Chapter 2

Basic Considerations

2.1 Preliminary Remarks

We will start with a single lens to illustrate the physics of imaging and will find the equivalence of material and geometry parameters for the image quality.

2.1.1 Optical element and wavefront propagation

Optical imaging is generally performed by lenses, that is, by pieces of glass of thickness d and two properly shaped glass–air surfaces. Such a component can be considered as a "black box," hopefully transparent, which transfers an optical input wave of amplitude $A_{\mathrm{In}}(r)$ and phase $\Phi_{\mathrm{In}}(r)$,

$$U_{\mathrm{In}}(r) = A_{\mathrm{In}}(r) * \exp[i * \Phi_{\mathrm{In}}(r)],$$

into an outgoing wave,

$$U_{\mathrm{Out}}(r) = A_{\mathrm{Out}}(r) * \exp[i * \Phi_{\mathrm{Out}}(r)],$$

where i is the imaginary unit and r is the radial coordinate normal to the light propagation direction z. We describe here the most simple case of a monochromatic wave originating from a far distant point source.

Such a lens alters the amplitude and the phase. We observe the phase change in Fig. 2.1, where a plane wave is transformed into a spherical wave. The spherical wave converges to an intensity spot at a distance f, called the focal length of the lens. Although an amplitude change results in a light intensity loss, this is ignorable here, and the phase term is much more important. It determines the optical quality, that is, the sharpness and contrast of an image. We see in Fig. 2.1 that a wavefront error of the perfect spherical output wave would result in an undesirable spot broadening.

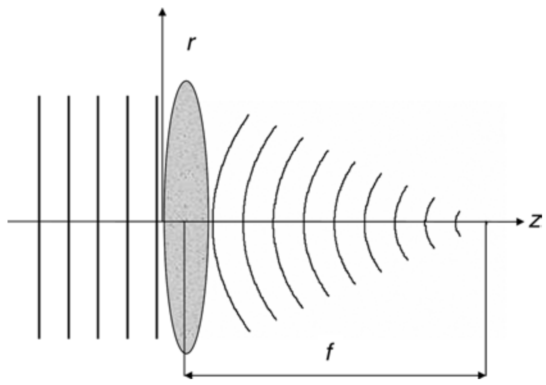

Figure 2.1 Lens action, showing an incident plane wave transformed into a spherical wave.

Let us further assume in our example an input plane wave with $\Phi_{\text{In}}(r) = 0$. Then the phase change between "in" and "out" is given by

$$\Delta\Phi_{\text{Out}}(r) = \Phi_{\text{Out}}(r) - \Phi_{\text{In}}(r) = \Phi_{\text{Out}}(r).$$

If light travels a distance z, which is exactly one wavelength λ or an integer multiple of it, the phase change is 2π. For arbitrary z in air, we obtain $\Phi_{\text{Out}}(r) = 2\pi/\lambda * z$ (r), while inside glass, with refractive index n, the wavelength is shorter, $\lambda' = \lambda/n$, and we obtain

$$\Phi_{\text{Out}}(r) = 2\pi/\lambda' * z(r) = 2\pi/\lambda * [n(r) * z(r)] = 2\pi/\lambda * \text{OP}(r),$$

where OP is the abbreviation for "optical path," $n * z$.

We are interested in how the incoming plane wave is deformed by the element, so we consider in the output plane the phase difference or, equivalently, the optical path difference (OPD) between position r and the optical axis $r = 0$,

$$\Phi_{\text{Out}}(r)/(2\pi/\lambda) = \text{OP}(r) - \text{OP}(0) = \text{OPD}(r),$$

and call this phase error also the wavefront error.

A closer look at our element shows that three physical effects contribute to Φ_{Out}: both glass–air surfaces and the glass medium (Fig. 2.2). If the surface function at the input side is $z_1(r)$ and at the output side is $z_2(r)$, if d is the axial glass thickness and n the refractive index of the glass, then

$$\text{OP}(r) = \{z_1(r) + n(r) * \{d - [z_1(r) + z_2(r)]\} + z_2(r)\}.$$

Using $z_1(0) = z_2(0) = 0$ and $n(r) = n(0)$ we obtain

$$\Phi_{\text{Out}}(r)/(2\pi/\lambda) = \text{OPD}(r) = -[n(r) - 1] * [z_1(r) + z_2(r)].$$

Note, we ignore any z dependence of the refractive index.

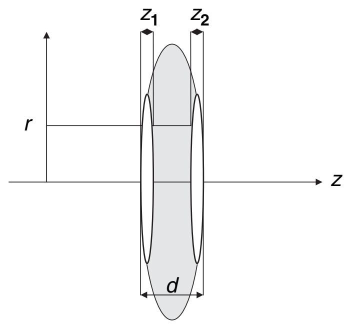

Figure 2.2 Phase contributions.

2.1.2 Optical design and tolerancing

The optical designer has to calculate the surface shape functions, the glass thickness, and has to select the glass material to obtain the "ideal" phase term Φ_{Design} to fulfil his specifications. The designer then has to specify the amount of wavefront degradation that he can accept. This allows him to tolerate quantitatively the fabrication errors.

By variation of the phase term above with respect to z_1, z_2, and n, we obtain

$$|\delta\Phi_{\text{Tol}}(r)|/(2\pi/\lambda) = |[n(r) - 1] * [\delta z_1(r) + \delta z_2(r)]| + |\delta n(r) * [z_1(r) + z_2(r)]|,$$

which has a contribution from two terms:

- The "perfect" material parameter n multiplied by surface tolerance values δz_1, δz_2 and
- Material tolerance value δn multiplied by "perfect" surface parameters z_1, z_2.

Obviously, process *and* material tolerances contribute equivalently to $\delta\Phi_{\text{Tol}}$.

2.1.3 Production and metrology errors

The production process introduces shape errors on both sides of the component and the refractive index of the delivered material shows deviations from the catalog value $\delta' n$. These errors must be properly added to the measurement errors, resulting in a forecast of the expected production errors.

The design office calculates $\delta\Phi_{\text{Prod}}$ from the reported production errors and compares it with the allowable tolerance value $\delta\Phi_{\text{Tol}}$. If the production value is greater than the tolerance value, the component is normally rejected. If it is an expensive component, the following possibilities exist:

- The component is accepted, because other components perform better than specified.

- The component can be remachined by the factory.
- Special adjustment means during system assembly may be used to compensate for the wavefront error.

2.1.4 System performance criteria

The design and tolerance work for realistic systems has to be performed at the system level, including all optical elements collectively, for many wavelengths and for many object points. This obviously leads to more complicated quality measures than the simple phase error. However, we will outline in Sec. 3.2 that all quality criteria can be traced back to physical phase errors, which we presented above for illustration purposes.

2.2 Definition of Aspherical Optical Elements

Aspherical optical surfaces deviate more or less pronouncedly from the spherical shape of standard optical surfaces. They are used in optical systems to increase imaging quality, to reduce construction size or the number of elements, to save weight, to simplify the assembly process, or to reduce the overall manufacturing costs.

Aspherical optical elements can be produced in several configurations: as one aspherical surface on a substrate (e.g., a parabolic reflector), as a combination of aspherical surfaces with spherical surfaces (e.g., aspherical lenses) or as a combination of several aspherical surfaces (bi-aspheric lenses, free shaped prisms; see http://www.olympus.co.jp/en/news/2004a/nr040126fslue.cfm).

As will be shown in Sec. 2.2.2, aspherical surfaces can be described by continuous mathematical functions. They can be rotationally symmetric, axially symmetric, or completely asymmetric (free-form surfaces). Dependent on the production volume, on the degree of asphericity, and on the required tolerance values, aspherical elements can be manufactured by a variety of production methods (Chapter 5), for example, by casting and injection molding of plastics, by blank pressing of glass or by precise machining (diamond turning of metals or polymers, grinding and polishing of metals, optical glasses, crystals, ceramics).

2.2.1 Basic characteristics of aspherical elements compared with spherical elements

2.2.1.1 Quality of the surface form

Spherical surfaces are characterized by a constant curvature value, and thus can be manufactured using large-format tools. These tools are state of the art and operate, when properly driven, over a long period of time without significant quality degradations. Additionally, the tooling heads move in a rather stochastical way, which avoids the generation of "zonal" artefacts in the surface structure. Consequently, very high form accuracies can be achieved, even with relatively simple machines.

In the case of aspheres, the local curvature changes across the surface, requiring small tooling heads for grinding and polishing. These tools are more sensitive to the deteriorations that destabilize the process. Very accurate machine kinematics and complex correction procedures are required, and the risk of generating artefacts is rather large. Additionally, very precise measuring methods with accuracies in the range 500 nm to below 1 nm are indispensable. Because several correction loops must often be performed, artificial ripples in the surface structure cannot be avoided completely and must be carefully tolerated.

2.2.1.2 Quality of the surface texture

The small-area working tools mentioned, in combination with the deterministic tool path, with little room for stochastic movements, tend to decrease the quality of the surface texture. In order to achieve the same high degree of polishing as obtained with spherical surfaces, more technical efforts are necessary. For example, grinding must be performed with smaller grain sizes, of 10 μm down to 3 μm, and with small tool pressure, leading to long working times. The polishing times are also much longer than those needed for equivalent spherical surfaces. Recent progress in polishing technology, such as magnetorheological polishing techniques (and the appropriate polishing fluids), which is used for finishing all kinds of optical surfaces, yields both high-quality surface form and texture.

2.2.1.3 Quality of positioning in optical systems

A spherical surface is, due to its rotational symmetry, uniquely described by its center of curvature. This has the advantage that a lens with two spherical surfaces has a unique optical axis, which is the line connecting both centers of curvature. The centring of such a lens, that is, the alignment of its optical axis with respect to a mechanical axis, can be performed after the manufacturing of the optical surfaces with virtually no limitation of the centring precision.

Aspherical surfaces, in contrast, have only one datum axis given by the design. In the case of rotationally symmetric surfaces, it is the symmetry axis, but in the case of asymmetric surfaces it is an axis that hits the surface at a certain point and with defined direction to the normal of the surface at this point. Therefore it must be guaranteed during the manufacturing process that the centring of the second surface with respect to the first axis is in tolerance. Thus, when combining an aspherical surface with a spherical surface, the center of curvature of the sphere must ideally lie on the axis of the asphere, but for the case of bi-aspheric lenses, both axes must ideally be collinear. The permissible deviations from the ideal case have to be specified.

Another aspect that has to be considered is that the spatial position of the axis of an asphere depends on the method with which the form deviation is estimated [1].

Due to these constraints, special techniques and equipments are necessary for the manufacturing and testing of aspherical optical elements (see Chapters 5 and 6).

2.2.2 Mathematical representation of aspherical surfaces

2.2.2.1 Basic equation according to ISO 10110—Part 12

The standard ISO 10110—Part 12 describes surface functions of second order with axial symmetry as

$$z = f(r) = \frac{\frac{r^2}{R}}{1 + \sqrt{1 - (1+\kappa)\left(\frac{r}{R}\right)^2}} + \sum_{n=2}^{m} A_{2n} \cdot r^{2x},$$

where r is the lateral coordinate, z the sagitta error, and R the paraxial surface radius. The conic constant κ is 0 for spheres, −1 for parabolas, <-1 for hyperbolas, between −1 and 0 for oblate and >0 for prolate ellipses. Details and the complete mathematical description can be found in the Chapter 10.

2.2.3 Specifying tolerances for aspherical optical elements

2.2.3.1 Surface form

Tolerancing specifies the maximum permissible deviation values of the manufactured actual form from the designed or theoretical form. Figures 2.3 and 2.4 show measured profiles. The global deviation, shown in Fig. 2.3, may be understood as the deviation of a best-fit radius from the theoretical value. This can be tolerated similarly to spherical surfaces, according to ISO 10110—Part 5, by specifying the permissible value of the sagitta error.

Rotationally symmetric deviations, as shown in Fig. 2.4, can be limited by indicating the permissible rotationally symmetric irregularity provided by this standard. Nonrotationally symmetric deviations can also be limited by specifying the permissible total irregularity.

As can be seen in the measured profiles, additional local deviations with strong gradients occur, which must be limited by an additional tolerance for the maximum allowable angular deviation of the local normal from the theoretical normal. This deviation is called “slope error” (or surface tangent error).

Parts 5 and 12 of ISO 10110 give the rules on how to indicate these form tolerances in the drawings of optical elements. The standards also specify the units of

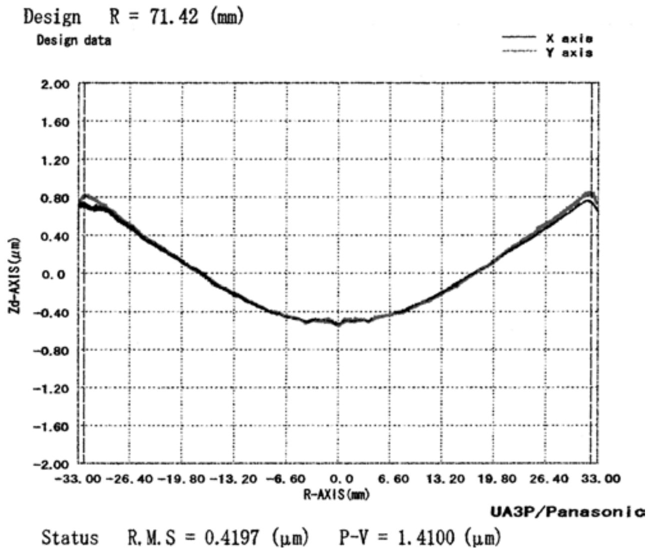

Figure 2.3 Global deviation from the specified surface shape function (rotationally symmetric asphere).

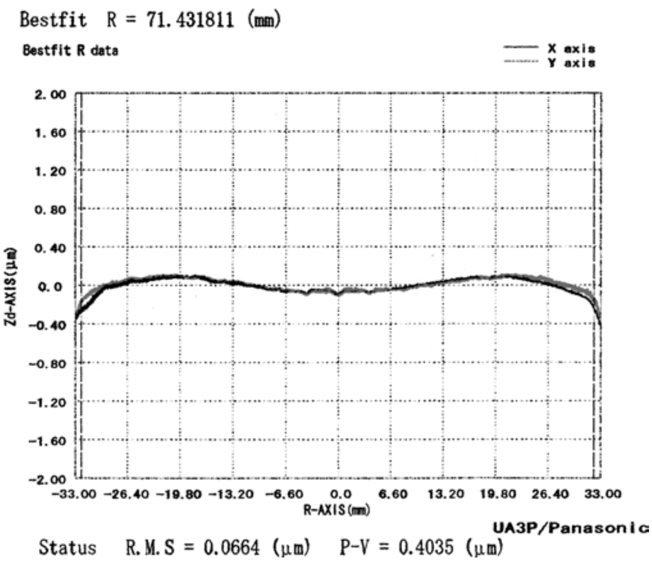

Figure 2.4 Local deviation from the best-fit radius, which varies by $+0.012$ mm from the specified value.

tolerance indications and give information on testing of optical elements, especially by interferometric methods.

Alternatively, the permissible form deviations can be specified according to ISO 1101 as tolerance zones, inside which the manufactured surfaces must be contained. The boundary surfaces of the tolerance zone are tangential surfaces to spheres, the diameters of which are the tolerance width, and the centers of which are located on the theoretical surface. As previously stated, a permissible slope error must also be indicated, to limit spatial oscillations of the surface within the boundary surfaces.

2.2.4 Surface texture

The tolerance for surface texture is indicated according to ISO 10110–Part 8. The required quality is specified by indicating one of four polishing grades. The polishing grades are related to certain maximum allowable numbers of pits in the surface, which can be detected by scanning a given distance on the surface, for example, by using a stylus of appropriately small tip radius.

The tolerancing of the surface texture of aspherical surfaces is the same as for spherical surfaces.

2.3 Drawing Indications

Figure 2.5 shows an example of a drawing of an aspherical lens element. The design equation, with its constants and coefficients, is given in the field of the drawing. The coordinate axes are indicated in the drawing. An abbreviated table with some function values is shown for information. It is especially useful to check for the correct signs of the constants and the coefficients. The indications are arranged in tabular form, according to ISO 10110-10. This prevents the drawing from being overloaded. The indications refer to the left and right surface and to the material data, given at the center of the table. The permissible form deviations are specified following the error code 3/, and tolerances for the position deviations of the surfaces follow the error code 4/.

The form tolerances of the asphere are given according to ISO 10110-5 as 3/4(0.8/0.4), which means a sagitta error of 4 fringes (@ $\lambda = 546\,\text{nm}$), a total irregularity of 0.8 fringes, and a rotational symmetric irregularity of 0.4 fringes are permissible. Because the axis of the asphere is the datum axis, no tolerance for the tilt angle is specified following error code 4/. The runout of the outer cylinder is limited to ≤ 0.005, according to ISO 1101.

For the slope tolerance, no error code exists. Therefore the tolerance is indicated as a text note in the field of the drawing, according to ISO 10110-12.

2.4 Information Exchange over Aspherical Elements

Currently, information about how to describe and characterize aspherical optical elements is produced and distributed in many different formats. This extends from using different mathematical formulas in Optic Design programs and the deduced technical drawings, to different user interfaces of measurement instruments (e.g., interferometers, profilometers) and manufacturing machines (e.g., generators,

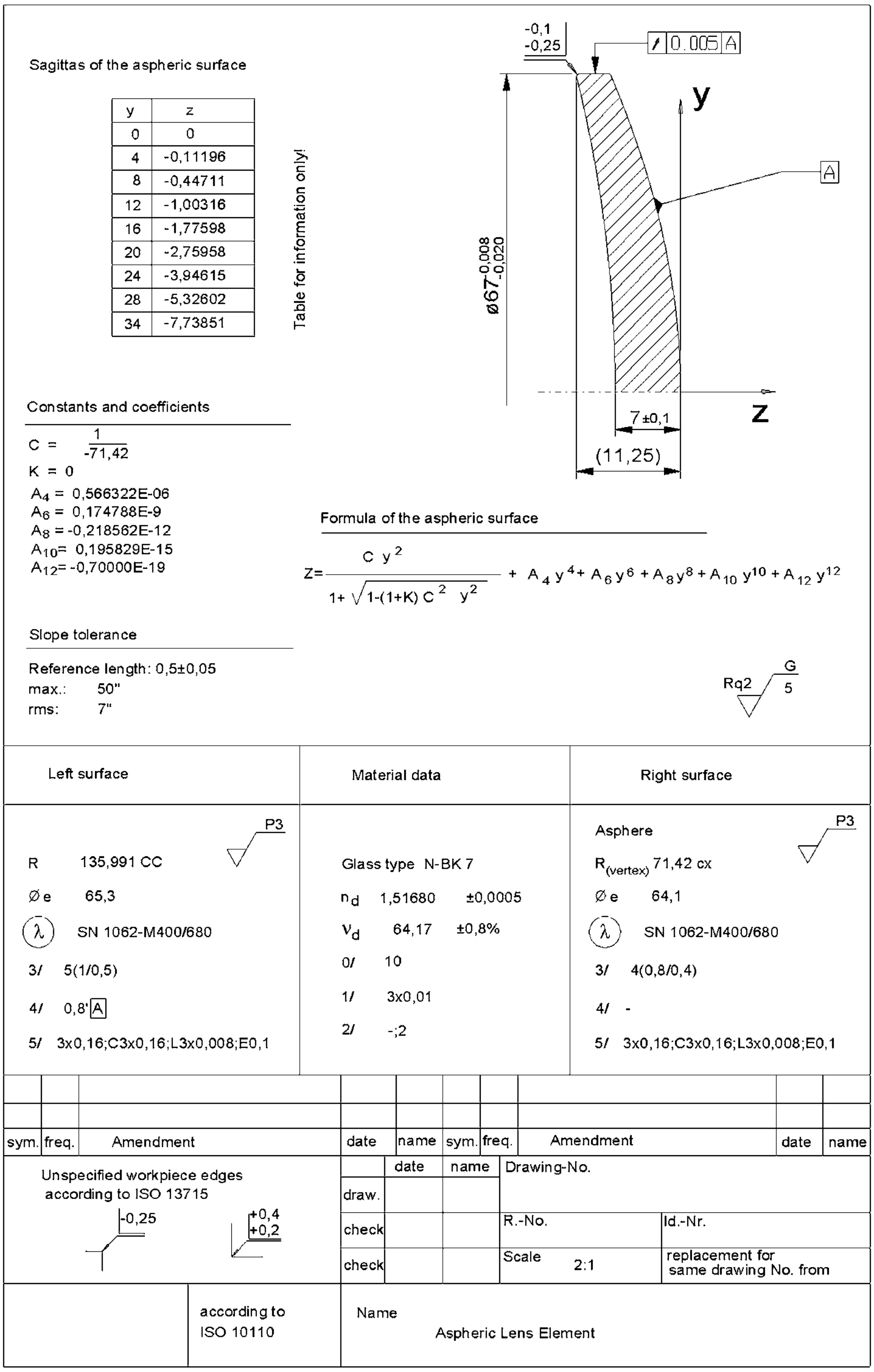

Figure 2.5 Drawing of an aspherical lens element according to ISO 10110.

polishing machines). It may be that the data in the drawing for an aspherical surface can be put directly into one device, while for another device the signs of some parameters have to be altered. Although the correct understanding and handling of the data is still under the control of companies, communication between different institutions is often difficult, time-consuming, and risky as a result of this inconsistent representation.

Unfortunately, the strict obedience of the relevant international standard ISO 10110–Part 12 in its present version does not give definitive safety, as it does not contain a sign convention. The recommendation of this standard to indicate the coordinate system and to add a numerical table of some surface function values in the technical drawing should therefore be considered.

2.5 Study about Surface Errors

In the following, we want to demonstrate the importance of the comments made above, mainly the need to communicate extensively between design and production. It will be shown later, in Chapter 5, that a large variety of cost-attractive manufacturing technologies exist today. However, each method has its performance limits, which lead unavoidably to residual surface deviations from the ideal form. The designer is well advised to know these limits in advance. This enables him to judge how reliably his specifications can be realized by the fabrication process.

2.5.1 Aspherical laser collimator

We consider an application with one plano-convex aspherical lens. The lens should image the emitting area of a laser diode to infinity (i.e., to collimate the laser beam in Fig. 2.6). The lens has a focal length F = 30 mm and a f-number of 1.8, equivalent to a free aperture diameter of about 16 mm.

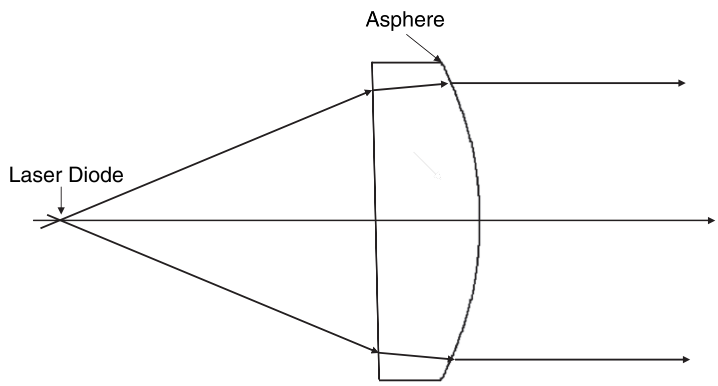

Figure 2.6 Aspherical lens for laser collimation.