Proc. of SPIE Vol. 0406, Optical Specifications: Components and Systems, ed. W.J. Smith, R.E. Fischer (Apr 1983) Copyright SPIE

Optical specifications and tolerances for large optics

Robert E. Parks
Optical Sciences Center, University of Arizona
Tucson, Arizona 85721

Introduction

The term "large" has specific implications about types of optics and their uses. In this paper large optics are defined and figure, beauty, and blank characteristics are specified with tolerances for each. Because it is useless to discuss specifications for properties that cannot be measured, tests for each specification are described. The specification of a large optic takes far more thought and insight than is the case for small optics because of the high cost and demanding applications of large optics.

Large optics and their ramifications

Large optics can be defined as any optic 50 cm (20 in.) in diameter or larger. Such a size boundary is appropriate for two reasons: (1) An optic this large can be moved around the shop only with mechanical aid; it is too big to be safely handled by one man alone. (2) Large optics require sophisticated mounting techniques. They are too big to simply polish on a rug pad support and to test on a simple plywood stand.

Having made this definition let us examine the consequences implied by the 50-cm size boundary. There are virtually no pieces of refractive glass larger than 50 cm, so large optics are primarily mirrors. Since mirror systems tend to have fewer elements than do refractive systems, they are usually aspheres because each surface must correct both for power and asphericity in the wavefront. These mirrors are difficult to make accurately, so large optics in general are expensive and thus there is little margin for error in the mirror specifications.

In addition to the high cost of large optics, there are relatively few facilities that are equipped to fabricate them. Not only are the raw materials for large optics expensive and the labor costs high because of the level of skill required by the opticians, but the tooling and polishing equipment is specially designed, one-of-a-kind machinery.

A final characteristic of large optics is that their mechanical deflections are large compared to optical tolerances if they are not supported correctly. The 6:1 ratio of diameter to thickness necessary for adequate stiffness for optical working and support loses all meaning for large optics. In addition, few applications can afford the luxury of a 6:1 ratio when the cost of the blank is taken into consideration. Thus the method of support during fabrication, testing, and use is critical to the specification of large optics. Any organization that wants to know the cost of a large optic for which the support method during use has not been determined and specified is misleading itself. No competent optical fabrication company will quote on the job without guidance in this area.

Implications of the use of mirrors

The use of mirrors rather than refractive elements affects the way the surfaces are toleranced. The contribution of surface error to wavefront error is roughly 4 times as great for a mirror as it is for a lens. The effective index change at a lens surface is n-1 whereas it is 2n at a mirror surface. Thus tighter surface tolerances are required on mirrors than lenses in general.

For scattering the situation is worse by at least a factor of 2 for mirrors. With a lens or window surface, any defect or scattering site can be expected to scatter at least half the light completely out of the beam in the backward direction. With a mirror used at near normal incidence, the backward direction is the forward direction as well. Consequently all the scattered light goes in the direction of the detector. Therefore, the specification on scattering should be written to allow only half the scattering sites as would be allowed on a lens surface used for the same purpose.

Thinking regarding coatings must also change when considering mirrors. If a system contains lenses, the coatings should make the surfaces seem to vanish so that there are no ghost images due to the roughly 4% Fresnel reflectance per surface. With mirrors it is desirable to have the highest possible reflectivity to maximize the light throughput of the system. For many types of systems there is a more important reason: the light that is not reflected is either absorbed or scattered. For high-energy systems neither of these two alternatives is attractive. Even in the less dramatic cases, the less light there is that is not reflected, the less problem there is dealing with it. Often it becomes a problem of balancing a coating to have the least detrimental effects rather than optimizing some of the more desirable features of the coating.

The sensitivity of mirrors to support conditions has already been mentioned, yet support is virtually never mentioned with regard to lenses. The reason is that if a lens has been figured correctly and then it is warped in its cell, both of its surfaces move together in a way to largely cancel the effect of the distortion. There is no net thickness change in the lens even though it is distorted, and consequently there is no change in optical path length to first order. With mirrors, there is a direct one-for-one change in optical path (times 2) as the mirror surface is distorted from its nearly ideal surface at the time of manufacture.

Finally, mirrors are generally made of glass. The types of glass used for making mirrors have a coefficient of thermal expansion from 10 to 100 times less than that of the metal structure in which they are usually mounted. This disparity increases the severity of the support problem for the mirror. Not only must it be supported without strain, but it must be done in a way that constrains the mirror against rigid body motion relative to the cell yet allows for the difference in thermal expansion between the mirror and its cell. The need for a thorough mechanical analysis should be obvious.

Uses of large optics and their effect on specifications

There are three categories of uses for large optics: terrestrial telescopes, satellite-borne telescopes for both earth resources use and astronomy, and high-energy laser applications. Because of the widely differing environment for each of these applications, the figure quality, beauty, and support methods for large optics will differ. Table 1 illustrates how the use dictates the importance of certain specifications.

For example, optics to be used in a 10.6-μm laser fusion experiment can have relaxed figure requirements compared to those used in the visible. Also the support system is simplified because the mirrors will always be used in one orientation relative to gravity. On the other hand, the support system may be complicated by shock and/or thermal loads. At the other extreme is an application such as the Space Telescope. Because there is no atmosphere, the surfaces should be as smooth and free of scattering sites as possible. Also the telescope will be pointed close to the sun so as not to limit the sky area that can be observed and to minimize scattering. This again calls for a minimum of beauty defects and high spatial frequency roughness.

Figure 1 shows a single-mirror telescope in a possible environment. It is the function of the engineer to determine what the performance of this telescope will be in the presence of the perturbations shown. An approach often used is to approximate the root mean square (rms) wavefront errors due to these perturbations and root sum square the results to determine the total rms error in the wavefront as it approaches the detector. These results are combined with the detector characteristics, and the system resolution is determined for various contrast objects.

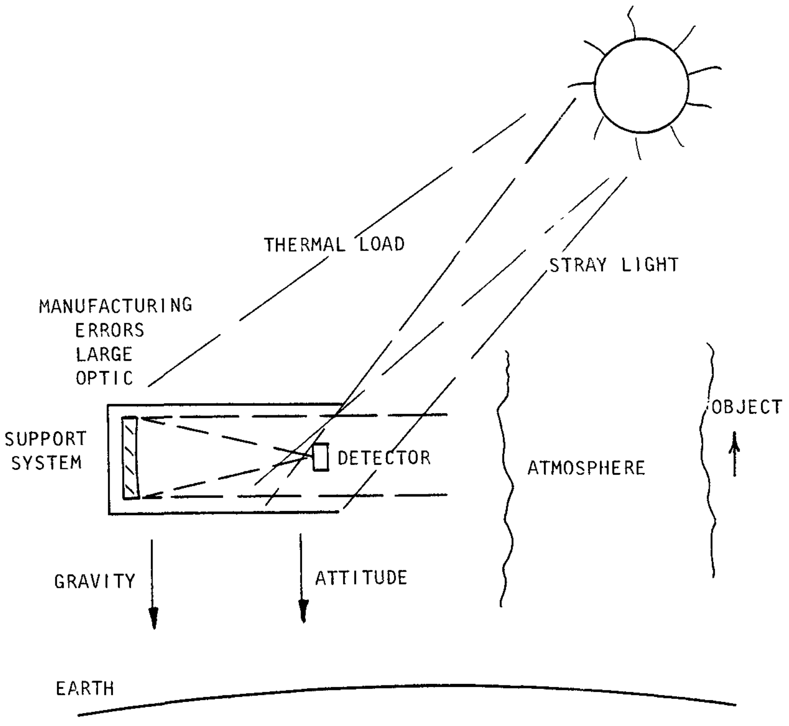

Figure 1. Operating environment of optical system.

Table 1. Application of Specifications According to Optical System Type.

	Quality aspect			
Use	Figure	Beauty and roughness	Blank design	Support method
Terrestrial telescope	Atmosphere limited	Not too important	Low thermal inertia Light weight	Changes relative to gravity
Space telescope	Must be good	Quite important, scattering	Light weight CTE Homogeneity	Gravity release
High power laser	Wavelength dependent	May be critical, Mirror damage	Shock and thermal loads	Application dependent

This simple one-parameter approach has been effective in the past, and there are many successful optical systems that have been designed using this method. The advantage of the method is that the logic behind the choice of tolerances is simple to understand. The method has the disadvantage that it tends to obscure the perturbations that affect the system performance the most, and it does not indicate clearly those areas where the tolerances may be loosened without harm. These disadvantages become more severe with larger systems and as more active mechanisms are included to counteract the perturbations on the systems. Such complex systems require more than single-number inputs around which a servo system must be designed.

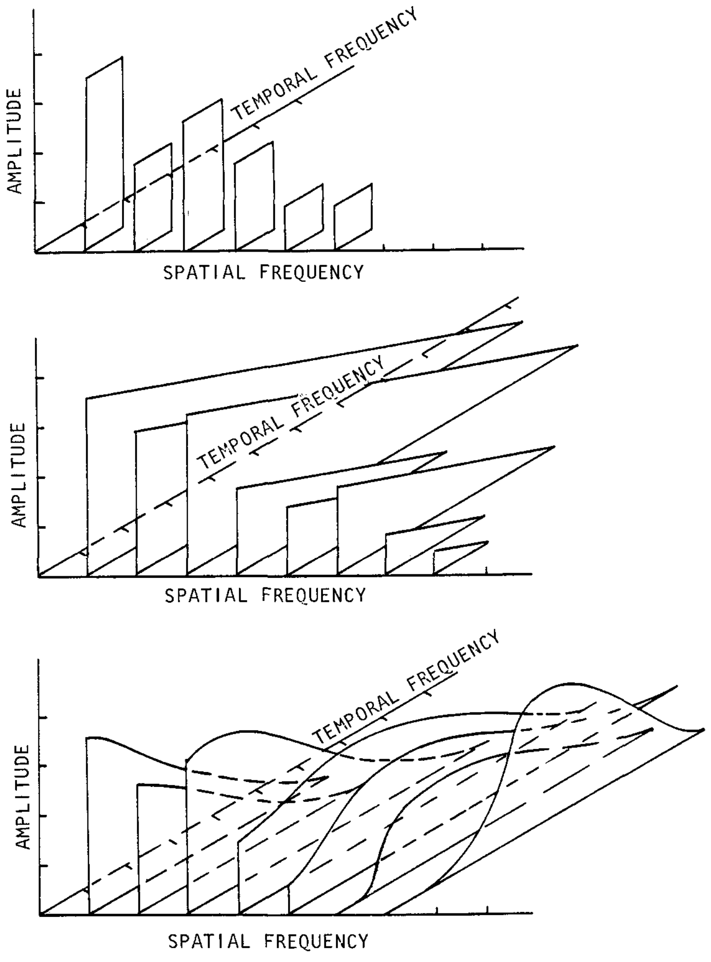

Figure 2. Amplitude of wavefront errors as a function of temporal and spatial frequency.

A somewhat more sophisticated approach can be illustrated by using the telescope support system. Assume that a terrestrial telescope is in an alt-azimuth mount and that the primary mirror sees a change from axial to radial support as the telescope goes from zenith pointing to horizon. The deformation of the primary can be modeled by any of several finite-element structural analysis programs. If these deflections are then fit with the same type of polynomials used for describing aberrations, for example the Zernike polynomials, each of the polynomial terms could be plotted as a function of elevation. If the mirror support system responds as expected, only the lower order polynomial terms will be significant, and they will vary monotonically with the angle of elevation.

Similarly, a thermal analysis could be made of the mirror blank to determine how its shape will be affected by diurnal variations in temperature. These variations would also be represented by low-order spatial frequency terms of the polynomial and have a slowly varying temporal frequency. These concepts are shown in Figure 2 where the spatial frequency domain is plotted as the x axis and the temporal domain is plotted into the page. The amplitude of the terms is plotted in the z direction.

This type of representation shows two things. First, by using an optical polynomial to represent the deflections of the mechanical model, those deformations that are the equivalent of the low-order aberrations can possibly be corrected by optical means. For example, in the terrestrial telescope, an S-shaped error is caused by a deflection of the surface of the primary mirror as the telescope goes from zenith pointing to horizon. A contour map of this error in two dimensions is similar to that for coma. The question then arises, would it be possible to compensate for this effect in a Cassegrain telescope with a lateral translation of the secondary mirror as this motion can occur as the instrument is tipped to the horizon. The answer may be determined by performing a similar mechanical analysis on the structure and matching the results with an optical analysis of the magnitude of coma as a function of decentering of the secondary.

Second, the temporal variation of each of the spatial components of the wavefront produced by the perturbed telescope can be specified. In the example above, if the question was what is the response of the primary mirror to a high-energy laser pulse during a fusion experiment, this information could then be used to determine the spatial and temporal frequency response necessary for a servo system to help correct the induced errors. It would then be obvious what parts of the spatial errors could be compensated by a correction scheme of a particular temporal response.

Likewise, one could attempt to represent atmospheric turbulence in a similar way to determine what spatial components of the atmosphere could be corrected in what parts of the telescope. Admittedly the atmosphere may not be well represented by a polynomial function, but using the same argument, the telescope active control system will have difficulty correcting for that part of the atmosphere that cannot be represented by a polynomial function. On the other hand, that part of the atmosphere that can be represented by tilt and focus, for example, could be compensated by simply using rigid body motion of the secondary. If it could be shown that a 50% improvement could be made in image size by this relatively simple correction scheme, this approach would probably be followed rather than some more complex approach. This method of atmospheric compensation is being implemented at the Multiple Mirror Telescope on Mount Hopkins, south of Tucson.

Clearly all the above has nothing to do with specifying the optics, but until such a study is made of the system response to likely perturbations, it is premature to discuss the tolerances on the optics.

The terrestrial telescope is a prime example of this argument. There is currently new thinking about how a telescope primary that has to look through the atmosphere should be specified. Until recently no specifications were put on any but the lowest order spatial frequency errors and even then only on the wavefront itself. (In fairness to astronomers, they have always been interested in how much energy would fall in a disc of a certain angular size, but this is a difficult measurement to make with accuracy in a test situation.) It is now recognized that slope errors greater than those likely to be introduced by the atmosphere will be detrimental to the performance of the telescope. New specifications based on this study are now being formulated for terrestrial telescopes. To do any less careful study of telescopes and large optical systems for other uses would be irresponsible.

Figure specification

Only low-spatial-frequency surface irregularities are considered in this section. Low spatial frequency errors are those that could be specified by the first 24 or 36 terms of a two-dimensional Zernike polynomial, for example. This corresponds to roughly 8 cycles per diameter radially and 4 cycles in a full revolution azimuthally.

Low-spatial-frequency figure errors are commonly referred to as irregularity and represent the departure of a surface or wavefront from a best fit spherical reference surface. In practice, the departure from a reference spherical wavefront is used (rather than departure from a best fit parabola, for example) because the optical test setup ultimately presents an interferometer with a spherical or very nearly spherical wavefront.

Although a spherical surface is seldom encountered in a large optical system, it is logical to start with this surface. With respect to low-spatial-frequency figure errors, a sphere should be specified only in terms of its nominal radius and its irregularity over some clear aperture. While the radius of curvature of a sphere can be held tightly within the ability to measure the radius, it is costly and time consuming to hold this tolerance too closely. Unless otherwise necessary, a typical tolerance might be 1% of the nominal radius.

Although a sphere is a natural figure to polish, that is, it is the figure of revolution obtained by simply rubbing two surfaces together, a lack of rigidity in the mirror support and in the lap generally limits the figure on a large sphere to 0.05 wave rms surface in the visible unless the sphere is worked locally as though it were an asphere. This local working obviously adds to the cost. One of the best large optics made to date is the Space Telescope primary mirror. When finished it was about 0.017 wave rms surface figure. If one is tempted to specify a large sphere to 0.01 wave rms surface, think twice unless preparations are made to embark on a research project. Also leave a clear aperture inside the physical diameter of the blank by approximately 5 mm on the radius. If this margin is not left outside the clear aperture, the worst error on the optic will be the rolled edge.

The next most simple surface to specify would seem to be a flat. This surface is a special case of a sphere with an infinite radius. However, it is one of the more difficult surfaces to make and test accurately. Fortunately, flats are unnecessary in most optical systems and every effort should be made to eliminate them. Since a flat has no power, it cannot be used to correct any aberrations within a system. It can only add a nonperfect surface and thus to some extent degrade the system. Furthermore, it is in the worst possible place in the system to cause degradation. Not only do errors in the surface affect the wavefront one for one but even the slightest error in absolute radius causes a figure error that cannot be removed by refocusing. It is our advice to point the entire system and eliminate flat(s) in front. At first this may seem to be a difficult way of attacking the problem of scanning, but in any large optical system it will invariably be the most cost-effective solution.

For those who insist upon a large flat, be sure that the optical company that will fabricate the flat has the software to properly reduce the test data. Also insist that a test plan be prepared so that the flat meets the required specifications. The details of data reduction for flats are beyond the scope of this paper, but a method for reducing this type of data may be found in Reference 1.

When an asphere is specified, either a null test is done, such as the test of a parabola against a flat, or a "null lens" is used to compensate for the aspheric figure by making the surface appear as though it were spherical to the test device. For most of the surfaces that are conic sections of revolution, there are geometric null tests such as those illustrated in Appendix 4 of Reference 2.

If the geometric null test is to be used, the surface should be specified in terms of the tolerances allowed in the test conjugates. The parabola has only one conjugate, its focal length. It can then be specified completely in terms of the conjugate distance between the mirror vertex and the focus of the test device such as an interferometer and the figure error representing the difference between the "nulled" surface and a near perfect reference sphere. The parabola is "figured" when it appears to the interferometer to look like a perfect sphere to within the figure error tolerance.

For an ellipse or hyperbola, two conjugates must be specified. One is usually considered a reference dimension or is fixed and the tolerance is applied to the other. It is up to the lens designer to decide what this tolerance should be. Only he can change this conjugate in his system design and see its effect on system performance as the mirror figure changes due to variations in this parameter. The vertex radius and conic constant of the desired asphere should appear on the mirror drawing only as reference dimensions. The reason for this is that neither of these attributes of the asphere can be measured directly, whereas it is easy to measure and hold the test conjugates of the mirror.

In the case of a mirror that must be tested against a null lens, the design of a suitable null lens and the distance from the null lens to mirror should be specified. The aspheric sag equation or coefficients should be listed only as reference quantities. Although this method of specifying an asphere is not common practice, it is suggested here because many optical fabrication shops do not have lens designers who are competent to design a null lens for a general asphere. Further, even if they could design the null lens, they are in no position to tolerance it because most often they do not know the final system parameters or performance requirements.

If the mirror that needs a null lens for testing is specified in terms of the null lens required, then the tolerances can be applied to the null lens, a reasonably simple lens to analyze and build. The lens designer can vary the null lens tolerances, determine the figure of the mirror made to the perturbed null lens, and then see how that mirror affects his system performance. The designer will discover prior to fabrication of the mirror whether anyone can reasonably be expected to polish the mirror using the null lens. By using this method of applying tolerances, the fabricator can do the job in which he is expert and the lens designer has no one but himself to blame if the finished mirror does not perform as expected (provided of course that the null lens has been built and used according to specification).

In spite of the design and tolerancing, it may become obvious that the quality desired in the mirror is as good or better than the quality that can be expected in the test optics. This will become all the more apparent the faster the mirror is. In this case a method of separating the errors in the test optics from those in the mirror must be used. The mirror under test may be rotated about its axis of symmetry relative to the test optics (or null lens). Those figure errors that rotate with the mirror obviously are contained in the mirror while those that remain stationary are in the test optics. This method is fully described in Reference 3 and makes use of the properties of the Zernike polynomials. It is clear that this method will separate only nonrotationally symmetric errors. A similar technique can be used to separate the rotationally symmetric errors by performing a lateral shear between the test optics and the element under test.

Mid spatial frequency errors

There are surface errors whose lateral extent places them in a region between figure errors and surface roughness. These are the mid spatial frequency surface errors. This type of error is prevalent in aspheric surfaces where it is necessary to use small laps to produce the desired figure. For the sake of definition it can be assumed that this type of surface error has a spatial extent of between 10 and 100 periods per diameter of the optic being tested. This is too high a spatial frequency to be fit to most polynomial interferogram reduction routines yet too long in spatial extent to be the cause of scattering. A good discussion of this type of error and its effect on image quality may be found in an article by Wetherell.[4]

Unfortunately errors of this type have not been treated theoretically to any extent and there exists little information on how to specify this type of error. At the same time these errors can cause a loss in system performance and are becoming a greater concern as aspheric mirrors become steeper and more difficult to make.

One recent attempt to specify the mid spatial frequency errors has been made by the Royal Greenwich Observatory.[5] The specification calls for slope errors to be less than a criterion based on the measured statistical properties of atmospheric turbulence. The thought behind this specification was that it was unnecessary to hold slope errors more tightly than those errors that naturally occur in the atmosphere through which the telescope is looking.

While the derivation of the specification is beyond the scope of this paper, the specification may be stated in the form of Table 2. For errors of various correlation distances (the mirror in this case was 4.2 m in diameter) there were maximum allowable phase errors. By dividing the phase error by the correlation distance, the specification is converted into a maximum slope error requirement. As may be seen from the table, the allowable slope errors increase with shorter correlation distances but not very rapidly. This is a reasonable specification from a fabrication standpoint because in actual practice the slope errors of a mirror will increase with decreasing spatial extent.

Once a specification has been written, it then becomes necessary to measure the slope errors on the surface. There are two rather easily implemented methods, one using the same interferograms used to determine the figure, the other a photometric method now being developed. In the interferometric method, rather than digitizing perhaps 20 points along a fringe across the diameter of the optic under test as would typically be the case for measuring figure, some 200 to 400 points are measured along a fringe at evenly spaced intervals. These data are then spatially shifted in a computer and subtracted from themselves. This in effect determines the derivative or slope of the original surface data for the particular shift distance. A simplified example of the process is illustrated in Figure 3. For the example given, the maximum value of the derivative or slope increases as the shear or correlation distance is decreased. The advantage of this method is that no additional test hardware is needed to implement this test method and only the most simple software is necessary. This technique is sure to become more widely used as appropriate slope specifications are determined.

Table 2. Slope Error Specification for the 4.2-m Primary for the Herschel Telescope.

Correlation distance	2 cm	8 cm	32 cm	128 cm
Phase errors (wavelength rms)	0.024	0.075	0.24	0.95
Slope errors (μrad rms)	0.70	0.54	0.44	0.43

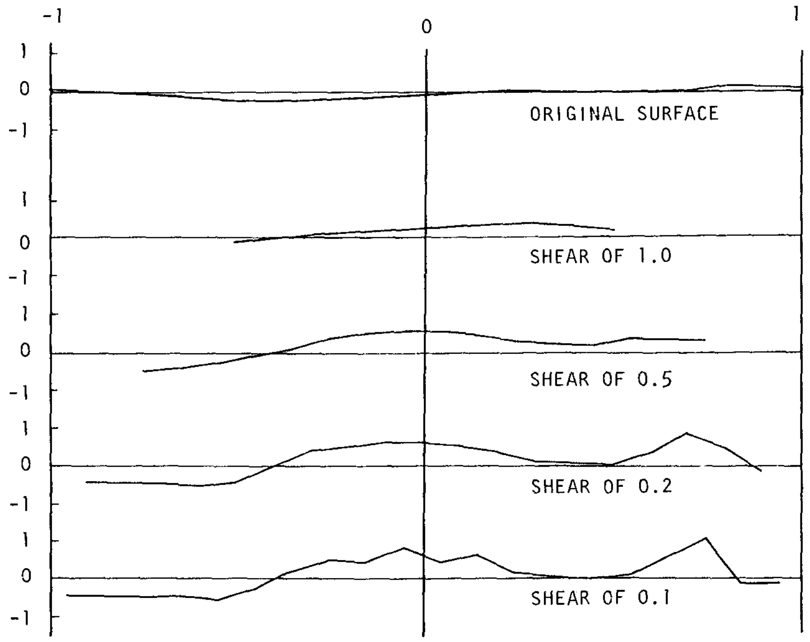

Figure 3. Example of shearing surface data to obtain slope error data.

The other approach to determining mid spatial frequency surface errors is perhaps better suited to in-process testing or rapid inspection. In this method, the intensity of the out-of-focus image in the return test beam of light is examined. If the system has no mid spatial frequency surface errors, the out-of-focus image will be a circle of light of perfectly uniform intensity except at the very edge of the spot where diffraction will produce rings. If there is surface waviness, these errors will act as weak lenslets causing some parts of the beam to be more intensely focused and other areas less so. This gives rise to a variation in intensity within the out-of-focus image. By measuring the variation in intensity as a function of correlation distance within the out-of-focus image, the magnitude of the waviness on the surface may be determined as a function of its spatial frequency. This method is in its infancy, but it shows promise as an easily implemented method of inspection.

High spatial frequency errors

The high spatial frequency surface errors and/or beauty defects are a somewhat better understood than the mid spatial frequency errors. At least there are standards for beauty defects, but specification in terms of these standards has little to do with system performance. The meaning and interpretation of the common military standards for beauty defects (MIL-O-13830A and MIL-C-48497(MU)) are discussed in Reference 6.

These military specifications or "scratch and dig" standards are often used for mirrors as well as the refracting optics for which they were originally intended. There is a hazard here in terms of the dig specification. Many mirror substrate materials, in particular the low expansion glassy materials, tend to have a number of small bubbles. These bubbles must be included in the count of digs in the surface under the military specifications. For a low scatter surface, one may be tempted to call out a 40-20 surface, for example. The optician may have no trouble producing a surface meeting the 40 scratch criterion, but the glass substrate itself may preclude meeting the 20 dig requirement independent of how well the surface is polished. A study of the glass bubble specification must be made before applying the dig specification as it would normally be applied to a refracting optic. Also see Reference 6 for a comparison of the relative surface areas affected by application of various sizes scratches and digs.

In addition to the familiar military specification for beauty or roughness defects, high spatial frequency roughness is often specified in terms of a surface roughness of, say, 30Å rms as might be determined using a profilometer, FECO or surface heterodyne micro-interferometry. This is a good specification if the optic is a few inches in size so that in will fit into a measuring device. While the optician will understand that a smooth surface is desired when such a specification is made, there will generally be little chance of verifying that the specification has been met because the appropriate test instrumentation simply does not exist that is compatible with large optics.

One of the better ways of dealing with this problem for large optics is to write the specification in terms of scattered light. First, scattering is the reason for applying a roughness specification. The roughness that produces the scattering affects the optical system's ability to image. Second, since light can measure at a distance, so to speak, large mirrors do not necessarily have to be brought to the test instrument but may be measured somewhat removed from the test device. There are also at least two somewhat standardized methods of measuring scattering: the total integrated scatter (TIS) method and the bidirectional reflectance distribution function (BRDF) method. While both methods are usually set up to measure only a small area of any surface, measurements can easily be made at many places on the surface. The TIS method is the simplest approach but there is some debate as to exactly what is being measured, and it also gives no information as to whether the surface has anisotropic scattering properties. With a BRDF measurement, a great deal of information about the surface can be determined, and for some surfaces it is possible to study the entire surface at once.

One thing to keep in mind with scattering measurements is that the final result will contain not only information about the surface but also what is on the surface. In a sense this is really the desired result functionally but an optician may be upset if a surface is rejected for scatter when in fact it has not been cleaned properly before measurement. Finally it should be mentioned that the out-of-focus image method may also find use in measuring surface roughness.

Turning now to a few common sense thoughts about beauty specifications, the ideas of good figure and good beauty at the same time are largely incompatible. The type of lap the optician would like to use and his method of applying it run counter to the approach he would use to achieve good beauty. It is the mark of a superior optician to be able to achieve both at once. Luck also plays a certain role as is seen by our next comment. The larger the optic is, the more likely the optician is to get a large scratch somewhere on the surface. This is because the likelihood of getting a particle that could scratch mixed into the process is proportional (among many other things) to the area of the optic. Also such a scratch has a greater likelihood of being overlooked immediately after it has happened because the optician has so much glass to look at between each polishing run.

This brings up the question of what is to be done if a scratch is found on a large optic. (Remember to weight all these comments about beauty defects more heavily the more aspheric or faster the large optic is. This just makes the consequences that much worse.) It may require weeks if not months on a very large mirror to remove the scratch and return to the state of polish when the defect was first discovered. Even at this point the odds of picking up another scratch before the mirror is finished are still high. The point being made is that it will cost vast sums of money to correct a defect that in terms of its surface area to the total mirror area may be one part in a billion. The cost of writing down a 60 scratch specification rather than an

80 scratch specification could cost hundreds of thousands of dollars. The designer must weigh this potential cost with the impact on system performance.

Conclusion

It is hoped that by describing the method of thinking about how to apply specifications to large optics that the point has been made that each note and each tolerance on the drawing impacts fundamentally on the quality to be expected from and the cost of the finished mirror. Each different specification carries so much more weight than it does in the case of a small optic. If it happens that an error was made in a specification on a 50-mm (2-in.) lens, even if several hundred have been made before the error is discovered, the dollar and schedule impact is relatively minor. The same is certainly not the case for large optics. Each specification and each tolerance on the drawing should be the subject of a small study on that one item alone. Every time a designer calls for what appears to be a tight tolerance, study the impact of that requirement on the methods necessary for fabrication.

It is also necessary for every specification given that the designer have a well-thought-out method for measuring whether the specification has been met. This must not only be a theoretical method but a practical one, *one that can be put into practice in real hardware. Perhaps the best indication of how good a particular* specification is lies in how easily it may be verified.

References

1. Shu, Ker-Li, Appl. Opt, Vol. 22, p. 1879. 1983.
2. Malacara, D., Optical Shop Testing, Wiley, New York. 1978.
3. Parks, R. E., "Removal of Test Optics Errors," Proc. SPIE, Vol. 153, pp. 56-63. 1978.
4. Wetherell, W. B., "Effects of Mirror Surface Ripple on Image Quality," Proc. SPIE, Vol. 332, pp. 335-351. 1982.
5. F. Graham Smith, private communication, 1983.
6. Parks, R. E., "Optical Component Specifications," Proc. SPIE, Vol. 237, pp. 462-463. 1980.