Beam propagation (M^2) measurement made as easy as it gets: the four-cuts method

Thomas F. Johnston, Jr.

Tolerance analysis shows that an efficient M^2 measurement plan is a first cut (beam diameter measurement) at 0.5–2.0 Rayleigh ranges to one side of the waist, which is matched by interpolation between second and third cuts to the opposite side. The waist is measured by a fourth cut halfway between the matched diameters, yielding an easy two-parameter curve fit for M^2. © 1998 Optical Society of America

OCIS codes: 120.4800, 140.3430, 350.5500.

1. Introduction

The beam propagation factor M^2 has become an often quoted parameter[1–3] in laser discussions since the introduction in the early 1990's of commercial instruments for its measurement. This factor is also called the beam quality or the times diffraction limit number,[1] or the ratio of the beam's divergence to that of a diffraction-limited beam of the same waist diameter. Yet obtaining accuracy in the determination of M^2 is surprisingly difficult because of a number of factors that are often overlooked, and many of these quoted values are at best only half a measurement. Often only the beam divergence is measured without confirming the beam's actual waist diameter and location, the substitution being made instead of values that are either calculated or supplied by the laser's manufacturer.

After designing[4] and using commercial beam propagation analyzers (the instruments that measure M^2) for several years and coming to appreciate the information they provide, I recently needed to accurately measure M^2 for a New Star Lasers (Auburn, Calif.) Model NS-600 laser, a pulsed Tm:Cr:Ho:YAG laser at 2.1 μm, in a laboratory where no commercial analyzer was conveniently available. Reproducible beam diameters could be determined by the traditional process of translating a knife edge (a razor blade) in small steps across the beam and repetitively pulsing the laser between steps to construct the knife-edge transmission function. This laser emits pulses of 2-J energy 10 ms apart in a 3-pulse burst, too brief a time interval for the laser rod to cool back to the quiesent state between pulses, giving the second and third pulse beams different propagation constants than the first pulse beam. Together with the two independent orthogonal propagation planes (the horizontal or X plane and the vertical or Y plane that contained the flash lamp that determined the symmetry axis of the beam) this meant there were actually six beams to characterize for this laser.

The traditional process is tedious and slow in contrast to the commercial instrument[4] (which collects 100 times as much information in 1/200 of the time). The data for each knife-edge diameter required 20 laser shots, or approximately 15 min, to acquire. The minimum number of diameter measurements for an M^2 determination is four (shown below) or requires an hour of data gathering (with a second hour for data analysis and reduction). The amount of work to be done on the six beams led to the question of how few a number of cuts (diameter measurements) can provide the beam propagation factor to a desired accuracy (how to make M^2 measurement as easy as it gets). This gave insight into why incorrect M^2 values are often quoted—what factors that are required for accuracy in measuring M^2 are often overlooked (highlighted in italics as they occur below). The minimum number of steps (illustrated on the NS-600 first pulse beam), the logic for those steps, and the tolerances required in their execution for an accurate M^2 measurement are discussed here.

When this research was done, the author was with Olenellus Engineering, 409 Pinetree Lane, Colfax, California 95713. He is now with Melles Griot, 2251 Rutherford Road, Carlsbad, California 92008.

Received 3 November 1997; revised manuscript received 10 March 1998.

0003-6935/98/214840-11$15.00/0

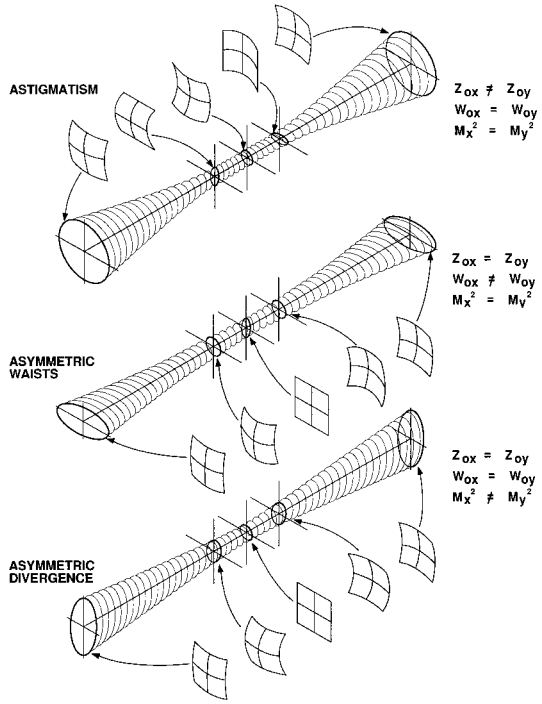

Fig. 1. Illustration of the three common types of beam asymmetry possible for a multimode beam. The window insets show the wave-front curvatures along the beam path. For the astigmatic beam there are two points with cylindrical wave fronts but none where the wave front is plane.

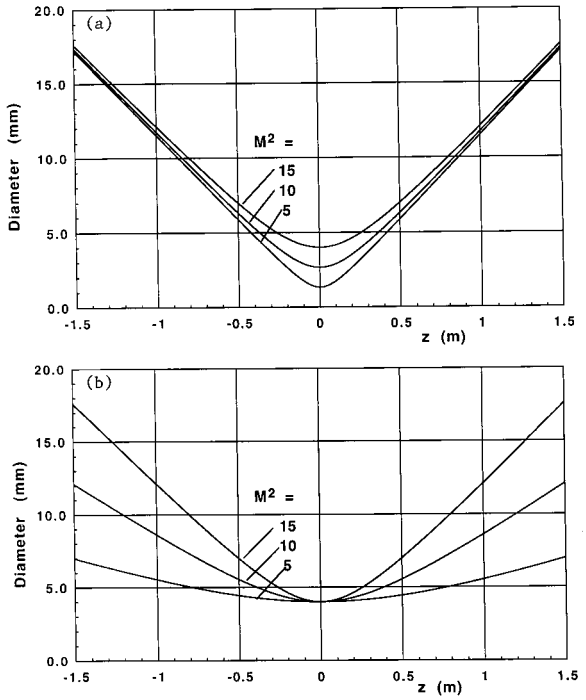

Fig. 2. Beams of (a) constant divergence and (b) constant waist diameter illustrate the consequence of $M^2 \neq 1$. The beam must be sampled both near and far from the waist to distinguish between these possibilities. The figures are drawn with values appropriate to the $\lambda = 2.1$-μm beam in the text.

2. M^2 Model

The usefulness of the model for a multimode laser beam (any real beam with $M^2 \geq 1$, as opposed to an idealized, perfect Gaussian beam with $M^2 = 1$) is that once M^2 is known, the propagation of the multimode beam can be analytically[5] described. The equations giving this description derive[5] from those for a Gaussian beam[6] from the fact that the real beam, of diameter $2W$ at propagation coordinate z from the waist (at coordinate z_0 and of diameter $2W_0$) is everywhere M times larger[5] than the embedded Gaussian. The embedded Gaussian beam is the fundamental mode that would be generated in the same resonator as the real beam. Here we use the convention[3,5] that uppercase quantities refer to the multimode beam and lowercase quantities to a Gaussian beam. Thus

$$2w_0 = 2W_0/M, \tag{1}$$

$$2W(z) = 2W_0\{[1 + (z - z_0)^2/{z_R}^2]\}^{1/2}. \tag{2}$$

The result in Eq. (2) is true because the multimode beam is made up of a superposition of Laguerre–Gauss resonator modes, all of which have the same waist location and Rayleigh range[1] or scale length for beam expansion with propagation of

$$z_R = \frac{\pi {w_0}^2}{\lambda} = \frac{\pi {W_0}^2}{M^2\lambda}. \tag{3}$$

Here λ is the laser wavelength. In propagating a distance z_R from the waist the multimode beam expands by a factor of $\sqrt{2}$, or the beam cross-sectional area doubles for a round beam. There is a set of three constants, $2W_0$, z_0, M^2, for each of the X and Y planes. As M^2 is conserved in propagation through any nonaberrating optical system,[7] for many systems the beam in the whole optical train is characterized once M^2 is known.

In the model above there are two independent propagation planes fixed in space (the X and Y planes). Although this covers most laser beams, it excludes those whose orthogonal axes rotate or twist around the propagation axis (termed beams with general astigmatism[8]) such as can be generated in nonplanar ring or out-of-plane folded resonators.

What are often overlooked when one first encounters this generalization of adding multimode beams to the familiar propagation analysis for Gaussian beams are the implications of the additional degree of freedom introduced with the factor M^2. The multimode beam divergence is given by the ratio of M^2 over $2W_0$ as

$$\Theta = (4\lambda/\pi)\frac{M^2}{2W_0} \tag{4}$$

and is no longer fixed by the inverse of the waist diameter alone, as it is for a Gaussian beam by

$$\theta = (4\lambda/\pi)\frac{1}{2w_0}. \tag{5}$$

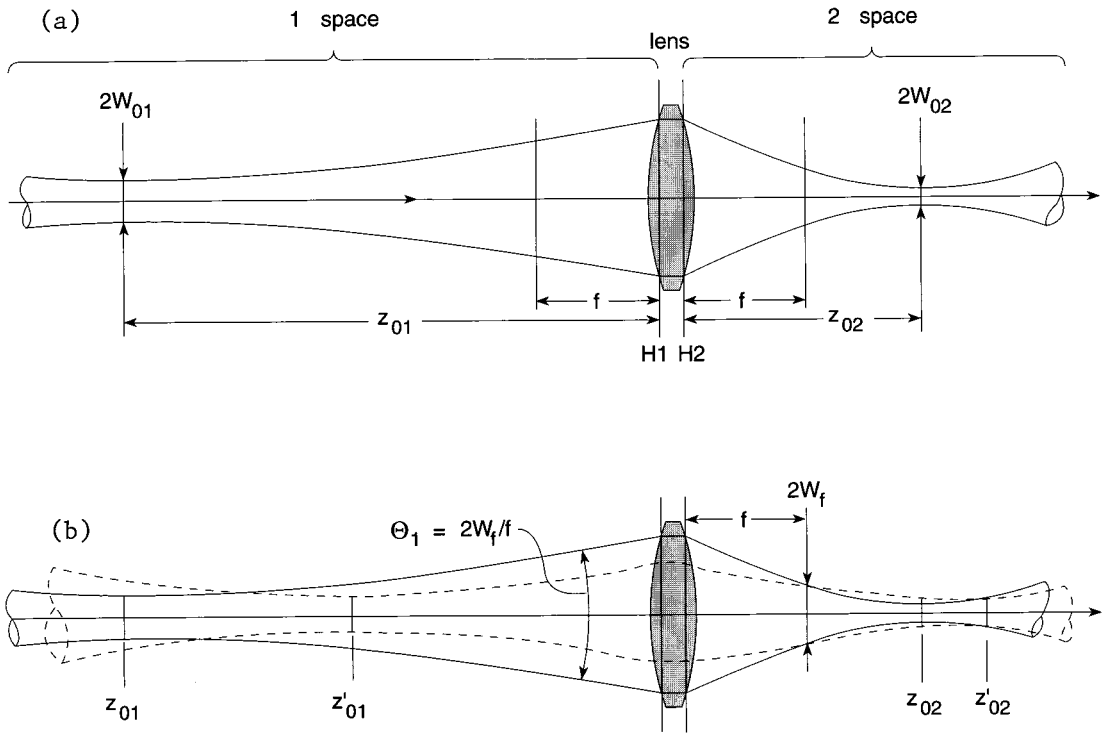

Fig. 3. (a) Definitions of the quantities in the transformation of a Gaussian beam through a lens. The principal planes of the lens are H1 and H2. (b) Illustration of the result that measurement of the beam diameter at the focal plane on the output side of the lens yields the beam divergence of the input beam, regardless of the location of the input waist.

Beams may be free of astigmatism, have the same waist diameters in the two propagation planes, yet be out of round in the far field—they simply have a higher mode order in one plane than the other (see Fig. 1, bottom example). With $M^2 > 1$, there are three types of beam asymmetry, no longer only two, with asymmetric divergence being added to the two Gaussian beam asymmetries of asymmetric waist diameters and astigmatism (Fig. 1).

In determining M^2 experimentally this requires that the beam waist diameter be measured directly (not inferred from a divergence measurement). In Fig. 2(a) several beams are plotted, all with the same divergence (and therefore M^2/W_0 = constant) but with differing M^2 values (and therefore the Rayleigh range z_R varies proportionally with W_0). This last proportionality comes from Eqs. (3) and (4) that give

$$z_R = 2W_0/\Theta. \tag{6}$$

From diameter measurements only at large distances from the waist, it would be impossible to distinguish between these curves and determine M^2.

Conversely, it would be impossible to distinguish between the curves of Fig. 2(b) from diameter measurements only near the waist and determine the beam divergence Θ. In Fig. 2(b) several beams are plotted, all with the same waist diameter but with differing values of M^2 [and therefore the divergence varies proportionally with M^2 by Eq. (4), and z_R varies inversely with M^2 by Eq. (3)]. What is called the normalizing Gaussian is defined by choosing the waist diameter and location equal to the multimode waist diameter and location, $2w_{0n} = 2W_0$ and $z_{0n} = Z_0$. This has a divergence given by Eq. (5), and the ratio of Eq. (4) to Eq. (5) then gives the simple expression for M^2 as the normalized divergence:

$$M^2 = \Theta/\theta_n. \tag{7}$$

To measure M^2, the beam diameter must be sampled both far from the waist (to determine Θ) and near the waist (to determine θ_n).

To locate the waist position, diameter measurements are performed on both sides of the waist to fix the position halfway between two equal diameters lying on opposite sides, which is the basis of the four-cuts method. To gain access to the beam on both sides of the waist, in general a lens must be inserted in the beam to form an auxiliary waist. The lens should be aberration free (typically, at $f/20$ or a smaller aperture) so as to not change the M^2 value of the transmitted beam. Measurements are made on that auxiliary beam, and the measured constants are transformed back through the lens to obtain the desired constants for the input beam. This was necessary for the 2.1-μm laser as the nominal waist location is at the output coupler, which excludes access to the beam on one side of the waist.

The lens transform equations[6] designate the space to the right behind the lens where the auxiliary waist is measured as 2 space (subscript 2) and the input beam on the left as 1 space (subscript 1) as shown in Fig. 3(a). The lens effective focal length is f, the

waist locations z_{01} and z_{02} are measured from the respective principal planes of the lens (with distances to the right as positive for z_{02} and distances to the left as positive for z_{01}). Then the lens transformation factor Γ is

$$\Gamma = f^2/[(z_{02} - f)^2 + {z_{R2}}^2], \tag{8}$$

$$2W_{01} = \Gamma^{1/2}(2W_{02}), \quad z_{R1} = \Gamma z_{R2},$$
$$z_{01} - f = \Gamma(z_{02} - f). \tag{9}$$

This is the notation used in the ModeMaster manual (available[4] from Coherent, Inc.). Equivalent expressions are derived in Ref. 9 in a slightly different notation by use of a parameter α, where $\alpha^2 = \Gamma$.

An immediate application of the above results is to demonstrate that the input beam divergence Θ_1 can be measured by finding the beam diameter $2W_f$ at precisely one focal length f behind the lens (in 2 space) from

$$\Theta_1 = 2W_f/f. \tag{10}$$

The proof follows by computing in 2 space $2W_f = 2W(z_2 = f)$ from Eq. (2) and using Eqs. (6), (8), and (9):

$$2W_f = 2W_{02}\{[1 + (f - z_{02})^2/{z_{R2}}^2]\}^{1/2}$$
$$= 2W_{02}\left(\frac{f}{z_{R2}}\right)\left(\frac{1}{\Gamma^{1/2}}\right)$$
$$= 2W_{01}\left(\frac{f}{z_{R2}}\right)\left(\frac{1}{\Gamma}\right) = 2W_{01}\left(\frac{f}{z_{R1}}\right) = \Theta_1 f,$$

which is Eq. (10). As shown in Fig. 3(b), as the input waist location varies, the lens transform equations operate to keep the output beam diameter at one focal length from the lens constant at the value $\Theta_1 f$ (or in more abstruse terms, the image at the focal plane of the lens is always the Fourier transform of the far-field angular spread of the beam).

3. Multimode Beam Diameters

Unlike Gaussian beams in which the $1/e^2$ beam diameter definition is universally used and understood, in the past a variety of different diameter definitions[5] have been employed. (Fortunately, all multimode diameters at least reduce to the Gaussian $1/e^2$ diameter for an $M^2 = 1$ beam.) Different definitions give different numerical values for the beam diameter, and consequently for M^2 it is like having different currencies for different countries and one has to know the exchange rate. Since the recommendation[10] by the committee on beam widths of the International Standards Organization (ISO) was made to standarize on the second-moment diameter definition, there has been growing agreement to do so. This diameter definition (symbol $D_{4\sigma}$) gives the beam width as four times the standard deviation of the transverse irradiance distribution, as sampled by the transmission through a pinhole translated across the beam. The second-moment diameter is the choice with the best analytical and theoretical support, but it is sensitive to noise in the wings of the distribution in its experimental evaluation. Careful checks on the effects of the noise that is present in the measured distribution are necessary to trust the experimental answer. Therefore the older methods will likely persist, and the strategy is to use these experimentally more forgiving methods along the propagation axis for most of the data gathering. Then at one plane in the far field of the beam (away from diffractive overlay from the mode-selecting aperture of the laser) one should carefully measure the ratio of diameters to establish the conversion to second moments.

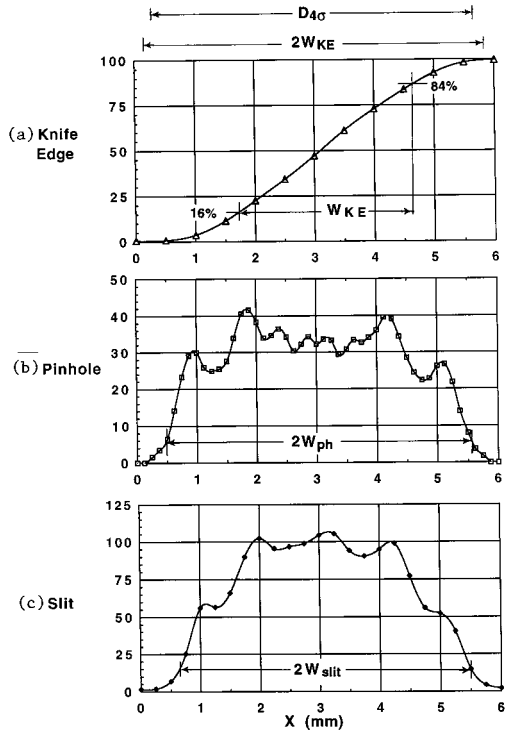

Fig. 4. Profiles of the 2.1-μm beam taken in the focal plane of the lens for different apertures translated across the beam. (a) The transmission (percent) past a knife edge, the 100% level is 2 J. (b) The transmitted energy (millijoules) through a pinhole. (c) The transmitted energy (millijoules) through a slit.

This has been done here with the plane chosen to intercompare diameters at one focal length (52.3 cm) behind the high-quality silica lens (Newport SPX031) forming the auxiliary waist. With Eq. (10) data taken at this plane, we can obtain an independent cross check against our final results on the input beam divergence. The beam diameter was measured by four different methods on the horizontal axis (X plane). The lens was 39 cm from the laser output coupler and the laser energy set to a nominal 2-J energy output per pulse at a manually triggered pulse repetition rate of less than 0.5 Hz.

The results are shown in Fig. 4. The knife-edge diameter (a) is defined as twice the edge translation distance between the 15.9 and 84.1% transmission points (this reduces to the $1/e^2$ diameter for a Gaussian input). This diameter is the largest and is mea-

Table 1. Conversions between Diameter Definitions for the First Pulse 2.1-μm Beam

Definition	Knife Edge	Second Moment	Pinhole	Slit
Diameter $2W_f$ (mm)	5.67	5.37	5.11	4.86
Divergence Θ_1 (mrad)	10.8	10.3	9.76	9.29
Relative divergence	1.06	1	0.952	0.905
Relative M^2	1.11	1	0.906	0.819

sured at high signal-to-noise ratio as the full beam energy is detected at maximum transmission.

The slot or pinhole profile in Fig. 4(b) gives a smaller diameter and shows the ring structure of this high-M^2 cylindrically symmetric mode. (The slot is a pinhole lengthened perpendicularly to the translation direction.) The pinhole diameter can be defined by one finding the highest (100%) point of the profile and then taking the profile width between the clip points at $1/e^2 = 13.5\%$ down from this. The slot was elongated vertically approximately six times its width (of 0.21 mm) to increase the detected signal-to-noise ratio. Nevertheless, the peak energy was only approximately 42 mJ (compared with 2 J detected in the first method). The pinhole profile is sensitive to centering errors (the centerline through the circular beam spot must be found by X scanning at several Y heights to find the maximum horizontal diameter).

The second-moment diameter is computed from the standard deviation of the pinhole transmission distribution. For the data plotted in Fig. 4(b), this gives $D_{4\sigma} = 5.37$ mm.

The narrowest diameter is obtained with the slit profile in Fig. 4(c), given as the distance between 13.5% clip points. The slit width here was 0.21 mm. This profile is not subject to centering errors and gave a peak energy of 106 mJ, but the profile gives a less intuitive picture of the mode structure.

Dividing these diameters by $f = 52.3$ cm gives the listed divergence of the input beam in the different diameter currencies in Table 1. For each diameter definition there is also listed the relative divergence (ratio of diameter to second-moment diameter) and relative M^2 (square of the ratio of diameters, as M^2 is always determined by a product of two diameters). The basic propagation data are taken with the knife-edge diameter because this gives the largest detected signal-to-noise ratio (highest measured transmitted energy) and has no position sensitivity. Both an X- and a Y-oriented knife edge were mounted in a common transverse plane, and data were taken with both at each propagation distance to aide in determining beam astigmatism.

The first often overlooked factor is *the need to specify and consistently use a given higher-order-mode diameter definition in making an M^2 measurement.*

4. Four-Cuts Method

Each diameter can be measured to a finite fractional precision g, which ultimately fixes the tolerances for all the measurements made to determine M^2. Typ-

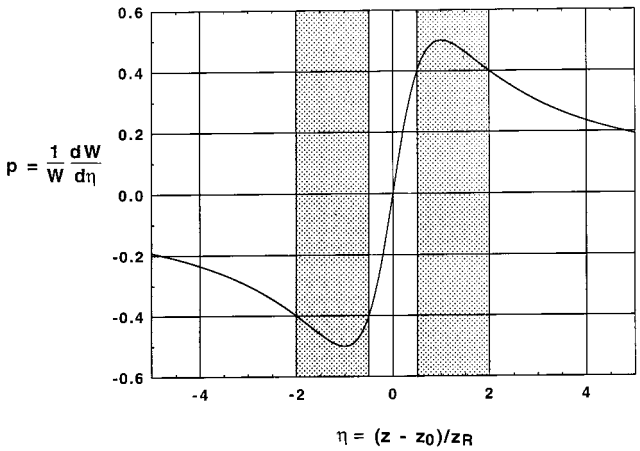

Fig. 5. Fractional change in beam diameter p as a function of normalized propagation distance η from the waist. Cuts made to locate the waist in the shaded regions benefit from a fractional change of 80% or more of the maximum. This requires a minimum of a Rayleigh range of access to the beam around the waist.

ically, diameter measurements are performed to $g = 1$–2%, giving a fractional precision h for the beam propagation factor (which goes as the product of two diameters) of $h = 3$–5%. A beam diameter measurement is termed a cut for the motion of an aperture (or cutting) across the beam to find the distance between features on the transmission function. We define a normalized propagation distance variable η by

$$\eta = (z - z_0)/z_R. \tag{11}$$

For an error in locating the waist position along the propagation axis, η_0, to cause a waist diameter fractional error less than g (assuming $g \ll 1$), one obtains from Eq. (2) the tolerable positional error

$$\eta_0 \leq \sqrt{(2g)}, \tag{12}$$

which is approximately 1/7 of a Rayleigh range for $g = 0.01$.

To find the waist position to this precision, cuts must be made far enough from the waist to see a beam diameter growth significantly greater than g. Diameter change with propagation is null at the waist. To precisely locate a null requires observations far from the null[11] where the diameter variation (and its reversal in sign) can be reliably detected. This takes cuts made at a sizable fraction of a Rayleigh range on both sides of the waist.

The optimum distances for these cuts are found from plotting the fractional change in beam diameter p versus the normalized distance from the waist as in Fig. 5. This fractional change p is computed from the differentiation of Eq. (2) as

$$p \equiv \frac{1}{W}\frac{dW}{d\eta} = \frac{\eta}{1 + \eta^2}. \tag{13}$$

A fractional change p of 80% or more of the maximum $p = 0.5$ at $z = \pm z_R$ is obtained if the waist-locating cuts are done at distances between +0.5 to +2.0 and

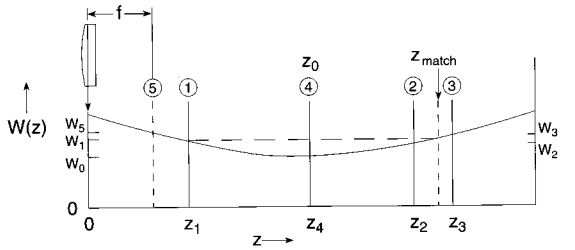

Fig. 6. Four-cuts method. The propagation plot of the beam behind the inserted lens is shown, and the circled numbers over the plot indicate the order of the cuts made to locate the waist. The propagation distance z_{match} of the diameter matching the first diameter at z_1 gives the waist location as halfway between these equal diameters.

-0.5 to -2.0 Rayleigh ranges from the waist. These are wide tolerances on locating these cuts, but all the cuts are far from the waist. *Access to at least a full Rayleigh range of propagation distance centered around the waist is needed*. Because measurements over this range for many lasers are not possible, access is gained to the waist lying either inside the resonator or at the surface of the output coupler by inserting the lens to make an auxiliary waist. *This requirement for an inserted lens, and later a transform to undo its effects, is the next often overlooked factor in performing an accurate M^2 measurement.*

In principle, three beam diameter measurements in each independent plane will fix M^2 for that plane, but one of these three cuts must be at the waist location within the tolerance of inequality (12). The problem is to precisely locate the waist with minimum additional cuts. To do this, a rough estimate is first made of the waist location and of the beam's Rayleigh range by use of visual inspection, burn paper, or manufacturer's specifications. In the case of the 2.1-μm laser, a series of spots on burn paper were made at 10-cm increments along the beam path behind the auxiliary lens. The burn diameters were plotted to estimate the auxiliary waist location as 85 ± 10 cm from the $f = 52.3$-cm lens and the auxiliary beam Rayleigh range as approximately 18 cm. Next the first cut at z_1 is made, approximately one z_R from the waist location toward the lens, and the diameter $2W_1$ there is determined from analysis of the knife-edge data. Then second and third cuts at z_2 and z_3 are made at approximately $0.9\,z_R$ and $1.1\,z_R$, respectively, on the other side of the waist, and these diameters are determined. At an intermediate distance z_{match} there will be a waist diameter that matches $2W_1$, and this is determined by interpolation (see Fig. 6):

$$z_{\mathrm{match}} = z_2 + \frac{W_1 - W_2}{W_3 - W_2}\,(z_3 - z_2), \tag{14}$$

$$z_4 = z_0 = \frac{(z_1 + z_{\mathrm{match}})}{2}. \tag{15}$$

The waist is located precisely halfway between z_1 and z_{match}, and the fourth cut is made there to directly measure the waist diameter and complete the minimum data needed to determine M^2. If the locating cuts are within the ranges determined from $|p| > |\,0.8(p_{\max})| = 0.4$, and the diameters are measured to a fractional error g, the error in the normalized waist location is no worse than $2.5g$, much less than the tolerance $\sqrt{(2g)}$ determined from inequality (12) because g is small. The measured waist diameter will then be correct to the fractional error g.

The value in locating the waist position to this tighter tolerance, as precisely as the diameter error g will allow, is that this reduces the number of unknown constants from three to two, which is important if the data are subsequently to be curve fit. The number of terms to evaluate in the curve fit drops by a factor of 4, and the remaining terms, some of which depend on $(z - z_0)^4$, are made accurate (see Appendix A). It is often advisable to take a fifth cut at $z_5 = f$ as done here. This balances the number of data points on either side of the auxiliary waist to improve the curve fit, as well as cross checking the input beam divergence.

In the experimental arrangement a 20% transmitting flat mirror was inserted into the beam to attenuate the beam energy and eliminate the knife-edge (razor-blade) erosion that initially occurred with repetitive full-energy pulses. The $f = 52.3$-cm lens was placed at 120 cm from the laser's output coupler, and behind this a dual-axis translation stage supported the X and Y razor blades in the common transverse plane. The X–Y stage slid in the propagation direction (z axis) along a rail set parallel to the auxiliary beam, and a steel metric tape permitted readout of the z-axis setting to 0.1 cm. A Molectron pyroelectric energy meter read the energy transmitted past the knife edge on each shot, the edge was translated an increment of the order of a tenth of a beam radius, and the next shot was taken.

For each shot the internal reading from the laser control circuitry was recorded as well to keep watch on the constancy of the output pulse energies throughout a cut. At the end of each cut, the razor blades were backed out of the beam and the energy was recorded for at least five shots at 100% transmission, the average of which was used to normalize the transmission data. Figure 4(a) shows an example of the raw data and its reduction by graphic interpolation to obtain $2W_{\mathrm{KE}}$. The same cut locations chosen by the four-cuts method for the X plane were used in the Y plane to aide in determining the beam astigmatism. Table 2 summarizes this data.

The four-cuts method is similar to the recommendation of the ISO draft standard[10,12] for beam parameter determination, except that in the ISO standard approximately ten diameter measurements on either side of the auxiliary waist are taken for each independent plane, and the data for each plane are curve fit to a second-order polynomial in the beam diameter. With four cuts per plane, and generally, a *weighted curve fit to the correct hyperbolic form is necessary* (see Appendix A). In commercial instruments, many more cuts are made[13] to ensure that

Table 2. Knife-Edge Diameters and Waist Locations Measured in 2 Space for the First Pulse 2.1-μm Beam

Cut Number	z (cm)	$2W_X$ (mm)	$2W_Y$ (mm)
5	52.3	5.57	5.57
1	59.5	4.75	4.50
	83.3		[a]
4	85.0	2.69	2.66
	85.1	[a]	
2	105.5	4.06	4.36
3	110.5	4.72	4.80

[a]Denotes the interpolated waist location from the four-cuts method.

diameter samples are taken within the propagation distance tolerance of inequality (12) for an accurate waist diameter measurement.

5. Data Analysis

The data in Table 2 are now to be fit to the propagation equation, Eq. (2), to determine the beam constants in 2 space. The data are first plotted and a Rayleigh range is graphically determined for both beams (Fig. 7). A graphic solution consists of one using the diameter of the measured point closest to the waist as the waist diameter, then laying in a smooth curve of an approximate hyperbolic form symmetrically about the known waist location for each axis (here with a French curve). Next, horizontal chords are marked off at heights $\sqrt{2}$ times the waist diameters. The lengths of these chords are twice the respective Rayleigh ranges $2z_{RX}$ and $2z_{RY}$, which are measured and give M^2 from the inverse of Eq. (3):

$$M^2 = (\pi W_0^2)/(\lambda z_R). \tag{16}$$

This is termed the initial graphic solution, and from Table 2 and Fig. 7 these are $z_{RX} = 17.6$ cm and $z_{RY} = 17.8$ cm, yielding the knife-edge results $M_X^2 = 15.4$ and $M_Y^2 = 14.9$.

These initial graphic solutions were originally intended as the inputs for a two-parameter analytical curve fit to determine final values of $2W_0$ and M^2. The transmission curve data points were reproducible to 1–1.5%, giving $g = 1.5$–2% reproducibility to the knife-edge diameters determined from them. For a cut to measure the waist diameter within 2%, the miss distance η_0 of the cut must be less than 0.20, as determined from inequality (12). The Rayleigh ranges for the 2-space beams of Fig. 7 are greater than 17 cm, which means that a 3.4-cm miss distance is the smallest to make a significant diameter error. From Table 2 the Y axis has the largest miss distance of 1.7 cm.

As the Y axis was expected to have the largest error, a weighted curve fit was first performed with this data (Appendix A), with a result of 1.9% for the goodness of fit, the root-mean-square (rms) residual divided by the weighted mean diameter (mean fractional error). The rms residual may be similarly calculated for the initial graphic plot [the same equation applies, Eq. (A11), but with $a = b = 0$] with a goodness-of-fit result for this axis of 2.3%. The closeness of these results suggests that an improved or corrected graphic fit may suffice for the analysis.

The corrected graphic solution uses the fact that a better estimate of the true waist diameter is known than just the diameter of the closest measured point. By Eq. (2), if the miss distance of the closest point $i = 4$ is η_0, then the best estimate of the waist diameter is

$$2W_0 = \frac{2W_4}{(1 + \eta_0^2)^{1/2}}. \tag{17}$$

The corrected graphic solution uses the initial Rayleigh range and waist values in Eqs. (11) and (17) to obtain a corrected waist diameter and finds the chord length at height $\sqrt{2}$ times this corrected diameter on the plot to determine $2z_R$ and M^2 from Eq. (16). The chords in Fig. 7 are for the corrected graphic solutions; only the Y-axis case gave a significant correction over the initial graphic solution. This corrected solution, designated by double primes, is $2W_{0Y}'' = 2.648$ mm, $z_{RY}'' = 17.28$ cm, $(M_Y^2)'' = 15.17$ and gives a fractional error of 1.8%, which is indistinguishable from the curve fit solution. This means that the initial graphic solution should suffice for the (better) X-axis data.

In hindsight this result is not unexpected. It is a consequence of the four-cuts measurement plan. The waist diameter is directly measured, and if the initial estimate is reasonable for the beam's Rayleigh range, the other points are near the end points of the z_R chord drawn on the graph. Then the propagation plot amounts merely to an analog interpolation to find these end points.

Table 3 summarizes the 2-space solutions for the two beams (by use of the Y-axis fit from Appendix A), the computed transform constant from Eq. (8), and finally the 1-space solutions from Eq. (9) divided by the knife-edge conversion factors from Table 2 to put the ending result into second-moment currency.

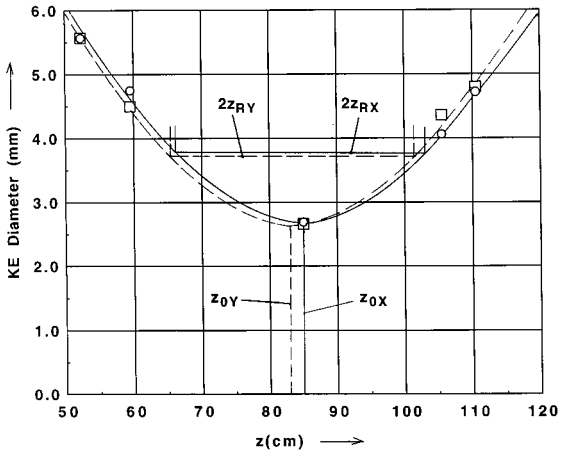

Fig. 7. Graphic analysis of the auxiliary beam propagation data. The chords giving the Rayleigh ranges for the X- and Y-plane beams are drawn at heights on the plot (diameters) larger than the waist by $\sqrt{2}$.

Table 3. First Pulse Beam Constants Measured in 2 Space and Transformed to 1 Space

Quantity	Plane		Fractional Error (%)
2-space (auxiliary beam) knife-edge units	X	Y	
z_0 (cm)	85.1	83.3	4 (of z_R)
$2W_{02}$ (mm)	2.69	2.66	2
z_{R2} (cm)	17.6	17.4	4
M^2	15.4	15.2	4
Γ	1.974	2.170	2
1-space (input beam) second-moment units	X	Y	
M^2	13.8	13.6	5
$2W_0$ (mm)	3.58	3.70	3
z_0 (cm)[a]	+2.9	−0.4	6 (of z_R)
z_R (cm)	34.7	37.8	6
Θ (mrad)	10.3	9.85	6
$2W_0/M$ (mm)	0.964	1.00	6
Astigmatism $(z_{0Y} - z_{0X})/z_{RR}$	−0.09		8
Waist asymmetry $2W_{0Y}/2W_{0X}$	1.03		6
Divergence asymmetry Θ_Y/Θ_X	0.95		8

[a]Referenced to the output coupler at 120 cm in front of lens f.

The average of the Rayleigh ranges for the two beams z_{RR} was used to normalize the beam astigmatism.

From the estimated fractional error of 2% for auxiliary beam diameter measurement, the other fractional errors are estimated as double that for quantities determined by two correlated diameter measurements, and the square root of the sum of the squared errors for the other quantities that are determined by uncorrelated products or quotients. The input beam is seen to be within 5% of round in both spatial and angular dimensions, with a slight astigmatism, and both X and Y waists are located essentially at the output coupler. The quantity $2W_0/M$ is the embedded Gaussian waist diameter.

6. Curve Fits

Although a curve fit may not be necessary in the four-cuts method, with more data points (always desirable) it is the only general way to account for all the data properly. Also, the curve fit residuals give the best estimate of the measurement accuracy. Finally, a common overlooked factor is *to use the wrong curve fit*, requiring discussion of the correct one (explained in detail in Appendix A).

The correct theoretical propagation curve is the hyperbolic form, Eq. (2), but it is not enough to just fit to this form. From past experience[4] a weighted curve fit is used, with the weight of the ith squared residual in the least-squares sum an inverse square power of the measured diameter $2W_i$. There are three arguments for use of this curve fit weighting.

In principle, in a weighted curve fit, the weights should be the inverse squares of the uncertainties in the original measurements.[14] In a variety of cw lasers and repetitively pulsed lasers, the diameter deviations from the fitted curve are observed to increase with the measured diameter (the fractional errors increase as distances from the waist increase). This is probably due to the longer time to measure a larger diameter, giving a larger effect from beam amplitude noise.

The second argument results from an experimental study[4] of different weightings in the presence of beam amplitude noise induced on a fundamental mode ion laser source with known $M^2 = 1.03$–1.05. During the 30-s run[13] to measure M^2 with a ModeMaster propagation analyzer, controlled noise was produced by manually dithering the laser's tube current. The set of propagation data for a run was then curve fit to Eq. (2) five times by use of five different weighting factors. Weight c_i for the ith diameter measurement was $c_i = (2W_i)^n$ and the five curve fits were for $n = -1, -0.5, 0$ (unity or equal weighting), $+0.5$, and $+1$.

Data runs were repeated with different levels of noise. The results showed that the inverse weightings, $n = -1$, -0.5, and to a lesser extent equal weights ($n = 0$), would give stable M^2 values within 3% of the correct value for 5% peak-to-peak amplitude noise. However, the positive power weightings $n = +0.5$, $+1$ gave increasingly larger M^2 errors of 4–5% and 12–19%, respectively, because of an increasingly larger sum of diameter residuals with increasing n. With larger amplitude noise, the M^2 errors for positive power weightings grew rapidly and nonlinearly with increasing n. The diameter residuals were observed to increase with the size of the diameter as stated above.

A commonly encountered curve-fitting mistake is to use a polynomial curve fit for the square of the beam diameter versus propagation distance. This attempts to take advantage of the fact that $W(z)^2$, the square of Eq. (2), is a polynomial in z. Polynomial curve fits are widely available in commercial software, and curve fits to a hyperbolic form are much less available. *By doing this one makes the mistake of using a positive power weighting function to fit the propagation data, which gives unreliable results*. If $2W_i$ is the measured diameter and $2W_i'$ the exact diameter, with $2\delta_i$ the (small) deviation, then in the W^2 polynomial curve fit, the ith term is $(W_i)^2 = (W_i' + \delta_i)^2 = (W_i')^2 + 2W_i'\delta_i$. That is, the residual $(W_i)^2 - (W_i')^2$ from the exact polynomial curve is weighted by $2W_i'$ in the fit, a positive power of W_i'.

Equally prone to unreliable results is the attempt to locate the waist by a fit to one-sided data, where the waist is located within a quarter of a Rayleigh range of the end of the sampled propagation region. Small errors in diameter measurements or weightings can then move the fitted waist location in either direction by large fractions of a Rayleigh range.

The third argument for an inverse-power weighting is that mathematically the least fractional error in M^2, from Eq. (6), results if the fractional errors from the denominator and numerator are roughly balanced. The residuals from the more numerous cuts far from the waist (which determine beam divergence and numerator) would swamp that from the

few cuts (or single cut) at the waist (which determines the normalizing Gaussian divergence and denominator) without an inverse weighting. In the four-cuts method an inverse square weighting approximately halves the influence of the three or four far points relative to the unity weight of the waist point to give a rough balance.

7. Test of the Beam Constants

In its intended application the NS-600 laser output is coupled into a 400-μm-diameter silica fiber through a fiber-focusing triplet lens of effective focal length $f = 15.0 \pm 0.5$ mm. The focused beam is difficult to measure, having a Rayleigh range of a fraction of a millimeter and an energy density high enough to drill most materials. One goal in modeling the direct output was to be able to do analytical tolerancing studies of the focused beam.

The lens transform equations, Eqs. (8) and (9), starting from the 1-space constants of Table 3, predict, after focusing by the $f = 15$-mm lens, the results $2W_{0X} = 157$ μm, $z_{RX} = 599$ μm, $2W_{0Y} = 148$ μm, and $z_{RY} = 540$ μm, with the beam being within 5% of round in spatial and angular dimensions with only 3% astigmatism. (This assumes that the triplet lens aberrations cause no significant increase in M^2.)

To verify these predictions, an experiment was performed to measure directly the knife-edge diameter of the focused beam on one axis in the horizontal plane. The polished edge of a BK-7 glass cube, originally fabricated as part of a cube beam splitter, was used as a refractive knife edge that could tolerate the focused beam (after the direct beam had been attenuated to 80 mJ/pulse). A digital micrometer drove the X axis of an X–Z stage carrying the cube into the beam.

Because of the difficult microphonic nature of the experiment, the knife-edge data on the focused beam could not be gathered with the control and economy of the four-cuts method. Seven points were measured, with two near the waist, and to determine the best constants the data were curve fit as in Appendix A. The results were a best-fit diameter $2W_{02X} = 153$ μm (quite close to the 157-μm prediction), $z_{RX} = 551$ μm (versus 599 μm), and $M_X^2 = 16.0$ (versus 15.4 assumed in the prediction). The fractional error of residuals over the weighted mean measured diameter was 3.7%. Thus the predictions of the beam focal properties from direct beam measurements and the M^2 model were quite acceptable.

8. Summary of the Method and Overlooked Factors

Beam propagation measurement is made as easy as it gets if the minimum number of steps are accepted and executed:

(a) Use a forgiving method for the multiple diameter measurements along the propagation axis; then at one plane in the far field carefully measure the conversion to second moments and apply this to the final results.

(b) Use a lens of known focal length to form an auxiliary (and accessible) waist.

(c) Sample the auxiliary beam at between 0.5 and 2.0 Rayleigh ranges on both sides of the waist and at the waist.

(d) Match a diameter on one side by interpolating between diameters measured to the other side to locate the waist precisely and sample its diameter directly—the four-cuts method.

(e) Use the corrected graphic analysis to reduce the propagation data or (as in Appendix A) use a weighted least-squares curve fit to the correct hyperbolic form, Eq. (2), with the inverse squares of the measured diameters as weights.

(f) Transform the auxiliary beam constants back through the lens to obtain the propagation constants of the original beam, accounting for the compounding of errors in all the steps.

The often overlooked factors, to be wary of, or what not to do are

(a) Use an experimentally difficult diameter measurement method, neglect to state what method was used, or apply different methods and mix results inconsistently.

(b) Measure whatever portion of the beam that is directly available (this is usually data on only one side of the waist, which give an unreliable propagation fit).

(c) With the waist inaccessible, assume its location, calculate its diameter, or trust generic laser specifications (the result then is only half of an M^2 measurement).

(d) In sampling the beam, miss the waist location by a distance more than a quarter of a Rayleigh range, or determine the beam divergence from a diameter taken within a distance less than half of a Rayleigh range from the waist.

(e) Fit the data to a polynomial form, the square of Eq. (2), instead of the correct weighted hyperbolic one (yielding an unreliable fit).

(f) Neglect the compounding of errors in computing M^2 from the 2-space diameter measurements, the lens transform to the input beam constants, and the conversion to the standard second-moment currency.

Appendix A. Weighted Least-Squares Curve Fit of Propagation Data

1. General Form

The traditional method to fit data points to a curve of general form is given in Ref. 14 (Problem 8.4) and Ref. 15 (Section 13.37) and is reviewed here to derive a particularly efficient form for the beam propagation fit in Subsection A.2. The general method with two undetermined constants fits a data set of i points y_i to $y = f(x, A, B)$ to fix the most probable values of the constants A and B. Initial values, plus small corrections, are assumed:

$$A = A_0 + a, \quad B = B_0 + b, \tag{A1}$$

Table 4. Weighted Least-Squares Curve Fit, Auxiliary Beam *Y*-Axis Data

Cut Number i	z_i (mm)	$z_i - z_0$ (mm)	$2W_i$ (mm)	$2W_i'$ (mm)	F_i (mm)	P_i	Q_i	R_i	S_i	T_i	$c_iF_i^2$
5	523	−310	5.57	5.351	0.2190	7.35×10^{-3}	3.978	3.37×10^{-3}	3904	1.824	1.55×10^{-3}
1	595	−238	4.50	4.448	0.0523	17.25×10^{-3}	5.504	1.53×10^{-3}	1756	0.487	0.135×10^{-3}
4	850	17	2.66	2.672	−0.0122	0.1413	0.230	-1.72×10^{-3}	0.374	−0.0028	0.021×10^{-3}
2	1055	222	4.36	4.258	0.1019	19.58×10^{-3}	5.434	3.27×10^{-3}	1508	0.908	0.546×10^{-3}
3	1105	272	4.80	4.865	−0.0654	1.33×10^{-3}	5.554	-1.57×10^{-3}	2314	−0.655	0.186×10^{-3}
						$\Sigma = 0.1988$	20.700	4.88×10^{-3}	9482	2.561	2.43×10^{-3}

and the residuals from the theoretical curve are given by

$$d_i = y_i - f(x_i, A, B). \tag{A2}$$

The residuals are expanded in a Taylor series, keeping only the first-order correction terms

$$d_i = F_i - a\frac{\partial f(x_i, A_0, B_0)}{\partial A} - b\frac{\partial f(x_i, A_0, B_0)}{\partial B} \tag{A3}$$

with

$$F_i = y_i - f(x_i, A_0, B_0). \tag{A4}$$

The derivatives are given by the symbols

$$u_i = \frac{\partial f(x_i, A_0, B_0)}{\partial A}, \quad v_i = \frac{\partial f(x_i, A_0, B_0)}{\partial B} \tag{A5}$$

to convert Eq. (A3) to

$$d_i = F_i - au_i - bv_i. \tag{A3'}$$

Next the weighted sum of the squares of the residuals is formed and minimized. The ith squared residual term is multiplied by the assigned weight c_i, the sum over the i data points taken, and the least square or minimum of $\Sigma c_i d_i^2$ can be found by setting the derivatives with respect to a and b of this sum equal to zero:

$$\sum c_i d_i \frac{\partial d_i}{\partial a} = 0, \quad \sum c_i d_i \frac{\partial d_i}{\partial b} = 0. \tag{A6}$$

Using Eq. (A3′) gives the normal equations

$$\begin{aligned} -2\sum c_i(F_i - au_i - bv_i)u_i &= 0, \\ -2\sum c_i(F_i - au_i - bv_i)v_i &= 0 \end{aligned} \tag{A7}$$

of the form

$$Pa + Qb = R, \quad Qa + Sb = T, \tag{A8}$$

where

$$\begin{aligned} P &= \sum c_i u_i^2, \quad Q = \sum c_i u_i v_i, \quad R = \sum c_i F_i u_i, \\ S &= \sum c_i v_i^2, \quad T = \sum c_i F_i v_i. \end{aligned} \tag{A9}$$

The solution to Eqs. (A8) is

$$a = \frac{RS - QT}{PS - Q^2}, \quad b = \frac{PT - QR}{PS - Q^2}, \tag{A10}$$

which completes the fit when these values are put back into Eqs. (A1). The root-mean-square residual is

$$s = \left(\sum c_i d_i^2 / \sum c_i\right)^{1/2} = \left[\left(\frac{\sum c_i F_i^2 - aR - bT}{\sum c_i}\right)\right]^{1/2}. \tag{A11}$$

2. Beam Propagation Form

Because the curve fit depends on derivatives of the theoretical curve with respect to the fitted constants, it is convenient to rewrite Eq. (2) of the text in terms of the full divergence angle Θ [by use of Eq. (6)] to make W_0 and Θ the fitted constants. This gives the simplest algebraic expressions for the derivatives. Thus

$$2W(z) = \{[4W_0^2 + \Theta^2(z - z_0)^2]\}^{1/2}, \tag{A12}$$

with

$$2W_0 = 2W_0' + a, \quad \Theta = \Theta' + b, \tag{A2'}$$

where W_0' and Θ' are the first-order values of auxiliary waist radius and divergence. The required derivatives evaluated at the point at $z = z_i$ are

$$u_i = \frac{\partial W}{\partial W_0} = \frac{2W_0'}{2W_i}, \quad v_i = \frac{\partial 2W}{\partial \Theta} = \frac{\Theta'(z_i - z_0)^2}{2W_i}. \tag{A13}$$

If the ith point is given weight $c_i = 1/4W_i^2$, the quantities of Eq. (A9) become

$$\begin{aligned} P &= \sum \frac{(2W_0')^2}{(2W_i)^4}, \quad Q = \sum \frac{2W_0'\Theta'(z_i - z_0)^2}{(2W_i)^4}, \\ R &= \sum \frac{2W_0'F_i}{(2W_i)^3}, \quad S = \sum \frac{(\Theta')^2(z_i - z_0)^4}{(2W_i)^4}, \\ T &= \sum \frac{F_i\Theta'(z_i - z_0)^2}{(2W_i)^3}. \end{aligned} \tag{A14}$$

3. Numerical Example

The Y-axis auxiliary beam data from Table 2 of the text is curve fit as an example of the procedure. It is convenient to use a spread sheet organized as in Table 4. From a graphic analysis of the data, the trial values are $2W_0' = 2.66$ mm, $z_0 = 833$ mm, and $z_R' = 177.6$ mm [which yields by Eq. (6) $\Theta' = 14.98$ mrad]. The theoretical form of Eq. (2) gives, for the ith data point, the diameter values

$$2W_i' = 2W_0' \{[1 + (z_i - z_0)^2/(z_R')^2]\}^{1/2}, \quad \text{(A15)}$$

which gives

$$F_i = 2W_i - 2W_i'. \quad \text{(A16)}$$

The letter quantities in Eqs. (A11) and (A14) were evaluated by summing the respective column entries, $P = \Sigma P_i$, etc., in Table 4.

Putting these quantities into Eqs. (A10) and (A2′) gives the fitted values $a = -0.00462$ mm, $2W_{0Y} = 2.655$ mm, $b = 0.280$ mrad, and $\Theta_Y = 15.26$ mrad. By use of Eqs. (6) and (16) of the text this gives $z_{RY} = 174.0$ mm and $M_Y^2 = 15.15$. Finally, the fit gives the rms residual s_Y from Eq. (A11), which uses $\Sigma c_i = 0.3190$ mm^{-2} as $s_Y = 0.0738$ mm. This can be compared with the weighted mean diameter

$$\langle 2W_Y \rangle = \frac{\sum 1/(2W_i)}{\sum 1/(2W_i)^2} = 3.81 \text{ mm} \quad \text{(A17)}$$

to give an estimate of the goodness of fit of $s_Y/\langle 2W_Y \rangle = 1.9\%$.

References and Notes

1. A. E. Siegman, *Lasers* (University Science, Mill Valley, Calif., 1986), Section 17.6.
2. T. F. Johnston, Jr., "M^2 concept characterizes beam quality," Laser Focus, 173–183 (May 1990).
3. W. T. Silfvast, *Laser Fundamentals* (Cambridge U. Press, Cambridge, UK, 1996), pp. 340–342.
4. The author was the chief engineer for the development of the ModeMaster beam propagation analyzer, a product of Coherent, Inc., Instruments Group, 2303 Lindbergh St., Auburn, Calif., 95602.
5. M. W. Sasnett, "Propagation of multimode laser beams—the M^2 factor," in The Physics and Technology of Laser Resonators, D. R. Hall and P. E. Jackson, eds. (Hilger, New York, 1989), Chap. 9, pp. 132–142.
6. H. Kogelnik and T. Li, "Laser beams and resonators," Appl. Opt. **5,** 1550–1566 (1966).
7. P. A. Bélanger, "Beam propagation and the *ABCD* ray matrices," Opt. Lett. **16,** 196–198 (1991).
8. J. Serna and G. Nemes, "Decoupling of coherent Gaussian beams with general astigmatism," Opt. Lett. **18,** 1774–1776 (1993).
9. D. C. O'Shea, *Elements of Modern Optical Design* (Wiley, New York, 1985), pp. 235–237.
10. "Test methods for laser beam parameters: beam widths, divergence angle, and beam propagation factor," ISO/TC 172/SC9/WG1, ISO/DIS 11146, available from Deutsches Institut für Normung, Pforzheim, Germany.
11. To drive home this point and show that it is nothing new (though often overlooked), the author calls this the Stonehenge Effect, after Fred Hoyle's convincing interpretation of stone placements made in Chap. 4 of his book *On Stonehenge* (Freeman, San Francisco, Calif., 1977). Hoyle shows that 5000 years ago the builders of this ancient monument understood that, to locate a null, you must look away from the null point. The monument was built to locate the exact day of the summer solstice, the day the position of the rising Sun in its northward march along the horizon reversed direction and turned southward. If the sighting stone had been placed directly to mark the position of the turnaround, there would have been an ambiguity of at least a day in locating the solstice because the Sun's motion per day is nil at that time. Instead, the sighting stone obscured a span of the horizon and was of a width and placement that the exact day of the solstice was marked as the day midway between the days of disappearance and reappearance of the Sun on the south side of the stone. This resolved the ambiguity in the sightings on the trio of days around the solstice when the Sun emerged on the opposite (north) side of the stone.
12. In support of the adoption of the ISO beam characterization standard (Ref. 10), Coherent, Inc. has stated that it will grant royalty-free license to its patent (U.S. patent 5,267,012) on the use of a lens to form an auxiliary waist for M^2 measurement of an astigmatic beam.
13. In the Coherent ModeMaster instrument, in the focus pass to determine M^2, cuts are made at 260 points along the auxiliary beam propagation path in each independent plane.
14. J. R. Taylor, *An Introduction to Error Analysis* (University Science, Mill Valley, Calif., 1982).
15. H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry* (Van Nostrand, New York, 1943), pp. 500–502.