Z-Scan Technique for Nonlinear Materials Characterization

Eric W. Van Stryland
CREOL
Center for Research and Education in Optics and Lasers
University of Central Florida
Orlando, Florida 32816-2700

and

Mansoor Sheik-Bahae
Department of Physics and Astronomy
University of New Mexico
Albuquerque, New Mexico, 87131

Abstract

We review Z-scan and related techniques for the measurement of the nonlinear optical properties of materials. The Z-scan is a simple technique for measuring the change in phase induced on a laser beam upon propagation through a nonlinear material. It gives both the sign and magnitude of this phase change, $\Delta \Phi$, which is simply related to the change in index of refraction, $\Delta n$. Additionally, a Z-scan can also separately determine the change in transmission caused by nonlinear absorption that is related to the change in the absorption coefficient, $\Delta \alpha$. Importantly, the determination of the nonlinear refraction from $\Delta n$ is independent of the determination of the nonlinear absorption from $\Delta \alpha$, within a quite broad range of these parameters. Thus, for third-order nonlinear responses the real and imaginary parts of the third-order nonlinear susceptibility, $\chi^{(3)}$, can be measured. However, Z-scan is sensitive to any nonlinear processes which result in $\Delta \alpha$ or $\Delta n$, so that great care must be taken in interpreting data taken with this or any other nonlinear materials characterization technique.

Keywords: Z-scan, EZ-scan, nonlinear optics, third-order susceptibility, nonlinear refraction, two-photon absorption, nonlinear absorption

I. Introduction

A variety of experimental techniques have been developed for measuring the nonlinear optical properties of materials. Since the nonlinear response can change drastically from one material to another, it is usually difficult to have one measurement that can unambiguously determine the dominant nonlinearity. Indeed there may be more than one strong nonlinearity present at a time. In general many different types of measurement techniques are useful, and often necessary, to unravel the nonlinearities present in a material and determine their origins. The Z-scan technique is a method which can rapidly measure both nonlinear absorption (NLA) and nonlinear refraction (NLR) in solids, liquids and liquid solutions.\(^1\)\(^2\) The information obtained from Z-scan measurements is
complementary to that obtained from other techniques, and this information should be analyzed in conjunction with other data whenever possible. In this paper we review only the Z-scan technique and its various closely related derivatives. Simple methods for data analysis are then discussed for “thin” and “thick” \(^3,4,5,6\) nonlinear media Z-scans, eclipsing Z-scan (EZ-scan) \(^7\), two-color Z-scans \(^8,9\), time-resolved pump-probe Z-scans \(^10,11\), and top-hat-beam Z-scans \(^12\).

Z-scan is a relatively simple method to separately measure the sign and magnitude of both NLR and NLA. It has gained rapid acceptance by the nonlinear optics community as a standard technique. In most experiments the index change, \(\Delta n\), and absorption change, \(\Delta \alpha\), can be determined directly from the data without resorting to computer fitting. However, it must always be recognized that this method is sensitive to all nonlinear optical mechanisms that give rise to a change of the refractive index and/or absorption coefficient, so that determining the underlying physical processes present from a Z-scan is not in general possible. A series of Z-scans at varying pulsewidths, frequencies, focal geometries etc. along with a variety of other experiments are often needed to unambiguously determine the relevant mechanisms. In this regard, we caution the reader that the conclusions as to the active nonlinear processes of any given reference using the Z-scan technique is often subject to debate.

2. **Method and Simple Interpretation**

![Figure 1](image1)

*Figure 1. Example of a closed aperture Z-scan performed on a sample showing self-focusing. The solid lines indicate the ray propagation for linear optics while the dashed lines show the nonlinear propagation. The upper figure shows the sample positioned at \(Z<0\) while the lower figure shows the sample at \(Z>0\).*

The usual “closed aperture” Z-scan apparatus (i.e. aperture in place in the far field) for determining nonlinear refraction is shown in Fig. 1 (example shown is for a sample exhibiting self focusing). The transmittance of the sample through the aperture is monitored in the far field as a function of the position, \(Z\), of the nonlinear sample in the
vicinity of the linear optics focal position. The required scan range in an experiment depends on the beam parameters and the sample thickness \( L \). An important parameter is the diffraction length, \( Z_0 \), of the focused beam defined as \( \pi w_0^2/\lambda \) for a Gaussian beam where \( w_0 \) is the focal spot size (half-width at the 1/e² maximum in the irradiance). A “thin” sample is defined as having a thickness \( L \leq n_0 Z_0 \) where \( n_0 \) is the linear index (later we add and discuss a further restriction on this definition that is usually automatically satisfied for a Z-scan). Although all the information is theoretically contained within a scan range of \( \pm Z_0 \), it is preferable to scan the sample for \( \approx \pm 5Z_0 \) or more to determine the linear transmittance. This requirement, as we shall see, simplifies data interpretation when the sample’s surface roughness or optical beam imperfections introduce background “noise” into the measurement system. In many practical cases where considerable laser power fluctuations may occur during the scan, a reference detector can be used to monitor and normalize the transmittance (see Fig. 2). To eliminate the possible noise due to spatial beam fluctuations, this reference arm can be further modified to include a lens and an aperture identical to those in the nonlinear arm. The position of the aperture is rather arbitrary as long as its distance from the focus, \( d \gg Z_0 \). Typical values range from 20\( Z_0 \) to 100\( Z_0 \). The size of the aperture is signified by its transmittance, \( S \), in the linear regime, i.e. when the sample has been placed far away from the focus at low energy. In most reported experiments, \( 0.1 < S < 0.5 \) has been used for determining nonlinear refraction. Obviously, the \( S=1 \) case corresponds to collecting all the transmitted light and therefore is insensitive to any nonlinear beam distortion due to

![Diagram](image_url)
nonlinear refraction. The experiment with S=1 is referred to as an “open aperture” Z-scan and allows direct measurement of nonlinear absorption ($\Delta\alpha$) in the sample. We first discuss measurement of $\Delta n$.

![Figure 3](image)

Figure 3. A typical Z-scan for positive ($\Delta\phi=0.5$, solid line) and negative ($\Delta\phi=-0.5$, dashed line) third-order nonlinear refraction using an aperture with $S=0.5$.

A typical closed aperture Z-scan output for a thin sample exhibiting nonlinear refraction, is shown in Fig. 3. For example, a self-defocusing nonlinearity, $\Delta n<0$ (dashed line in Fig. 3), results in a peak followed by a valley in the normalized transmittance as the sample is moved away from the lens in Fig. 1 (increasing $Z$). The normalization is performed in such a way that the transmittance is unity for the sample far from focus where the nonlinearity is negligible (i.e. for $|Z|>>Z_0$). The negative lensing in the sample placed before the focus moves the focal position further from the sample resulting in a smaller far field divergence and an increased aperture transmittance. On the other hand, with the sample placed after focus, the same negative lensing enhances the diffraction (increases the far field divergence) resulting in a reduced aperture transmittance. The opposite occurs for a self-focusing nonlinearity, $\Delta n>0$ (solid line in Fig. 3).

A very useful feature of the Z-scan technique is the ease by which the NLA and NLR can be separately determined and absolutely calibrated. The errors are primarily limited by errors in the determination of the irradiance, fluence and/or energy. However, as is the case with most nonlinear optical measurement techniques, the measured quantities are the nonlinearly induced $<\Delta n>$ and/or $<\Delta\alpha>$, where $<$ denotes a time-average over the time corresponding to the temporal resolution of the detection system. Accurate determination of nonlinear coefficients such as $n_2$ or $\beta$ is model dependent and can be influenced by competing nonlinearities.
Given a specific nonlinearity (e.g. an ultrafast \(\chi^{(3)}\) response), a Z-scan can be rigorously modeled for any beam shape and sample thickness by solving the appropriate Maxwell’s equations. However, a number of valid assumptions and approximations will lead to simple analytical expressions, making data analysis easy yet precise. Aside from the usual SVEA (slowly varying envelope approximation), a major simplification results when we assume the nonlinear sample is “thin” so that neither diffraction nor nonlinear refraction cause any change of beam profile within the nonlinear sample. This implies that \(L<<n_0Z_0\) and \(L<<Z_0/\Delta\Phi_0\) respectively where \(\Delta\Phi_0\) is the maximum nonlinearly-induced phase distortion. The latter requirement assures “external self-action” and simply states that the effective focal length of the induced nonlinear lens in the sample should be much larger than the sample thickness itself. In most experiments using the Z-scan technique we find that this second criterion is automatically met since \(\Delta\Phi_0\) is small. In Section 4 we will analyze thick sample Z-scans (\(L>Z_0\)) and show that for phase distortions that are small enough, simple expressions can still be derived for the Z-scan transmittance. Additionally we will show that for the sample to be safely regarded as “thin”, the first criterion for linear diffraction is more restrictive than it need be, and it is sufficient to replace it with \(L<n_0Z_0\).

The external self action limit simplifies the problem considerably, and the amplitude \(\sqrt{I}\) and phase \(\Delta\phi\) of the electric field \(E\) are now governed in the SVEA by the following pair of simple equations:

\[
\frac{d\Delta\phi}{dz'} = \frac{2\pi}{\lambda} \Delta n(I) \quad (1)
\]

\[
\frac{dI}{dz'} = -\alpha(I)I, \quad (2)
\]

where \(z'\) is the propagation depth in the sample and \(\alpha(I)\) in general includes linear and NLA terms. Note that \(z'\) should not be confused with the sample position \(Z\).

For third-order nonlinearities we take:

\[
\Delta n = \left\{ \frac{n_2}{2} |E|^2 \right\}_{esu} = \left\{ n_2 I \right\}_{MKS} \quad (3)
\]

\[
\Delta \alpha = \beta I, \quad (4)
\]

where \(n_2\) is the nonlinear index of refraction, \(E\) is the peak electric field (cgs), and \(I\) denotes the irradiance (MKS) of the laser beam within the sample. Here \(\beta\) denotes the third-order nonlinear absorption coefficient, which for ultrafast NLA is equal to the two-photon absorption (2PA) coefficient. \(n_2(esu)\) and \(n_2(MKS)\) are related through the conversion formula, \(n_2(esu) = (cn_0/40\pi) n_2(MKS)\), where \(c\) (m/sec) is the speed of light in vacuum. We note, however, that while we are using \(n_2\) here for any third-order nonlinearity, it may not be the best description of cumulative nonlinearities. These occur in, for example, reverse saturable absorbing (RSA) dyes. In such dyes linear absorption is followed by excited-state absorption (ESA) where the excited-state absorption cross
section is larger than the ground-state cross section. As the resulting change in absorption is best described by a cross section and not by a two-photon absorption coefficient, the index change, here due to population redistribution, is better described by refractive cross sections than by an \( n^2 \). Such an \( n^2 \) (or \( \beta \)) would change with the laser pulsewidth.\(^{15,16}\) This is discussed in more detail in Section 5.

Once the amplitude and the phase of the beam exiting the sample are known, the field distribution at the far-field aperture can be calculated using diffraction theory (Huygen’s principle). We will briefly review this procedure in Section 3 for a Gaussian beam. Simple analytical or empirical relations as obtained from those rigorous treatments are presented in this section. In most practical cases these relations present a convenient yet accurate method for estimating the nonlinear coefficients. In the remainder of this chapter \( n^2 \) always refers to \( n^2 \) (MKS).

### 2.1 Nonlinear Refraction without Nonlinear Absorption

We define the change in transmittance between the peak and valley in a Z-scan as \( \Delta T_{pv} = T_p - T_v \) where \( T_p \) and \( T_v \) are the normalized peak and valley transmittances as seen in Fig. 3. The empirically determined relation between the induced phase distortion, \( \Delta \Phi_0 \), and \( \Delta T_{pv} \) for a third-order nonlinear refractive process in the absence of NLA is,

\[
\Delta T_{pv} = 0.406(1 - S)^{0.27}\left|\Delta \Phi_0\right| \quad (5)
\]

where

\[
\Delta \Phi_0 = \frac{2\pi}{\lambda} n_2 I_0 L_{eff} \quad (6)
\]

with, \( L_{eff} = (1 - e^{-\alpha L})/\alpha \), and \( S \) is the transmittance of the aperture in the absence of a sample. \( \Delta \Phi_0 \) and \( I_0 \) are the on-axis (r=0), peak (t=0) nonlinear phase shift and the irradiance with the sample at focus (Z=0) respectively. The sign of \( \Delta \Phi_0 \) and hence \( n_2 \) is determined from the relative positions of the peak and valley with Z as shown in Fig. 3. This relation is accurate to within \( \pm 3\% \) for \( \Delta T_{pv} < 1 \). As an example, if the induced optical path length change due to the nonlinearity is \( \lambda/250 \), \( \Delta T_{pv} \approx 1\% \) for an aperture transmittance of \( S=0.4 \). Use of \( S=0.4 \) is a good compromise between having a large signal which averages possible beam nonuniformities.

The distance in Z between peak and valley, \( \Delta Z_{pv} \), is a direct measure of the diffraction length of the incident beam for a given order nonlinear response. In an standard Z-scan (i.e. using a Gaussian beam and a far-field aperture), this relation for a third-order nonlinearity is given by:

\[
|\Delta Z_{pv}| \approx 1.7 Z_0 \quad (7)
\]

This can be extremely useful since it gives the focal spot size of the beam for diffraction limited optics independent of the irradiance for small nonlinearities. In principle the Z-scan can be used to measure very small spot sizes by using very thin samples. For small \( \Delta \Phi_0 \), peak and valley are equidistant (\( \approx \pm 0.856 Z_0 \)) from the focus (Z=0). As \( \Delta \Phi_0 \)
increases, the peak and valley positions do not remain symmetric; the valley moving toward focus and the peak away so that $\Delta Z_{pv}$ remains nearly constant as given above. We must re-emphasize that the above relation is valid only for closed-aperture $Z$-scans involving an $n_2$-type nonlinearity, a good quality Gaussian beam ($M^2 \approx 1$), and thin nonlinear samples. Any departure from these conditions will give rise to a different characteristic $\Delta Z_{pv}$. Later on, we will briefly discuss cases involving thick samples, $\chi^{(5)}$-type nonlinearities, eclipsing and top-hat-beam $Z$-scans.

The linear relationship between $\Delta T_{pv}$ and $\Delta \Phi_0$ makes it convenient to include a time averaging factor which is not included in Eqs. 5 and 6 for pulsed inputs. Inclusion of this temporal averaging reduces the measured $\Delta T_{pv}$ by a factor, $A_T$, which generally depends on the pulse shape and the response time of the nonlinearity. For nonlinearities with response times much shorter than the pulsewidth (i.e. instantaneous nonlinearities), $A_T$ is given by:  

$$A_T = \frac{\int_{-\infty}^{+\infty} f^2(t)dt}{\int_{-\infty}^{+\infty} f(t)dt}$$

$$\text{(8)}$$

where $f(t)$ denotes the dimensionless temporal profile of the incident laser pulse. For a Gaussian temporal shape, this gives $A_T = 1/\sqrt{2}$ while a Sech$^2$ pulse gives $A_T = 2/3$. In the other extreme, where the time response of the nonlinearity is much larger then the pulsewidth, $A_T$ assumes a value of 1/2 for a third-order nonlinearity, independent of the pulse shape. Of course, in this case, the interpretation of $n_2$ changes. For example, in the case of reverse-saturable absorbers and under the approximations discussed in Ref. 16, $n_2 I_0$ is replaced by $\sigma_r F / 2 \hbar \omega$, where $\sigma_r$ is the excited-state refractive cross section and $F$ is the fluence. Cases involving higher-order nonlinearities, and/or with response times that are comparable to the pulsewidth, require proper averaging of $\Delta \Phi_0(t)$ according to Eq. 8, and will not be discussed here.

2.2 Higher Order Nonlinearities:

Although many observed nonlinear optical effects give index changes proportional to the irradiance ($\Delta n \propto I$), we often encounter higher order effects where $\Delta n \propto I^\eta$, with $\eta > 1$. For example, a fifth order NLR, ($\chi^{(5)}$-type nonlinearity where $\eta = 2$) becomes the dominant mechanism in semiconductors when $\Delta n$ is induced by two-photon generated free-carriers. 17 For this type of nonlinearity, where $\Delta n = n_4 I^2$ is assumed, we can derive simple relations that accurately characterize the $Z$-scan data. For a Gaussian beam and far-field aperture, these are given by:

$$\Delta T_{pv} \approx 0.21(1 - S)^{0.27}[\Delta \Phi_0]$$

$$\text{(9)}$$

and

$$|\Delta Z_{pv}| \approx 1.2 Z_0$$

$$\text{(10)}$$
where $\Delta \Phi_0 = kn_0 L'_{\text{eff}}$ with $L'_{\text{eff}} = \frac{[1 - \exp(-2\alpha L)]}{2\alpha}$. In certain cases where competing $\chi^{(3)}$ and $\chi^{(5)}$ processes are simultaneously involved, the data analysis becomes more complicated. In Ref. 17 a procedure is given for separating the two processes using a number of Z-scans at different irradiances. This procedure makes use of simple relations of Eqs. 5 and 9 to estimate the nonlinear coefficients associated with both $\chi^{(3)}$ and $\chi^{(5)}$ processes.

### 2.3 Eclipsing Z-scan (EZ-Scan)

As the Z-scan method relies on propagation of a phase distortion to produce a transmittance change, the minimum detectable signal is determined by how small a transmittance change can be measured. The surprising interferometric sensitivity comes about from the interference (diffraction) of different portions of the spatial profile in the far field. Recently, it was realized that this sensitivity could be greatly increased by looking at the outer edges of the beam in the far field rather than the central portion as in the Z-scan. This is accomplished by replacing the apertures in Fig. 2 with disks that block the central part of the beam. The light that leaks around the edges appears as an eclipse, thus the name EZ-scan for eclipsing Z-scan. An analogous empirical expression to Eq. 9 for the EZ-scan is

$$\Delta \Delta \Phi_{\text{EZ}} \cong -0.68(1 - S)^{-0.44}[\Delta \Phi_0],$$

which is accurate to within $\pm 3\%$ for $|\Delta \Phi_0| \leq 0.2$ and a disk linear transmittance rejection $S$ in the range $0.98 > S > 0.995$, i.e. the fraction of light seen by the detector is $1 - S$.

![Figure 4. A comparison of EZ-scan data ($S=0.98$, solid line) and Z-scan data ($S=0.02$, dashed line) for a self-focusing nonlinearity with a phase shift of 0.1.](image)

Figure 4 shows a comparison of an EZ-scan with a Z-scan for a phase distortion of 0.1 radian. The relative positions of peak and valley switch from the Z-scan since light that is transmitted by an aperture is now blocked by the disk and vice versa.
above relation, as $S \rightarrow 1$ (large disks), the sensitivity increases significantly. Sensitivities to optical path length changes of $\pm \lambda / 10^4$ have been demonstrated as compared to $\pm \lambda / 10^3$ for Z-scan. For the range of $S$ given above, the spacing between peak and valley, $\Delta Z_{pv}$, is empirically found to be given by $\Delta Z_{pv} \approx 0.9 - 1.0 Z_0$, which grows to the Z-scan value of $\approx 1.7 Z_0$ as $S \rightarrow 0$. The enhancement of sensitivity in the EZ-scan, however, comes at the expense of signal photons as well as a reduction in accuracy and absolute calibration capability. This added uncertainty originates from the deviations of the actual laser beams from a Gaussian distribution, and the fact that we need to know $S$ very accurately. We, therefore, recommend using this technique only when the added sensitivity is required and with a known reference sample to calibrate the system.

2.4 Nonlinear Absorption

While NLA can be determined using a two parameter fit to a closed aperture Z-scan (i.e. fitting for both $\Delta n$ and $\Delta \alpha$), it is more directly (and more accurately) determined in an open aperture Z-scan. For small third-order nonlinear losses, i.e. $\Delta \alpha L = \beta I L_{eff} << 1$ with response times much less than the pulsewidth (e.g. two-photon absorption), and for a Gaussian temporal shape pulse, the normalized change in transmitted energy $\Delta T(Z) = T(Z) - 1$, becomes

$$
\Delta T(z) \approx -\frac{q_0}{2\sqrt{2}} \left\{ \frac{1}{1 + Z^2/Z_0^2} \right\},
$$

where $q_0 = \beta I L_{eff}$ ($q_0 << 1$). This mimics the Lorentzian distribution of the irradiance with $Z$ for a focused Gaussian beam as seen for the dashed line in Fig. 5. If the response time of the material is much longer than the pulsewidth used, the factor $2\sqrt{2}$ is replaced by $2$. This is independent of the temporal pulse shape. Of course, in this case, the interpretation of $\beta$ changes. For example, in the case of reverse-saturable absorbers and under the approximations discussed in Ref. 16, $\beta I_0$ is replaced by $\sigma F / h \omega$, where $\sigma$ is the excited-state absorption cross section.

2.5 Nonlinear Refraction in the Presence of Nonlinear Absorption

We can also determine NLR in the presence of NLA. As mentioned, this can be done by fitting the Z-scan with a two parameter fit or by separately measuring the NLA in a Z-scan performed with the aperture removed (i.e. open aperture Z-scan). This second method is more accurate since two single parameter fits give a higher accuracy than one two parameter fit. Within approximations elaborated in Ref. 2 (primarily that the Z-scan is not dominated by nonlinear absorption) a simple division of the curves obtained from the two Z-scans (closed/open) gives a curve that closely approximates what would be obtained with a closed aperture Z-scan on a material having the same $\Delta n$ but with $\Delta \alpha = 0$. This greatly simplifies determining $\Delta n$. An example of this division process is shown in Fig. 5. In lieu of this division, with $\Delta \alpha$ known from the open aperture results, the Z-scan with aperture in place ($S < 1$) can be used to extract the remaining unknown, namely $\Delta n$ by a direct fitting as described later.17
Figure 5. Calculations of closed (solid line, $S=0.5$) and open (dashed line) aperture Z-scan data along with their ratio (dotted line) for self-defocusing ($\Delta\Phi_0=-0.5$) accompanied by nonlinear absorption ($q_0=0.4$). The Lorentzian shaped curve (dashed line) is the result of an open aperture Z-scan.

2.6 Pump-Probe Z-scans

Pump-probe techniques (also referred to as excite-probe techniques) in nonlinear optics have been commonly employed in the past to deduce information that is not accessible with a single beam geometry. The most significant application of such techniques concerns the ultrafast dynamics of the nonlinear optical phenomena. There has been a number of investigations that have used Z-scan in a pump-probe scheme.$^{8,9,10,11}$ The general geometry is shown in Fig. 6 where collinearly propagating pump and probe beams are used. After propagation through the sample, the probe beam is then separated and analyzed through the far-field aperture. Due to collinear propagation of the pump and probe beams, we are able to separate them only if they differ in wavelength or polarization. The former scheme, known as a 2-color Z-scan, has been used to measure the nondegenerate $n_2$ and $\beta$ in semiconductors.$^{8,9}$ The time-resolved studies can be performed in two fashions. In one scheme, Z-scans are performed at various fixed delays between pump and probe pulses. In the second scheme, the sample position is fixed (e.g. at the peak or the valley positions) while the transmittance of the probe is measured as the delay between the two pulses is varied. The analysis of the 2-color Z-scans is naturally more involved than that of a single beam Z-scan. The measured signal, in addition to being dependent on the parameters discussed for the single beam geometry, will also depend on parameters such as the pump-probe beam waist ratio, pulsewidth ratio and the possible focal separation due to chromatic aberration of the lens.$^{8,10}$
2.7 Non-Gaussian Beams

While Gaussian beams are extremely convenient since their propagation is particularly simple (e.g. a Gaussian beam remains Gaussian throughout a linear optical system in the absence of aberrations), the output of many lasers do not possess a Gaussian profile in space. Zhao and Palffy-Muhoray\(^{12}\) derived the results of performing a Z-scan using a focused “top-hat” beam, where the profile at the initial focusing lens is approximately a step function (Heaviside function) in the radial coordinate \(r\) (i.e. \(\Theta(r_0-r)\) with \(r_0\) a constant). In practice, one can produce this type of beam profile by sufficiently expanding any spatially coherent optical beam and then use a circular aperture at the focusing lens. The lens focuses this beam to an Airy pattern in the absence of aberrations. The empirical expression relating \(\Delta T_{pv}\) to \(\Delta \Phi_0\) and aperture transmittance \(S\) is given by:\(^{12}\)

\[
\Delta T_{pv} \approx 2.8(1-S)^{1.14} \tanh(0.37\Delta \Phi_0),
\]

where \(\Delta \Phi_0\) is the peak nonlinear phase shift at the center of the Airy disk at the focal plane. For \(S=0\) and small \(\Delta \Phi_0\), the above expression gives \(\Delta T_{pv} \approx 1.036\Delta \Phi_0\), indicating approximately a 2.5 times larger sensitivity than for a Gaussian beam Z-scan. This enhanced sensitivity is due to the steeper beam curvature gradients encountered by the nonlinear sample at \(Z\) positions near the focal plane.

It is also possible to use a sample of known nonlinearity as a reference to calibrate a system using a beam of arbitrary profile. Reference\(^{18}\) shows a way to use a reference sample to obtain the relative NLA and NLR without regard to the laser beam characteristics. This also allows violation of the thin sample approximation as long as the reference sample has the same thickness as the sample under measurement, and the irradiance is adjusted such that the \(\Delta T_{pv}\)'s in both measurements are nearly equal. More generally, Z-scans using reference sample calibration are useful provided that the orders
of both nonlinearities are the same (e.g. both are $\chi^{(3)}$ type) and conditions and parameters of both experiments are kept nearly the same.

### 2.8 Background Subtraction

In all the possible situations discussed above, it is often beneficial to perform experiments at high and low irradiance levels (low enough that the nonlinear response is negligible) and subtract the two sets of data. This greatly reduces background signals due, for example, to sample inhomogeneities or sample wedge. A necessary condition for this background subtraction process to be effective is that the sample position be reproducible for both high and low irradiance scans (i.e. laterally, vertically and along $Z$). It is also important that the data sets be normalized before subtraction such that $T(|Z|>>Z_0)$ are made equal for high and low irradiance $Z$-scans. Experience shows that even when the signal is indistinguishable within a background that this subtraction can often uncover a usable signal.

### 3. Thin Nonlinear Medium Analysis

While the above analysis gives the dependence of NLA on the sample position $Z$, the analysis for NLR was restricted to $\Delta T_{pv}$. The $Z$ dependence for NLR can be obtained by straightforward numerical techniques as outlined below. We find it useful to fit the full $Z$ dependence of the $Z$-scan signals since there is information regarding the order of the nonlinearity in this $Z$ dependence as discussed in Section 1.

The irradiance distribution and phase shift of the beam at the exit surface of a sample exhibiting a third-order nonlinear refractive index are obtained by simultaneously solving Eqs. 1 and 2:

$$I(Z,r,t) = \frac{I(Z,r,t) \exp^{-at}}{1 + q(Z,r,t)}, \quad (14)$$

and

$$\Delta \phi(Z,r,t) = \frac{kn_2}{\beta} \ln[1 + q(Z,r,t)] \quad (15)$$

where $q(Z,r,t) = \beta(I(Z,r,t)L_{eff})$. Combining Eqs. 14 and 15 we obtain the complex field at the exit surface of the sample to be:

$$E_c = E(Z,r,t)e^{-at/2} (1 + q)^{(ikn_2/\beta)^{-1/2}}, \quad (16)$$

where $E(Z,r,t)$ is the incident electric field. The reflection losses can be safely assumed to be linear and hence will be ignored in this formalism. In evaluating the nonlinear coefficients, however, one should account for reflection loss of the first surface by taking the irradiance (i.e. $I_0$) to be that inside the sample.

In general for radially symmetric systems, a zeroth order Hankel transform of Eq. 16 will give the field distribution $E_a$ at the aperture which is placed a distance $d$ from the focal plane:
\[ E_a(Z, r, t, d) = \frac{2\pi}{i\lambda d'} \exp\left( \frac{i\pi r^2}{\lambda d'} \right) \int_0^\infty r'dr' \ E_e(Z, r', t - d'/c) \exp\left( \frac{i\pi r'^2}{\lambda d'} \right) J_0\left( \frac{2\pi rr'}{\lambda d'} \right) \]  

(17).

where \( d' = d - Z \) is the distance from the sample to the aperture plane. The measured quantity is the pulse energy or average power transmitted through the far-field aperture having a radius of \( r_a \). The normalized transmittance is then obtained as:

\[ T(Z) = \frac{\int_{-\infty}^{\infty} dt \int_0^{r_a} |E_a(Z, r, t, d)|^2 r dr}{U} \]  

(18),

where \( U \) is the same as the numerator but in the linear regime (i.e. for \( I=0 \)). In the case of an EZ-scan, the limits of the spatial integral in Eq. 18 must be replaced by \( r_d \) to \( \infty \) where \( r_d \) is the radius of the obscuration disk. It is generally more convenient to represent the aperture (or disk) size by the normalized transmittance (or rejection) \( S \) in the linear regime.

The formalism thus far presented is generally applicable to any radially symmetric beam. Here, however, we assume a TEM\(_{0,0}\) Gaussian distribution for the incident beam as given by:

\[ E(Z, r, t) = E_0(t) \frac{w_0}{w(Z)} \exp\left( -\frac{r^2}{w^2(Z)} + i\frac{\pi r^2}{\lambda R(Z)} + i\phi \right), \]  

(19)

where \( w(Z) = w_0(1+Z^2/Z_0^2)^{1/2} \) and \( R(Z) = Z+Z_0^2/Z \). The radially invariant phase terms, contained in \( \phi \), are immaterial to our calculations and hence will be ignored.

The integral in Eq. 17 can be analytically evaluated if we assume that \( |q|<1 \) (in Eq. 16) and then perform a binomial series expansion of \( E_e \) in powers of \( q \). Recalling that \( q I \propto \exp(-r^2/w^2) \), this expansion effectively decomposes \( E_e \) into a sum of Gaussian beams with varying beam parameters. This method of beam propagation known as Gaussian decomposition was first given by Wearie et. al. Following the expansion, we obtain:

\[ E_e = E(Z, r, t)e^{-ao/2} \sum_{m=0}^{\infty} F_m \exp(2m r^2 / w^2(Z)), \]  

(20)

where the \( F_m \), the factor containing the nonlinear optical coefficients, is given by:

\[ F_m = \frac{(i\Delta\phi_0(Z, t))m}{m!} \prod_{j=1}^{m} \left[ 1 + i\left( j - \frac{1}{2} \right) \frac{\lambda\beta}{2\pi m_2} \right], \]  

(21)

with \( F_0=1 \) and \( \Delta\phi_0(Z, t) = \Delta\phi(Z, r=0, t) \) denoting the on-axis instantaneous nonlinear phase shift. The Hankel transform of \( E_e \) will then give the field at the aperture plane as a sum of Gaussian beams:
\[ E_a(r, t) = E(Z, r = 0, t)e^{-\alpha d^2} \sum_{m=0}^{\infty} F_m \frac{W_{m0}}{w_m} \exp \left\{ -\frac{r^2}{w_m^2} + \frac{im^2}{\lambda R_m} + i\theta_m \right\}, \quad (22) \]

where the beam parameters of each Gaussian beam are as follows:

\[ w_m^2 = \frac{w^2(Z)}{2m+1}, \quad d_m = \frac{kw_m^0}{2}, \quad w_m^2 = \frac{g^2 + d^2}{d_m^2}, \]

\[ R_m = d \left[ 1 - \frac{g}{g^2 + d^2 / d_m^2} \right]^{-1}, \text{ and } \theta_m = \tan^{-1} \left[ \frac{g}{d / d_m} \right] \text{, where} \]

\[ g = 1 + d / R(Z) \quad (23) \]

Finally, the normalized transmittance can then be evaluated as given by Eq. 18. We should note here that with an incident Gaussian beam, the aperture transmittance can be given as \( S=1 -\exp\left(-2ra^2/wa^2\right) \) where \( wa^2=w_0(1+d^2/Z_0^2)^{1/2} \) is the linear beam radius at the aperture.

It is worthwhile to analyze the implications of the above results under a number of further approximations. In the absence of nonlinear absorption (i.e. \( \beta=0 \)), \( F_m = (i\Delta \phi(Z, t))^m / m! \) and the far-field beam deformation will be as a result of external self-action (self-focusing and self-defocusing). In that case we can write \( \Delta \phi_0 \) as:

\[ \Delta \phi_0 = \Delta \phi_0(0,0) \text{ is the peak-on-axis nonlinear phase shift, and } f(t) \text{ represents the irradiance temporal profile of the incident pulse.} \]

\[ \Delta \phi_0(Z, t) = \frac{2\pi}{\lambda} n_2 I(Z, t) = \Delta \Phi_0 \frac{f(t)}{1 + Z^2 / Z_0^2}, \quad (24) \]

where \( \Delta \Phi_0 = \Delta \phi(0,0) \) is the peak-on-axis nonlinear phase shift, and \( f(t) \) represents the irradiance temporal profile of the incident pulse. We find that this Gaussian decomposition method is useful for the small phase distortions detected with the Z-scan (or EZ-scan) method since only a few terms of the sum in Eq. 22 are needed. Figure 3 depicts calculated Z-scans for \( \Delta \Phi_0 = \pm 0.5 \) using the above formalism. The simple relations (Eq. 9 and 11) given earlier were obtained by empirically fitting the calculated results using the equations derived in this section. As was shown in Ref. 2, such empirical relations are exact in the limit of small \( \Delta \Phi_0 \) where only one nonlinear term in the expansion (Eq. 22) is retained.

With NLA present (i.e. \( \beta\neq0 \)), although no restriction is imposed on the magnitude of \( \Delta \Phi_0 \), the above formalism is valid only for \( q_0 = |\beta I_0 L_{\text{eff}}| < 1 \). Note that the coupling factor \( \lambda \beta / 2m_2 = q_0 / \Delta \Phi_0 \) in Eq. 21 is twice the ratio of the imaginary to real parts of the third-order nonlinear susceptibility, \( \chi^{(3)} \) (i.e. \( q_0 / \Delta \Phi_0 = \operatorname{Im}\{\chi^{(3)}\} / \operatorname{Re}\{\chi^{(3)}\} \)). The 2PA figure-of-merit (FOM) for all-optical switching has been defined as \( 4\pi \times \text{this value} \). Since the irradiance and effective length cancel in this ratio, this FOM can be deduced for third-order nonlinearities without knowledge of the irradiance or sample length as long as the thin sample approximation is valid.

The Z-scan transmittance variations can be calculated following the same procedure as described previously. As is evident from Eqs. 21-22, the absorptive and refractive contributions to the far field beam profile and hence to the Z-scan transmittance are coupled. When the aperture is removed, however, the Z-scan transmittance is insensitive
to beam distortion and is only a function of the nonlinear absorption. The total transmitted fluence (i.e. energy per unit area) in that case (S=1) can be obtained by spatially integrating Eq. 14 without having to include the free space propagation process. The resultant normalized transmittance for a pulse with a temporal profile \( f(t) \) can then be derived as:

\[
T(Z,S=1) = \frac{1 + Z^2 / Z_0^2}{q_0} \left[ \ln \left( \frac{1 + q_0 f(\tau)}{1 + Z^2 / Z_0^2} \right) \right] \int_{-\infty}^{+\infty} f(\tau) d\tau.
\] (25)

For \( |q_0| < 1 \), this transmittance can be expressed in terms of the peak irradiance in a summation form more suitable for numerical evaluation. Assuming a Gaussian temporal profile (i.e. \( f(\tau) = \exp(-\tau^2) \)) this can be written as:

\[
T(Z,S=1) = \sum_{m=0}^{\infty} \frac{(-q_0)^m}{(1 + Z^2 / Z_0^2)^m (m + 1)^{3/2}}.
\] (26)

Thus, once an open aperture (S=1) Z-scan is performed, \( \beta \) can be unambiguously deduced. With \( \beta \) known, the Z-scan with aperture in place (S<1) can be used to extract the remaining unknown, namely the coefficient \( n_2 \). Note that \( \Delta T \), as given by Eq. 12 in Section 2, is simply the \( m=1 \) term in Eq. 26.

### 4. “Thick” samples

It is apparent from relations derived so far, that a way to obtain larger Z-scan signals (\( \Delta T_{pv} \)) is to increase \( \Delta \Phi_0 \) through either stronger focusing (shorter \( Z_0 \)) or thicker samples (larger \( L \)). In either case, we recall that the validity of these relations becomes questionable once the thin sample criterion (\( L \ll n_0 Z_0 \)) is violated. In this section we address this problem and analyze a general case in which no limitation is imposed on the sample length. The rigorous treatment of this problem involves numerical solutions to nonlinear wave equations and will not be discussed here. In addition to numerical calculations, two types of approximate solutions, resulting in simple relations, have been reported. One involves an “aberration-free” approximation of the nonlinear wave equation, and the other treats the wave propagation exactly to first order in the nonlinear phase shift (\( \Delta \phi \)). The latter approach requires that \( \Delta \phi \) is small enough that no nonlinear beam distortion (self-action) occurs within the sample although linear diffraction does occur. This condition is controllable and can be satisfied in an experiment. In fact, often being faced with this limitation (low \( \Delta n \)), is the very reason that one resorts to thick sample conditions.

Following Hermann and McDuff\(^4\), and Tian et al.\(^5\), the on-axis (S=0) Z-scan transmittance of a thick nonlinear sample can be written as:

\[
T(x) \approx 1 + \Delta \Phi_{z_0} F(x, l),
\] (27)

where \( \Delta \Phi_{z_0} = (2\pi/\lambda)n_2 l_0 Z_0 \) is the nonlinear phase shift occurring within one \( Z_0 \), \( x = Z/Z_0 \), and \( l = L/(n_0 Z_0) \) is the normalized length of the sample. \( F(x, l) \) is given by.\(^4\)
Plots of $F(x,l)$ are shown in Fig. 7 for $l=1,2,5,8$ and 10. The analysis here is strictly correct for $n_0=1$, but does not account for the longitudinal shift of the linear focus for $n_0>1$. Since no useful information (regarding the nonlinear optical measurement) exists in the absolute position of the transmittance peak and valley, this analysis is sufficient and simple. The position of peak and valley are obtained by evaluating $dF/dx=0$, which gives:

$$X_{p,v} = \pm \sqrt{\frac{(l^2/2-10)+(l^2+10)^2}{6}}.$$

The peak-valley separation, therefore, is given by $\Delta Z_{pv}=2|X_{p,v}|Z_0$. As evident from Fig. 8 which shows $L/(n_0Z_0)$ as a function of $L/(n_0Z_0)$, this separation approaches $L/n_0$ for $L>>n_0Z_0$. All the above relations reduce to that derived for a thin sample when we let $l-\to0$. Moreover, as one would expect, it is seen in Fig. 8 that by increasing the sample thickness above $\approx2n_0Z_0$, the signal ($\Delta T_{pv}$) gradually levels off and ultimately becomes a constant. A useful quantity that illustrates this effect, is the effective length of a thick

\[
F(x,l) = \frac{1}{4} \ln \left\{ \left( \frac{(x+l/2)^2+1}{(x-l/2)^2+1} \right) \left( \frac{(x-l/2)^2+9}{(x+l/2)^2+9} \right) \right\}. 
\]
nonlinear medium defined as the length that can be attributed to the sample if it were to be regarded as thin in data analysis. Once we identify such an effective length, we can use the “thin” sample relation (Eq. 5) to quickly evaluate the \( n_2 \) coefficient for third-order nonlinearities. We, thus, define this length as 

\[
\ell_{\text{eff}} = \frac{\Delta T_{\text{pv}}(\text{thick})}{(0.406 \Delta \Phi Z_0)^3}
\]

which is evaluated from the above expressions as:

\[
\ell_{\text{eff}} = \frac{F(|X_{\text{pv}}|, l)}{2 \times 0.406} \quad \text{(30)}
\]

This is plotted in Fig. 8 together with a simpler empirical fit given by:

\[
\ell_{\text{eff}} = \frac{2.706[(l + 1)^{1.44} - 1]}{l^{1.44} + 3.924} \quad \text{(31)}
\]

Figure 8. The effective sample length, \( \ell_{\text{eff}} \) (\( L_{\text{eff}} \) in units of \( n_0 Z_0 \)), as a function of \( l=L/n_0 Z_0 \) (solid line) and empirical fit to this line (dotted line). The dot-dashed line is a straight line of slope one.

5. Interpretation

There are many physical processes which can lead to third-order nonlinearities (i.e. effects proportional to the input irradiance, fluence or energy). Ultrafast nonlinear absorption processes include multiphoton absorption\(^{22,23}\), stimulated Raman scattering\(^{24}\).
and AC-Stark effects. These lead via causality and Kramers-Kronig relations to the bound-electronic nonlinear refractive index, $n_2$. Cumulative (i.e. slow) nonlinearities include population redistribution from linear absorption (this includes saturable and excited-state or reverse saturable absorption and their refractive counterparts), reorientation of anisotropic molecules such as in CS$_2$, thermal refraction, electrostriction, etc. The Z-scan is sensitive to all of these nonlinearities including higher-order effects and cannot simply be used by itself to distinguish these nonlinear processes or separate fast from slow nonlinearities.

A key to distinguishing these processes is to pay particular attention to the temporal response. Ultrafast nonlinearities are easily analyzed as has been discussed. The use of pulsewidths much shorter than the decay times of excited states allows such cumulative nonlinearities to be more easily analyzed. As we shall show, in this regime, the excited-state nonlinearities are fluence dependent, while the ultrafast effects remain irradiance dependent. The explicit temporal dependences of the nonlinearities can be obtained from, for example, degenerate four-wave mixing experiments or time resolved Z-scan experiments which can separate $\Delta n(t)$ and $\Delta \alpha (t)$. These time-resolved experiments can, in principle, separate slow and fast nonlinear responses. In addition, nondegenerate nonlinearities can be determined from pump-probe, four-wave mixing, or 2-color Z-scan experiments. These nondegenerate nonlinear responses are also useful in distinguishing various contributing nonlinear mechanisms.

We illustrate the potential problems associated with interpreting nonlinear measurements with a single example of comparing excited-state absorption (reverse saturable absorption) and two-photon absorption signals. The equation describing 2PA in the presence of residual linear absorption is:

$$\frac{dI}{dz} = -\alpha I - \beta I^2. \quad (32)$$

Excited states created by linear absorption in molecules are characterized by a

$$\Delta \alpha \approx \frac{\alpha \sigma}{\hbar \omega} \int I(t') dt'. \quad (15,16)$$

By temporally integrating the resulting equation for $dI/dz'$, we find the fluence $F$ (energy per unit area) varies with $z'$ as

$$\frac{dF}{dz'} = -\alpha F - \frac{\alpha \sigma}{2 \hbar \omega} F^2. \quad (33)$$

Notice that this equation is exactly analogous to the equation describing 2PA loss (Eq. 32) with the fluence replacing the irradiance and $\alpha \sigma/2 \hbar \omega$ replacing $\beta$. Therefore, since in most experiments the pulse energy is detected, excited-state absorption initiated by linear absorption and 2PA will give nearly identical results for loss as a function of input energy (microscopically ESA can be considered as the limit of 2PA with a resonant intermediate state). The difference between Eqs. 32 and 33 when determining the transmitted energy is in the temporal integral over the pulse for 2PA. For ESA this integral has already been performed. In other words, in order to determine which of these nonlinearities is present, the temporal dependence must be measured in some way. An analogous problem exists with excited-state refraction and the bound electronic $n_2$. Additionally, as seen in many semiconductors and in some organic materials, the excited states can be created by
nonlinear absorption (e.g. 2PA) leading to fifth-order nonlinear absorption and refraction, further confusing interpretation.

Equation 33 is only valid for low fluence where the changes in transmittance are small. For higher fluence, saturation of the ground state absorption process (or even the excited state absorption) can occur. In such cases the best approach is to solve the system of rate equations to determine the level populations and then use these in the loss equation (Eq. 34) in terms of absorption cross sections, $\sigma_{ij}$, or phase equation (Eq. 35) using refractive cross sections ($\sigma_{r})_{ij}$:

$$\frac{dI}{dz} = -\sum_{i=1, j>i}^{N} \sigma_{ij} \Delta N_{ij} I \quad (34)$$

$$\frac{d\phi}{dz} = \sum_{i=1, j>i}^{N} (\sigma_{r})_{ij} \Delta N_{ij} \quad (35)$$

where $\Delta N_{ij}$ is the population difference between two levels ($N_i - N_j$) coupled by an absorption cross section $\sigma_{ij}$. The nonlinear refraction is due to the redistribution of level populations and the sign depends on the frequency position with respect to the resonance frequency as well as on whether the loss increases or saturates. For many materials (e.g. organic reverse saturable absorbers) $\Delta N_{ij}$ can be replaced by the population of the lower level $N_i$, since the upper level rapidly decays to an intermediate level (e.g. in the vibration/rotation band). Temporal and spatial integrals of Eqs. 34 and 35 also need to be performed numerically. This procedure, of course, leads to Z-scans where the loss or refraction are not described by the third-order analysis given in this paper. This can also be said of the simple 2-level saturation model which is only described by a third-order response for small inputs.

It is important to note the importance of accurately measuring the laser mode and pulse parameters. For example, 2PA is irradiance dependent. Thus, given the pulse energy, we need to know both the beam area (i.e. spatial beam profile) and the temporal pulsewidth (i.e. temporal shape) in order to determine the irradiance. Any errors in the measurement of irradiance translate to errors in the determination of $\beta$ as well as several other nonlinear coefficients.

There are several other papers that report methods or analysis for Z-scans that we have not yet mentioned. For example, Herman et. al. discuss factors that affect optical limiting in thin samples with large nonlinearities related to Z-scans in Ref. 35. In Ref. 36, Hochbaum discusses the simultaneous determination of two or more nonlinear refractive constants in Z-scan measurements. Sutherland describes the effects of multiple internal reflections within a sample on the Z-scan signal in Ref. 37.

For some materials the light permanently or temporarily changes the optical properties so that the sample properties change within the duration of a Z-scan experiment. Oliveira et. al. discuss the analysis of such data. Petrov et.al. describe the use of a Z-scan in a reflection mode to determine changes of the complex dielectric function at surfaces. Kershaw describes his analysis of EZ-scan measurements in Ref. 40, and a method to enhance the sensitivity of a 2-color Z-scan is described in Ref. 41.
6. Conclusion

There are a several different methods and techniques for determining the nonlinear optical response of a material, each with its own weaknesses and advantages. It is advisable to use as many complementary techniques as possible over a wide range of input irradiance, power and fluence, using different pulsewidths and over a broad spectral range in order to unambiguously determine the active nonlinearities. Z-scan is one of the simpler experimental methods to employ and interpret. Despite the wide range of available methods, it is rare that any single experiment will completely determine the physical processes behind the nonlinear response of a given material. A single measurement of the nonlinear response of a material, at a single wavelength, irradiance and pulsewidth gives very little information on the material, and should not, in general, be used to judge the device performance of a material or to compare one material to another.

When measuring nonlinear material response, it is important to know the intended application. For example, if optical limiting with nanosecond pulses is the purpose, so that materials having large nonlinear loss for 10 ns pulses is desirable, the pulsewidth to be used is 10 ns. However, in order to determine the physics behind the nonlinear loss it may be useful to look at this loss using shorter or longer pulses.

Nonlinear absorption and refraction always coexist (although with different spectral properties) as they result from the same physical mechanisms. They are connected via dispersion relations similar to the usual Kramers-Kronig relations that connect linear absorption to the linear index, or equivalently, relate the real and imaginary parts of the linear susceptibility. The physical processes that give rise to NLA and the accompanying NLR include “ultrafast” bound electronic processes, “excited state” processes where the response times are dictated by the characteristic formation and decay times of the optically induced excited states, thermal refraction, etc. Ultrafast processes include multi-photon absorption, stimulated Raman scattering and AC-Stark effects. Excited-state nonlinearities can be caused by a variety of physical processes including absorption saturation, excited-state absorption in atoms or molecules or free-carrier absorption in solids, photochemical changes, as well as defect and color center formation. The above processes can lead to increased transmittance with increasing irradiance (e.g. saturation, Stark effect) or decreased transmittance (e.g. multiphoton absorption, excited-state absorption). A key to distinguishing these processes is to pay particular attention to the temporal response. One way of achieving this is the use of pulsewidths much shorter than the decay times of the excited states. In this regime, the excited-state nonlinearities are fluence (i.e. energy per unit area) dependent, while the ultrafast effects remain irradiance dependent.

The Z-scan has only recently been introduced. The use of the Z-scan technique as both an absolutely calibrated method for determining standards and as a relative measurement method is increasing. The database of nonlinear material properties is rapidly increasing, however, great care must still be employed when using numbers obtained from Z-scans as well as from other techniques primarily due to questions of data interpretation, e.g. is the response really third-order and ultrafast? The Z-scan signal as a function of irradiance and shape with respect to sample position, Z, can give useful information on the order of the nonlinearity as well as its sign and magnitude.
Acknowledgment

We gratefully acknowledge the support of the National Science Foundation over the past many years along with early support from the Defense Advanced Projects Research Agency and the Naval Air Warfare Center Joint Service Agile Program. The work presented in this paper represents many years of effort involving colleagues and many former and current students as well as post-doctoral fellows. We thank all those involved and acknowledge their contributions through the various referenced publications. We explicitly thank Edesly J. Canto-Said, J. Richard DeSalvo, Arthur Dogariu, David J. Hagan, Ali A. Said, Tai H. Wei, and Tiejun Xia for their many contributions.

References


