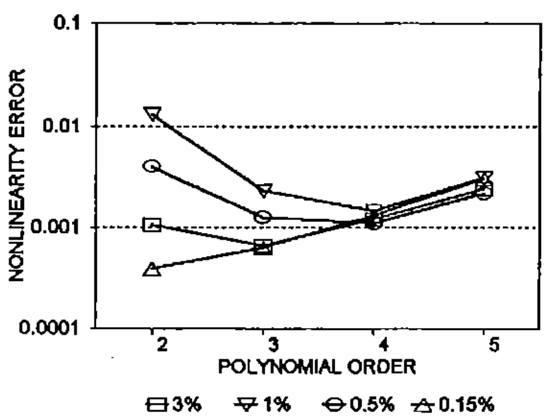

Figure 6. Combined error of superposition method for four response functions.

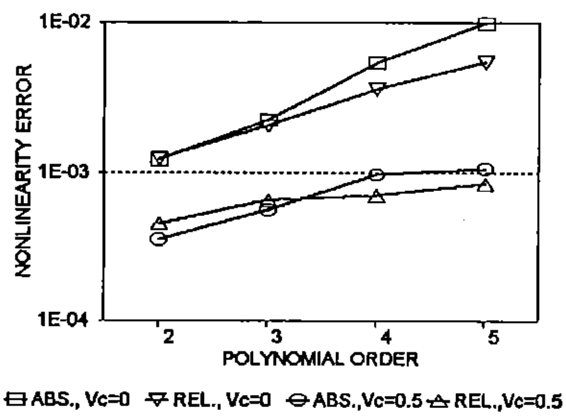

Figure 7. Errors when calibrated at zero and at mid-point for equal absolute and equal relative errors.

The figures shown above are all for data error of equal absolute error and for calibration point at 0 (V_c=0). Simulations for equal relative error and for calibration point at the middle of the measurement range (V_c=0.5) were also conducted. A comparison between NL(V;0) and NL(V;0.5) for both equal absolute and equal relative errors is shown in Figure 7 for attenuation method with known τ. The data have an error with standard deviation of 0.001 for curves of equal absolute error and 0.1% of the output V for curves of equal relative error. Equal relative error yields less error when higher order polynomials are used. Moving the calibration point from 0 to mid-point reduces the spread of the results approximately by half. All the methods benefit by nearly the same amount when moving the reference point from 0 to the middle.

Figure 8 demonstrates that, when the data have random scatter, increasing the number of data points will reduce the resultant error linearly with the square root of N, the number of data points. When the data are without error, increasing data points will have little effect.

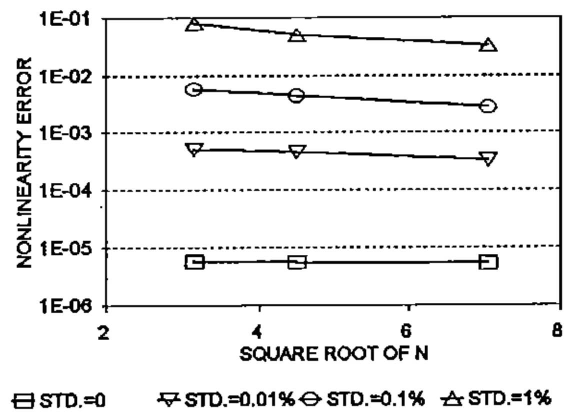

Figure 8. Effect of number of data points.

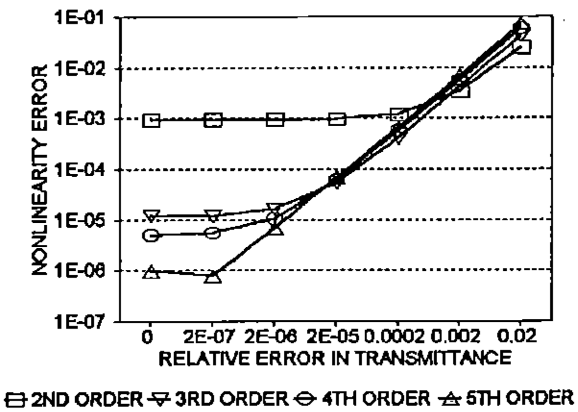

Figure 9. Effect of error in transmittance for attenuation method.

Figure 9 depicts the effect of the deviation of the transmittance τ from the true value. From this figure, we can determine what uncertainty in τ is tolerable to achieve the target accuracy of the nonlinearity measurement. Higher accuracy in the transmittance is required than the target accuracy of nonlinearity measurement. This makes the attenuation method an unfavorable choice against the other methods.

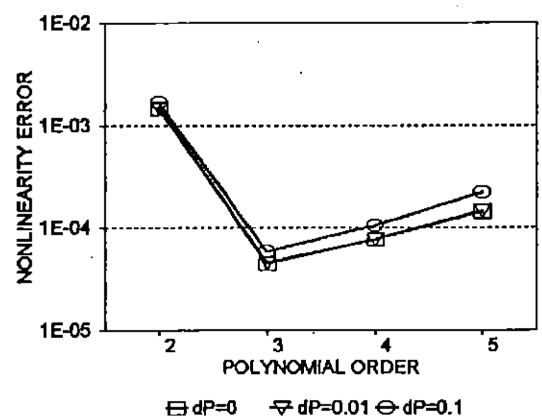

Figure 10. Error due to finite ac input for differential method.

It is assumed in the differential method that the ac input is very small, so the derivative of the conversion function is measured. In a real measurement, however, the ac input magnitude is limited at the low end by the signal-to-noise ratio and the lowest range of the ac meter. Errors due to the finite ac input are shown in Figure 10, where dP depicts the ac input. dP=0 is the case of true derivative, which can be realized by simulation. When the AC input is 1% of the maximum dc input, the induced error is negligible (the curves for dP=0 and dP=0.01 in Figure 10 virtually overlap). When the AC input is 10% of the maximum dc input, the error is noticeable but may be tolerable for the desired accuracy.

Conclusions

(1) Although detector nonlinearity is expressed with reference to the calibration point, it can be measured without first calibrating the detector and expressed with reference to any dummy calibration point. It can later be converted to nonlinearity with reference to the real calibration point. Dummy calibration at the middle of the measurement range will give results with less random spread than at the calibration point at 0.

(2) If the measurement condition can yield the same data accuracy for all the methods considered, we suggest using superposition method or differential method. However, the simulations do not tell whether the same conditions can be reached in a practical system for different methods. Many specific technical issues combined will determine the data accuracy we can have for a particular method. Once the data accuracy is known, even though they are different for different methods, we can still use the simulation results.

(3) Systematic error and random uncertainty behave differently in terms of the fitting polynomial. A compromise usually must be made. Because we can reduce the random error of the result by increasing the number of data points and the number of measurements, ensuring the systematic error well below the target accuracy should be the first consideration. Systematic error depends very much on the actual response function of the detector. A fourth-order polynomial is recommended if the nonlinearity is very small (at or less than 0.1%) and also the data standard deviation is very small (at or less than 0.01%). If the data standard deviation is not small and the detector nonlinearity is at or above 1%, third-order polynomial can be used. A fifth-order polynomial will yield a higher random uncertainty. Because using too many data points or measurements to reduce the random error is not practical and may cause other technical problems, a fifth- or higher-order of polynomial is not recommended unless the measurement system produces data with extremely low random noise.

Acknowledgements

Equation (2) is quoted from an unpublished document written by Mike Zander of the U.S. Army Missile Command, Alabama. Robert Gallawa and Paul Hale of NIST, Boulder, made useful comments and suggestions for the contents and the preparation of the manuscript.

References

1. "Calibration of Fiber-Optic Power Meters," Draft International Standard, IEC TC 86, June 25, 1992.

2. C.L. Sanders, "Accurate measurements of and corrections for nonlinearities in radiometers," J. Res. Nat. Bur. Stand. 76A, 437-453(1972).

3. R.G. Frehlich, "Estimation of the nonlinearity of a photodetector," Appl. Opt. 31, 5926-5929(1992).

4. L. Coslovi and F. Righini, "Fast determination of the nonlinearity of photodetectors," Appl. Opt. 19, 3200-3203(1980).

5. R.D. Sanders and J.B. Schumaker, "Automated radiometric linearity tester," Appl. Opt. 23, 3504-3506(1984).