Real Ray Tracing

Josep Arasa
University Polytechnique of Catalonia, Terrassa, Barcelona, Spain

Javier Alda
University Complutense of Madrid, Madrid, Spain

INTRODUCTION

Real ray tracing is an essential step to obtain numerical quality performances in optical systems and is also involved in its optimization procedure. Real ray tracing does a complete description of the light ray for a selected number of rays; moreover, real ray tracing mixes the intuitive approach of ray propagation with the right light. Real ray tracing provides a powerful tool to reach, at the same time, information about an individual light trajectory and about all the optical system.

Real ray tracing is based on the knowledge of Snell law and the geometric description of the surfaces used to specify the boundaries of different media glass. Both calculations are difficult to do; so in case of exact calculation, the reliability of a given algorithm or program is strongly based on how many rays are traced from the input to the output plane: the larger the number of rays, the better the results. Although the foundations of ray tracing are the simple rules of geometrical optics, the scientists and engineers involved in optical design have made successive refinements to include both the energy carried out by the light and also the wave nature of the electromagnetic radiation.

In this contribution, we present the rules of real ray tracing step-by-step; these rules are explained in a clear language and same examples are included for the first-time readers. Real ray tracing is presented in a self-explanation form to make this contribution readable in itself. The Snell law is revisited and adapted for its implementation in a numeric algorithm. The spot diagrams resulting from a real ray-tracing calculation are discussed. The wave nature of light propagation can be also approximated by using real ray tracing. Therefore we show the method to calculate a geometrical wave front and also its intrinsic limitations. Linked with the spot diagrams, it is possible to evaluate a point spread function. Also, an optical transfer function and a modulation transfer function are defined within this geometrical approach. The evaluation of the flow of energy through the system is discussed within the ray-tracing framework. Some hints about the inner mechanisms used by the optimization procedures of optical systems are discussed and related with a basic classification of the software packages currently used.

REAL RAY TRACING

Vectorial Form of the Snell Law

The real ray tracing is based on the application of the Snell law.[1–4] To treat this problem properly, a special formulation of the Snell law is necessary. The starting point is the coplanarity law of geometrical optics. It establishes that the input ray, the refracted ray, and a vector normal to the surface at the point of incidence are located in the same plane. This plane is also named as the plane of incidence. Within this plane (Fig. 1), without loss of generality, the Snell law is written as

$$n\sin(\theta) = n'\sin(\theta') \tag{1}$$

This expression is valid to trace the trajectory of any refracted ray. However, when practicing with real ray tracing, it is often more efficient to rewrite this expression in such a way that the prior knowledge of the incidence plane is not necessary. To do that, we calculate the cross product of two unitary vectors; one of them is pointing along the direction of the input beam, and the other is normal to the interface at the incidence point and pointing toward the medium where the light is coming from (Fig. 2). The modulus of this cross product is sin (θ)

$$|\vec{i}\times\vec{n}| = |\vec{i}||\vec{n}|\sin(\theta) \tag{2}$$

Then, the Snell law can be written as

$$n|\vec{i}\times\vec{n}| = n'|\vec{o}\times\vec{n}| \tag{3}$$

Real Ray Tracing Through an Interface Separating Dielectric Media

In this section, we consider that the media supporting the light trajectories are dielectric, linear, homogeneous, and

DOI: 10.1081/E-EOE 120027488

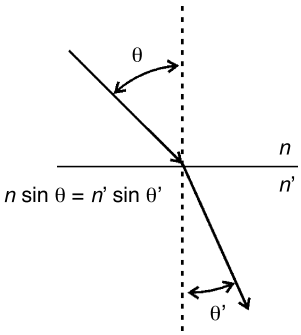

Fig. 1 Graphical representation of the Snell law in the plane of incidence. Here, a ray is refracted from a media having an index n to another media having an index n', where $n'>n$.

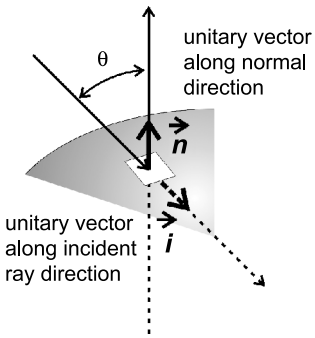

Fig. 3 Unitary vectors used to rewrite the Snell law in its vectorial form.

isotropic. The radiometric flux associated with the ray is not attenuated in the media. The radiometric balance in the interface is not taken into account as far as we are only interested in the location of the light trajectories refracted through the interfaces.

Under the previous conditions, a light ray is represented by a straight line plus a direction of propagation. To make the calculus easier, it is much better to use the parametric form of the straight-line equation:

$$\frac{x - x_0}{u_i} = \frac{y - y_0}{v_i} = \frac{z - z_0}{w_i} = t \tag{4}$$

where $u_i^2+v_i^2+w_i^2=1$. This previous equation can also be written in matricial form as:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} u_i \\ v_i \\ w_i \end{pmatrix} t - \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \tag{5}$$

where x_0, y_0, and z_0 are the coordinates of a point that belongs to the straight line. u_i, v_i, and w_i are the direction cosines of the straight line and t is the free parameter. This parametric form is very efficient because the use of the direction cosines speeds up the calculation of the cross product.

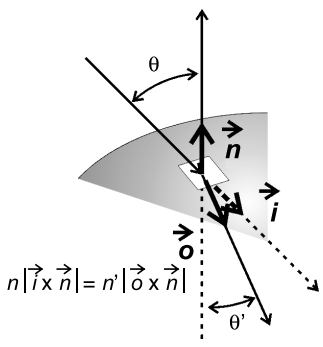

Fig. 2 Three-dimensional representation of the Snell law for a ray that is refracted from a medium having an index n toward a medium having an index n', where $n'>n$.

The next step is to find the intersection between the previous straight line characterizing the input ray and the surface representing the interface. When the interface is a spherical surface or any other closed surface, special attention has to be paid to consider the part of the surface actually interacting with the input ray. The intersection point is characterized by its coordinates: x_s, y_s, and z_s.

Another piece of the calculation is the equation of the straight line perpendicular to the interface at the intersection point previously obtained. This straight line is also expressed in its parametric form as:

$$\frac{x - x_s}{u_n} = \frac{y - y_s}{v_n} = \frac{z - z_s}{w_n} = t' \tag{6}$$

where $u_n^2+v_n^2+w_n^2=1$. This equation can also be given in matricial form as:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} u_n \\ v_n \\ w_n \end{pmatrix} t' - \begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} \tag{7}$$

where x_s, y_s, and z_s are the coordinates of the intersection point, u_n, v_n, and w_n are the direction cosines of the normal straight line, and t' is the free parameter.

Once we have the unitary vectors characterizing the input ray and the line perpendicular to the surface at the intersection point (Fig. 3), it is possible to calculate the sine of the angle of incidence as the modulus of the cross product of these two unitary vectors.

$$\sin(\theta) = (v_i w_n - w_i v_n)^2 + (w_i u_n - u_i w_n)^2 + (u_i v_n - v_i u_n)^2 \tag{8}$$

The refracted ray also passes through the intersection point (x_s, y_s, z_s). The straight line characterizing this refracted ray is:

$$\frac{x - x_s}{u_o} = \frac{y - y_s}{v_o} = \frac{z - z_s}{w_o} = t'' \tag{9}$$

where $u_o^2+v_o^2+w_o^2=1$. This relation can be written in matricial form as follows:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} u_o \\ v_o \\ w_o \end{pmatrix} t'' - \begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} \tag{10}$$

where x_s, y_s, and z_s are the coordinates of the point of incidence on the interface, u_o, v_o, and w_o are the direction cosines of the straight line, and t'' is the free parameter.

The sine of the refracted ray is given by the cross product of the unitary vectors characterizing the refracted ray and the normal ray (Fig. 2). These vectors are given by (u_o, v_o, w_o) and (u_n, v_n, w_n), respectively, and produce the following result,

$$\sin(\theta') = (v_o w_n - w_o v_n)^2 + (w_o u_n - u_o w_n)^2 + (u_o v_n - v_o u_n)^2 \tag{11}$$

Our goal is to obtain the equation describing the refracted ray. This ray passes through the point of incidence (x_s, y_s, z_s), and we only need to know the coordinates of the vector (u_o, v_o, w_o). These three unknowns need three equations to be found. The first one is given by the Snell law, $n\sin(\theta)=n'\sin(\theta')$, where $\sin(\theta)$ has been previously calculated and $\sin(\theta')$ is given by the last equation. A second equation is obtained from the coplanarity condition of the input ray, refracted ray, and normal ray. This condition is formulated by canceling the scalar product between the unitary vector characterizing the refracted ray, (u_o, v_o, w_o), the unitary vector obtained as the cross product of the unitary vector characterizing the input ray, (u_i, v_i, w_i), and the unitary vector perpendicular to the surface at the intersection point, (u_n, v_n, w_n),

$$u_o(v_i w_n - w_i v_n) + v_o(w_i u_n - u_i w_n) + w_o(u_i v_n - v_i u_n) = 0 \tag{12}$$

This condition can also be written as:

$$\begin{vmatrix} u_o & v_o & w_o \\ u_i & v_i & w_i \\ u_n & v_n & w_n \end{vmatrix} = 0 \tag{13}$$

The last equation is easily found when normalizing the unitary vector characterizing the refracted ray: $u_o^2+v_o^2+w_o^2=1$.

Unfortunately, as far as some of the variables are squared, the solution of these three equations is not unique. They produce four possible solutions (Fig. 4). The right solution is that one complying these two conditions: 1) it produces the highest value of the scalar product of the unitary vector representing the refracted beam and the unitary vector representing the input beam, and 2) the scalar product of the unitary vector characterizing the refracted ray and the unitary vector perpendicular to the surface is negative.

The procedure described in this section has to be applied at each interface of the optical system under analysis.

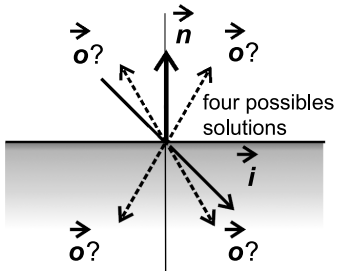

Fig. 4 Choice of the right solution for the refracted ray.

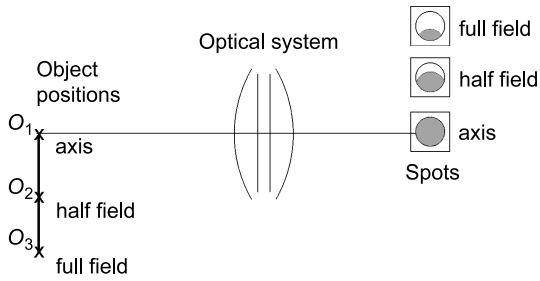

Fig. 5 The number of rays used to analyze and optical system varies when the object point changes. Usually, the image quality degrades when moving from the center to the edge of the field of view, and the number of involved rays also decreases.

Spot Diagrams in the Pupil Plane and in the Image Plane

The information provided by ray-tracing a single ray is clearly not enough to obtain a reliable description of the performance of a given optical system. A large and significant number of rays need to be traced along the system. Typically, a minimum bound for the number of rays is around 100 for systems showing a symmetry of

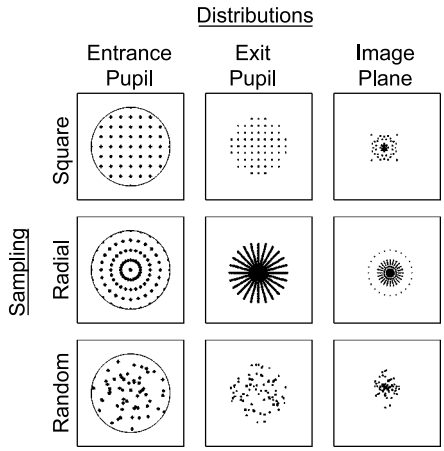

Fig. 6 Variation of the spot diagram at the exit pupil and at the image plane with respect to the method used to generate and sample the entrance pupil.

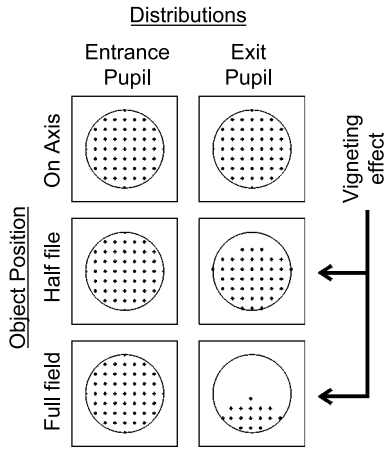

Fig. 8 Evidence of vignetting in a given optical system. The vignetting is related with a failure in the location and size of diaphragms and windows in the system.

revolution and around 1000 for systems without this symmetry. However, depending on the method of generating the rays used in the analysis, the total number of rays may be well increased by 1 or several orders of magnitude.

One of the simplest and most useful methods to choose those rays under analysis is to group them depending on the departure point. Each set of rays corresponds with the same object point. They are well suited to produce a spot diagram with clear meaning. The location of these spot diagrams is usually calculated at two important planes: the plane of the exit pupil and the image plane of the optical system.

The spot diagram is the representation of the coordinates of the intersection of the calculated rays with the selected plane. Obviously, the usefulness of these diagrams of impacts will strongly depend on the choice of the rays traced through the system. Following the previously mentioned method to group the rays by sharing the departure point, it is usual to select a few departure, or object, points to generate rays and produce meaningful spot diagrams.[5] One of these object points is usually located on the optical axis, another one is placed at the most extreme point in the field of view, and a third one is usually located in the middle point between the two previous object points (Fig. 5).

A useful sampling is depending not only with the number of rays used in the calculation, but also with the method to generate them. The ray-generating strategies can be structured, in a squared grid or in a radial grid, or random (Fig. 6). Every method has to fill the entrance pupil to obtain undistorted information about the system (Fig. 7).

There are three parameters that typically characterize the spot diagram obtained after a sound ray tracing at the image plane. These three parameters are the maximum extension, the averaged distance of every point to the centroid of the diagram, and a symmetry factor usually related with the eccentricity of the spot diagram. These parameters analyze the size of the image, the dimension of the spreading of the image, and certain degree of symmetry.

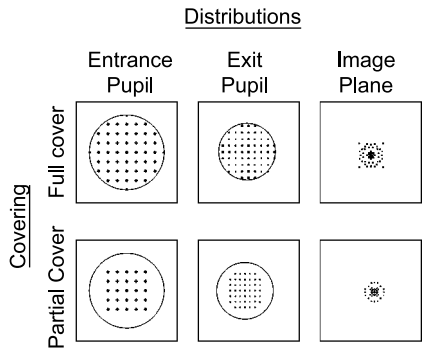

Fig. 7 Dependence of the quality of the spot diagram at the exit pupil and at the image plane with respect to the filling factor of the entrance pupil.

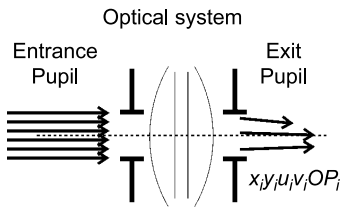

Fig. 9 Compiling of the necessary information used to obtain the wave front at the exit pupil (x_i, y_i, u_i, v_i, and OP_i).

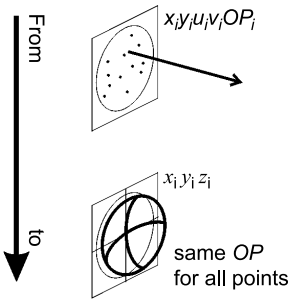

Fig. 10 Location of the wave front obtained from the information compiled at the exit pupil plane.

When the spot diagrams are obtained at the exit pupil plane, we are mainly interested in the uniformity of its distribution and in the filling of the exit pupil. A uniform distribution will provide sharp images or uniform defocused images. When the exit pupil is not homogeneously filled, we obtain information about the presence of vignetting in the system (Fig. 8).

Wave Front Evaluation and Aberration Fitting

The distribution and form of the spot diagrams provide visual and qualitative information about the performance of a given optical system. However, an appropriate analysis of the information given by the ray tracing may help to obtain quantitative information.

The wave front is defined as the locus in the space where the optical path measured from a given object point is constant. The wave front concept is intrinsically linked to the description of a continuous surface in the space.[6] Therefore any method used to obtain such a surface from a finite number of rays will face limitations and constraints. On the other hand, the wave front is usually evaluated at the exit pupil plane. Therefore the information obtained from the real ray tracing cannot be directly identified with the wave front, and vice versa.

To evaluate the wave front from the ray-tracing data, we need to account for the optical path of every ray from its departure point to the arrival to the plane of interest, usually the exit pupil plane. The information compiled for each ray is its location, its direction, and the optical path previously calculated (Fig. 9). To obtain the wave front is necessary to move back and forward a given distance to produce the same value of the optical path. The result is a collection of points (x_i, y_i, z_i) belonging to the wave front (Fig. 10).

Most of the times, a constant value is subtracted to the optical path defining a given wave front. This constant is usually the optical path of the ray passing through the center of the entrance pupil of the system. We should emphasize at this point that the previous wave front is evaluated on a geometrical sense only. It does not see any phase shift produced at the interface between media, and it is not possible to obtain diffractional information about the optical system. The wave front has been calculated using a discrete, nonregular grid. Therefore it does not allow an analytic representation. To do that, we fit the wave front to a given set of functions evaluated at a given regular grid. A very useful polynomial base is the Seidel polynomic set.[7] They are given as:

$$\phi(r, \phi, \eta') = {}_i w_{jk} \eta'^i r^j \cos^k \theta \tag{14}$$

where r and θ are the polar coordinates at the exit pupil of the system, η' is the normalized value of the field of view where the function is evaluated, and ${}_i w_{jk}$ are the aberration coefficients (Fig. 11). The first five correspond with the third-order aberration coefficients.

This base is directly related with an analytical representation of the aberration content of the optical system. By choosing a given order in the Seidel base, it is possible to quantify how far the real wave front is from the paraxial

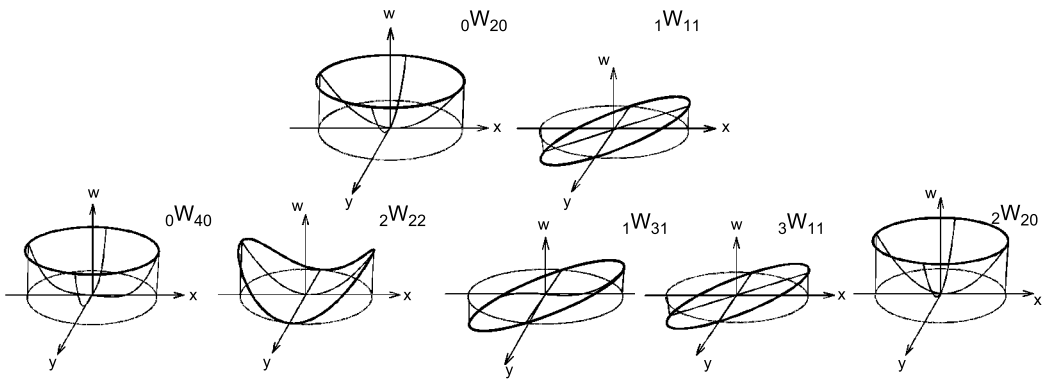

Fig. 11 Graphical representation of the first and third aberration coefficients—first order: longitudinal focal shift ($_0W_{20}$) and transverse focal shift ($_1W_{11}$); third order: spherical ($_0W_{40}$), astigmatism ($_2W_2$), coma ($_1W_{31}$), distortion ($_3W_{11}$), and field curvature ($_2W_{20}$).

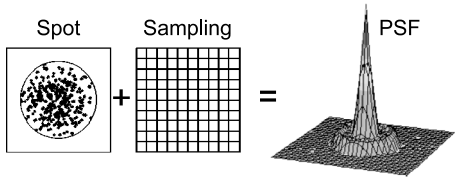

Fig. 12 The PSF is obtained from the sorting of the impacts on the image plane using a regular sampling grid.

behavior. The aberration coefficients characterize the whole optical system, but it cannot provide information about the importance of these coefficients at each one of the optical surfaces of the system.

Another possible representation is the monomial given by

$$W(z,y) = \sum_{n=0}^{n=k} \sum_{m=0}^{m=n} c_{ij} z^m y^{n-m} \tag{15}$$

This base is easy to fit and its use is widespread. However, it cannot relate the monomial coefficients with the fulfillment of the paraxial expectations of the optical system.

The Zernike polynomial base is probably one of the most widely used among the optics community. The Zernike base is expressed as,

$$W(r,\theta) = \sum_n \sum_m A_{nm} R_n^m r^k \cos^l \theta \tag{16}$$

where

$$R_n^m(r) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n+m}{2}-s\right)!} r^{n-2s} \tag{17}$$

The Zernike expansion allows very efficient fittings of the wave front. Some of the low-order Zernike components are directly related with the Seidel coefficients.[8] However, the Zernike base needs a regular sampling grid that has to be interpolated from the nonregular evaluation grid obtained from the real ray tracing. This interpolation procedure is affected by a loss in the accuracy of the results.

Geometrical Approximation to the Point Spread Function and to the Modulation Transfer Function

The point spread function (PSF) is defined by the image of a point source given by the system. This image is quantitatively described by its irradiance distribution. A first reading of this definition may indicate that ray tracing is not able to produce this result. Fortunately, there is a method to properly obtain the PSF from the diagram of impacts at the image plane. A basic step is to define a regular sampling grid on the image plane where equal size bins are located to fill the whole plane. Then, a large number of rays are generated with a random sampling procedure. If we consider that all the rays carry the same amount of energy, the number of rays contained in each bin, normalized with the total number of rays, produces a quantitative measure of the PSF (Fig. 12).

It is also possible to obtain the same distribution of irradiance by using a regular sampling at the entrance pupil. The number of rays has to be large enough to obtain a faithful PSF. A method to limit somehow the number of rays used in the calculation is to successively increase this number by an order of magnitude until the PSF profile stabilizes. The criterion to consider that the PSF has been stabilized is user-dependent.

The PSF is a very useful function to characterize the quality of an optical system. However, it is based on some basic assumptions that need to be fulfilled to extend properly its application. One of the most important is the isoplanatism requirement. This means that the PSF profile has to remain about the same for any location of the point source in the object plane. On the other hand, an optical system behaves as an isoplanatic system when its quality is very high. For these high-quality optical systems, the geometrical PSF does not provide meaningful information because the system is very close to the diffraction-limited case. Most of the optical systems are usually far away from the diffraction-limited case, and therefore they cannot be considered as isoplanatic. In these cases, the

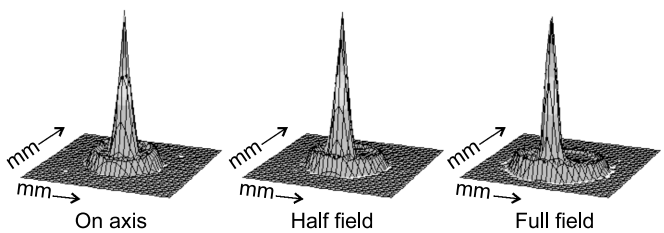

Fig. 13 Variation in the shape of the PSF obtained from the ray tracing for three different point sources.

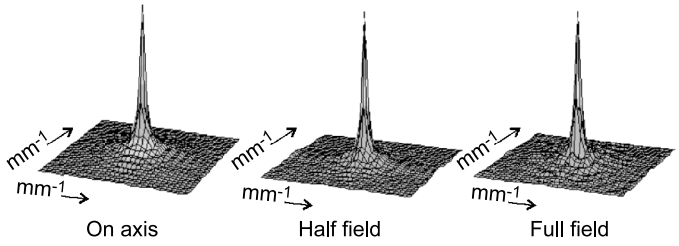

Fig. 14 Optical transfer function obtained from the previous results.

best solution is to obtain the geometric PSF for a collection of point sources on the object plane. Three points are usually enough (Fig. 13): one at the optical axis, one at the extreme of the field of view (full field), and a third one in a middle point in between the previous two points (half field).

Another useful function characterizing the system is the optical transfer function (OTF) and its modulus usually named as the modulation transfer function (MTF). The simplest way to evaluate the OTF is by Fourier-transforming the PSF. These two new functions (OTF and MTF) are affected by the same limitations and constraints than the PSF when evaluated from a ray-tracing calculation. This is because they only change the space of representation (Fig. 14).

The MTF provides information about the contrast losses produced by an optical system. A very useful calculation is to represent, as a function of the field of view, the contrast loss at some given spatial frequencies. As far as it is not possible to represent in a clear graph this contrast loss for every direction where the OTF may be evaluated, it is useful to restrain the analysis to two privileged directions: the tangential and the sagittal planes (Fig. 15).

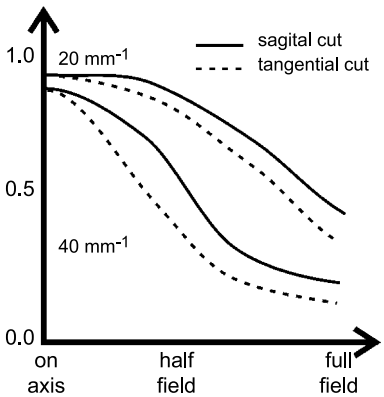

Fig. 15 Tangential and sagittal loss of contrast as a function of the field of view for two values of the spatial frequency: 20 and 40 mm^{-1}.

Energy Calculations by Using Real Ray Tracing

Ray tracing is intrinsically unable to produce energy calculations. It works very well when evaluating the trajectory of light. However, in the last section, we already consider a way to treat the energy transfer by attaching a constant value of energy to every traced ray. Then, by analyzing how the impact of rays has been distributed on the plane of interest, some insight in the energy distribution can be gained. This approach can be refined to improve such flux-of-energy calculation. A first improvement is made by assigning different energy values to each ray. Then, it is possible, for example, to simulate a Gaussian beam profile (Fig. 16). In this case, the normalization procedure is performed by using the total energy actually entering into the system.

A critical improvement in the energetic calculation is obtained when accounting for the energy balance at each interface. This balance depends on the angle of incidence, the indices of refraction of both media, and the state of

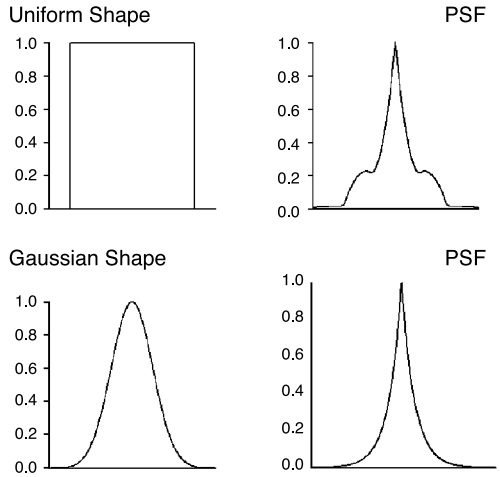

Fig. 16 The PSF changes when the energy distribution changes in the input beam. Typically, a Gaussian beam at the input produces narrower PSF than a uniform distribution.

polarization. The physical magnitudes describing this behavior are the transmittance and the reflectance:

$$T_{\parallel} = \frac{4\cos^2\varepsilon \times \sin^2\varepsilon'}{\sin^2(\varepsilon + \varepsilon') \times \cos^2(\varepsilon - \varepsilon')}; R_{\parallel} = 1 - T_{\parallel}$$

$$T_{\perp} = \frac{4\cos^2\varepsilon \times \sin^2\varepsilon'}{\sin^2(\varepsilon + \varepsilon')}; R_{\perp} = 1 - T_{\perp} \quad (18)$$

Ray tracing does not deal with the state of polarization of light. Therefore the usual solution is to consider the mean of the transmittance and transmittance for both polarizations:

$$T = (T_{\parallel} + T_{\perp})/2$$
$$R = (R_{\parallel} + R_{\perp})/2 \quad (19)$$

In the previous case, we conclude that changing the irradiance profile, the PSF and MTF also changed. Now again, the PSF and MTF significantly change when including the energy balance at the interfaces (Fig. 17).

The effect of the energy balance and the spatial distribution of energy obtained at the image plane and characterized by the PSF and the MTF can also be evaluated at the exit pupil plane. The representation on this plane is rather similar to that one used for the wave front.[9] To properly show all the features of interest, it is usual to present two graphs of the properties obtained from the ray tracing on the exit pupil plane (Fig. 18). One of the graphs represents the wave front, and the other shows the spatial distribution of energy as a grey-scale representation or by using points with size proportional to the irradiance distribution.

Although the state of polarization is not usually included in real ray tracing, it is possible to describe it as a couple of orthonormal vectors attached to each one of the rays. This strategy makes it possible to deal with polarizing media. On the other hand, it is possible to assign a geometrical length of coherence that allows to treat interference problems under a ray-tracing formalism. All these refinements need to be included in the very first stages of the ray-tracing calculation and also at the origin of the rays in the object plane. However, it is not possible to avoid the geometrical character of the ray, and therefore accurate diffractional analysis by using real ray tracing is intrinsically limited.

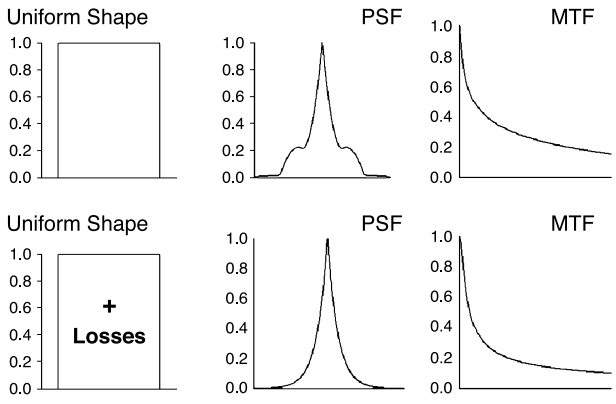

Fig. 17 The loss of energy due to the interfaces changes the form and value of the PSF and the OTF. The effect of this loss of energy is not merely a multiplicative factor.

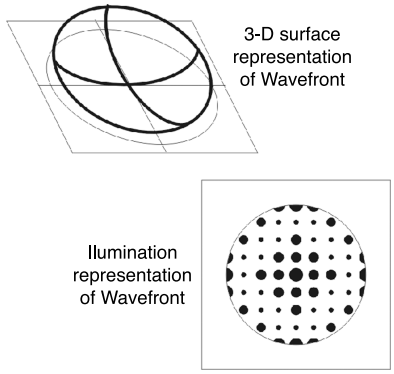

Fig. 18 Splitting of the information on the exit pupil. The wave front shape (left) and the spatial distribution of energy (right) are usually represented in different plots.

Optimization Procedures: Figures of Merit and Definition of Variables and Constraints

To make a sound optimization[10] of an optical system, it is necessary to identify its constitutive elements: the variables and their constraints, the figures of merit, the target value of the figure of merit, and the optimization algorithm and strategy. All these elements constitute the optimization procedure.

First of all, we should remind that any optimization procedure does not anticipate the optimum result of the system; we only know the result that we desire to obtain.

To begin the optimization, it is necessary to select and define the free variables. A free variable is a physical parameter of the optical system (radius of curvature, thickness, materials, location of the object and the image, etc.) that is allowed to vary within a given range during the optimization procedure. Not all the variables of an optical system are free variables. We can define two more categories of variables: the dependent variables and the tied variables. A dependent variable follows a known relation with another free variable. A tied variable can be used to maintain a given specification of the system, e.g., the value of the focal length of the system.

After defining what parameters are free, dependent, or tied, it is necessary to quantitatively specify what has to be the quality of the optimized optical system. This is one of the most critical stages of the optimization procedure. The quality of the optical system is defined by some characterizing functions and figures. Some examples are the wave front as a function depending with the aberration composition, the image quality defined in terms of the PSF or the OTF, or any other combination of these functions or other parameters quantifying the performance of the optical system. A wise definition of the figure of merit is very important. We have to balance the accuracy of the description and the feasibility of optimization. For example, by describing the quality of the system as the PSF evaluation at 10 different points, we would obtain a very precise description of the behavior of the system; however, the optimization of such a figure of merit would be impracticable. On the contrary, if a relaxed figure of merit is defined, e.g., by specifying a single parameter given by the first coefficient of the Seidel aberrations, the result of the optimization procedure will be so loose that it will be hard to find the targeted solution.

After selecting the merit function, it is necessary to fix the acceptable value of this function. The definition of this target will depend on the selection of the merit function and the expectations about the performance of the system.

When using real ray tracing, a typical function to describe the quality of the system is obtained by selecting a few object points where the rays are departing from. Then, the diagram of impact on the image plane is analyzed and four parameters are defined as:

$$Q_1 = \frac{1}{N\delta_R}\sum_{i=1}^{N}\Delta d_i \qquad Q_2 = \frac{1}{\delta_R}\sqrt{\sum_{i=1}^{N}\Delta d_i^2}$$

$$Q_3 = \frac{\delta_R}{N}\sqrt{\frac{1}{\Delta d_i + \delta_R}} \qquad Q_4 = \delta_R\sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{(\Delta d_i + \delta_R)^2}} \tag{20}$$

where d_i is the distance between the location of the impact for the ray i and the location of the centroid for the whole diagram of impacts. N is the total number of rays traced and δ_R is the spatial resolution expected for the system and defined using the Rayleigh criterion. These four coefficients (Q_1, Q_2, Q_3, and Q_4) are calculated for each object point. When $Q_1=Q_2=0$ and $Q_3=Q_4=1$, the optical system is considered as an aberration-free system. When the aberrations are present and a Strehl ratio of 0.73 is considered as an acceptable value, the individual values of these parameters are $Q_1=0.47$, $Q_2=0.66$, $Q_3=0.17$, and $Q_4=0.18$. Therefore the targeted values of optimization will be considered as those previously mentioned.

Another important choice is the type of optimization algorithm that will be used along the process.[11–13] At this point, it is important to remind that ray tracing is not providing analytic relations among variables. This analytical approach is possible when using Seidel aberrations. Therefore the functions and figures of merit are of numerical nature. This fact limits the type of algorithm usable for optimizing the results. They will be of genetic type, and some examples are those based in random sampling, global searching, or evolving. The simplest algorithm applied for optical system optimization is the simplex one, which also finds application in some other fields.

A basic flow graph for the optimization of an optical system by ray tracing is shown in Fig. 19.

There are a wide variety of software packages in the marketplace devoted to the analysis and optimization of optical systems. Not all the available packages perform all the types of calculation shown in this contribution. Then, as a simple guide for the reader, we may classify the

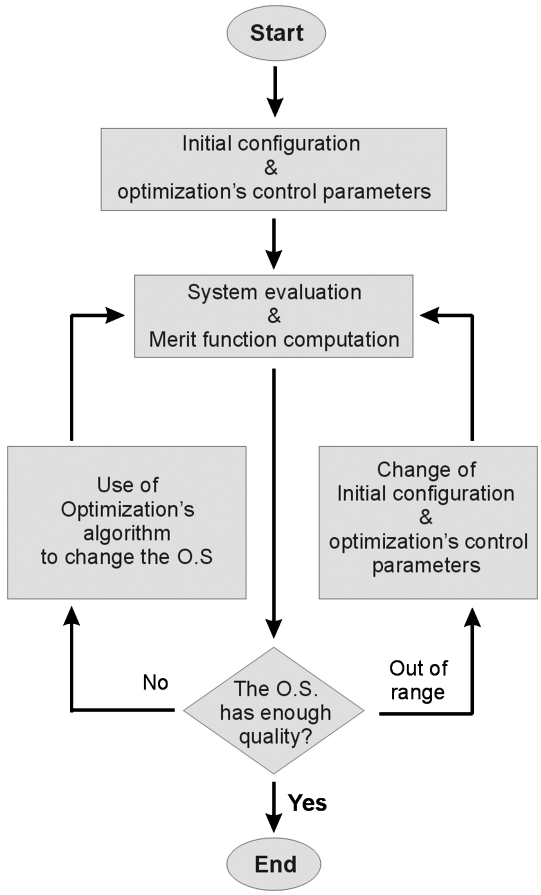

Fig. 19 Basic flow chart for optimizing an optical system (O.S.).

software packages into three different groups: the first one contains those packages devoted to the geometrical ray tracing, some of them may include PSF or OTF calculations; there is a second group that also performs calculations about the flux of energy through the system; a third group is formed by those packages able to optimize the optical system performance. The difference between the second and third group is sometimes difficult to define because some packages of the second group contain simple optimization tools, and some optimization packages of the third group include energetic evaluations. The most important factor to properly select a ray-tracing package is to keep always in mind the specific application of the package. Then, we will find that it is not always necessary to have a very sophisticated optimization procedure simultaneously working with a high-performance energetic evaluation.

CONCLUSION

In this article, we have revisited, using a feasible structure, the concepts and rules involved in real ray tracing. Real ray tracing has been presented as a powerful tool to model the actual trajectories of light when the paraxial approach is surpassed. The analysis of the outputs of real ray tracing has allowed the definition of geometric versions of functions usually defined within diffractional models, e.g., the wave front surface, the point spread function, the optical transfer function, and the modulation transfer function. When energetic considerations are included in the analysis, it has been possible to analyze losses and another kind of effects. The main ideas inspiring the optimization procedures have been described at the end of the article along with a basic classification of software packages currently used in optical design.

REFERENCES

1. Mouroulis, P.; Macdonald, J. Gaussian Optics. In *Geometrical Optics and Optical Design*; Oxford University Press: New York, 1997. Chapter 4.2.
2. Fincham, W.H.A.; Freeman, M.H. *Optics*; Cambridge University Press, 1974.
3. *Military Handbook 141, Optical Design*; US Department of Defense: Washington, DC, 1962.
4. Laikin, M. *Lens Design,* 3rd Ed., expanded and revised; Marcel Dekker, 2001.
5. Smith, W.J. *Modern Optical Engineering,* 3rd Ed.; McGraw-Hill: New York, 2000.
6. Mahajan, V.N. *Selected Papers on Effects of Aberrations in Optical Imaging*; SPIE Press, 1994. (V MS 74).
7. Born, M.; Wolf, E. *Principles of Optics,* 7th (extended) Ed.; Cambridge University Press, 1999.
8. Tyson, R.K. Conversion of Zernike aberration coefficients to Seidel and higher-order power-series aberration coefficients. Opt. Lett. **1982**, *7*, 262.
9. Pizarro, C.; Arasa, J.; Royo, S.; Tomàs, N. Inclusion of energetic features in wavefront descriptions. SPIE Proc. **2000**, *4093*, 261–271.
10. Hearn, G.K. The evolution of optimization algorithms. SPIE Proc. **1992**, *CR41*, 54–65.
11. Goldberg, D.E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley Publishing Company, Inc., 1989.
12. Michalewicz, Z. *Genetic Algorithms + Data Structures = Evolution Programs*; Springer-Verlag, 1996.
13. Yamamoto, K.; Chen, X. Genetic algorithm and its application in lens design. SPIE Proc. **1996**, *2863*, 216–221.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016