OPTICAL AND ELECTRO-OPTICAL ENGINEERING SERIES

MODERN LENS DESIGN

A Resource Manual

WARREN J. SMITH
GENESEE OPTICS SOFTWARE, INC.

ROBERT E. FISCHER & WARREN J. SMITH, Series Editors

Modern Lens Design

A Resource Manual

Warren J. Smith
Chief Scientist
Kaiser Electro-Optics, Inc.
Carlsbad, California

Genesee Optics Software, Inc.
Rochester, New York

McGraw-Hill, Inc.
New York St. Louis San Francisco Auckland Bogotá
Caracas Lisbon London Madrid Mexico Milan
Montreal New Delhi Paris San Juan São Paulo
Singapore Sydney Tokyo Toronto

Library of Congress Cataloging-in-Publication Data

Smith, Warren J.
Modern lens design : a resource manual / Warren J. Smith and Genesee Optics Software, Inc.
p. cm.—(Optical and electro-optical engineering series)
Includes index.
ISBN 0-07-059178-4
1. Lenses—Design and construction—Handbooks, manuals, etc.
I. Genesee Optics Software, Inc. II. Title. III. Series.
QC385.2.D47S65 1992
681'.423—dc20 92-20038
CIP

1 2 3 4 5 6 7 8 9 0 DOC/DOC 9 8 7 6 5 4 3 2

ISBN 0-07-059178-4

The sponsoring editor for this book was Daniel A. Gonneau, the editing supervisor was David E. Fogarty, and the production supervisor was Suzanne W. Babeuf. It was set in Century Schoolbook by McGraw-Hill's Professional Book Group composition unit.

Printed and bound by R. R. Donnelley & Sons Company.

Contents

Preface

This book had its inception in the early 1980s, when Bob Fischer and I, as coeditors of the then Macmillan, now McGraw-Hill, Series on Optical and Electro-Optical Engineering, were planning the sort of books we wanted in the series. The concept was outlined initially in 1982, and an extensive proposal was submitted to, and accepted by, Macmillan in 1986. At this point my proposed collaborators elected to pursue other interests, and the project was put on the shelf until it was revived by the present set of authors.

My coauthor is Genesee Optics Software, Inc. Obviously the book is the product of the work of real people, i.e., myself and the staff of Genesee. In alphabetical order, the Genesee personnel who have been involved are Charles Dubois, Henry Gintner, Robert MacIntyre, David Pixley, Lynn VanOrden, and Scott Weller. They have been responsible for the computerized lens data tables, lens drawings, and aberration plots which illustrate each lens design.

Many of the lens designs included in this book are from OPTICS TOOLBOX® (a software product of Genesee Optics Software), which was originally authored by Robert E. Hopkins and Scott W. Weller. OPTICS TOOLBOX is a collection of lens designs and design commentary within an expert-system, artificial-intelligence, relational data base.

This author's optical design experience has spanned almost five decades. In that period lens design has undergone many radical changes. It has progressed from what was a semi-intuitive art practiced by a very small number of extremely patient and dedicated lovers of detail and precision. These designers used a very limited amount of laborious computation, combined with great understanding of lens design principles and dogged perseverance to produce what are now the classic lens design forms. Most of these design forms are still the best, and as such are the basis of many modern optical systems. However, the manner in which lenses are designed today is almost completely different in both technique and philosophy. This change is, of course, the

result of the vastly increased computational speed now available to the lens designer.

In essence, much modern lens design consists of the selection of a starting lens form and its subsequent optimization by an automatic lens design program, which may or may not be guided or adjusted along the way by the lens designer. Since the function of the lens design program is to drive the design form to the nearest local optimum (as defined by a *merit function*), it is obvious that the starting design form and the merit function together uniquely define which local optimum design will be the result of this process.

Thus it is apparent that, in addition to a knowledge of the principles of optical design, a knowledge of appropriate starting-point designs and of techniques for guiding the design program have become essential elements of modern lens design. The lens designs in this book have been chosen to provide a good selection of starting-point designs and to illustrate important design principles. The design techniques described are those which the author has found to be useful in designing with an optimization program. Many of the techniques have been developed or refined in the course of teaching lens design and optical system design; indeed, a few of them were initially suggested or inspired by my students.

In order to maximize their usefulness, the lens designs in this book are presented in three parts: the lens prescription, a drawing of the lens which includes a marginal ray and a full-field principal ray, and a plot of the aberrations. The inclusion of these two rays allows the user to determine the approximate path of any other ray of interest. For easy comparison, all lenses are shown at a focal length of approximately 100, regardless of their application. The performance data is shown as aberration plots; we chose this in preference to MTF plots because the MTF is valid only for the focal length for which it was calculated, and because the MTF cannot be scaled. The aberration plots *can* be scaled, and in addition they indicate what aberrations are present and show which aberrations limit the performance of the lens. We have expanded on the usual longitudinal presentation of spherical aberration and curvature of field by adding ray intercept plots in three colors for the axial, 0.7 zonal, and full-field positions. We feel that this presentation gives a much more complete, informative, and useful picture of the characteristics of a lens design.

This book is intended to build on some knowledge of both geometrical optics and the basic elements of lens design. It is thus, in a sense, a companion volume to the author's *Modern Optical Engineering*, which covers such material at some length. Presumably the user of this text will already have at least a reasonable familiarity with this material.

There are really only a few well-understood and widely utilized principles of optical design. If one can master a thorough understanding of these principles, their effects, and their mechanisms, it is easy to recognize them in existing designs and also easy to apply them to one's own design work. It is our intent to promote such understanding by presenting both expositions and annotated design examples of these principles.

Readers are free to use the designs contained in this book as starting points for their own design efforts, or in any other way they see fit. Most of the designs presented have, as noted, been patented; such designs may or may not be currently subject to legal protection, although there may, of course, be differences of opinion as to the effectiveness of such protection. The reader must accept full responsibility for meeting whatever limitations are imposed on the use of these designs by any patent or copyright coverage (whether indicated herein or not).

Warren J. Smith

Modern Lens Design

Chapter

1

Introduction

Modern Lens Design is intended as an aid to lens designers who work with the many commercially available lens design computer programs. We assume that the reader understands basic optical principles and may, in fact, have a command of the fundamentals of classical optical design methods. For those who want or need information in these areas, the following books should prove helpful. This author's *Modern Optical Engineering: The Design of Optical Systems,* 2d ed., McGraw-Hill, 1990, is a comprehensive coverage of optical system design; it includes two full chapters which deal specifically with lens design in considerable detail. Rudolf Kingslake's *Optical System Design* (1983), *Fundamentals of Lens Design* (1978), and *A History of the Photographic Lens* (1989), all by Academic Press, are complete, authoritative, and very well written.

Authoritative books on lens design are rare, especially in English; there are only a few others available. The Kingslake series *Applied Optics and Optical Engineering,* Academic Press, contains several chapters of special interest to lens designers. Volume 3 (1965) has chapters on lens design, photographic objectives, and eyepieces. Volume 8 (1980) has chapters on camera lenses, aspherics, automatic design, and image quality. Volume 10 (1987) contains an extensive chapter on afocal systems. Milton Laikin's *Lens Design,* Marcel Dekker, 1991, is a volume similar to this one, with prescriptions and lens drawings. Its format differs in that no aberration plots are included; instead, modulation transfer function (MTF) data for a specific focal length and *f* number are given. Now out of print, Arthur Cox's *A System of Optical Design,* Focal Press, 1964, contains a complete, if unique, approach to lens design, plus prescriptions and (longitudinal) aberration plots for many lens design patents.

This book has several primary aims. It is intended as a source book

for a variety of designed lens types which can serve as suitable starting points for a lens designer's efforts. A study of the comparative characteristics of the annotated designs contained herein should also illustrate the application of many of the classic lens design principles. It is also intended as a handy, if abridged, reference to many of the equations and relationships which find frequent use in lens design. Most of these are contained in the Formulary at the end of the book. And last, but not least, the text contains extensive discussions of design techniques which are appropriate to modern optical design with an automatic lens design computer program.

The book begins with a discussion of automatic lens design programs and how to use them. The merit function, optimization, variables, and the various techniques which are useful in connection with a program are covered. Chapter 3 details many specific improvement strategies which may be applied to an existing design to improve its performance. The evaluation of a design is discussed from the standpoint of ray and wave aberrations, and integrated with such standard measures as MTF and Strehl ratio. The sample lens designs follow. Each presents the prescription data, a drawing of the lens with marginal and chief rays, and an aberration analysis consisting of ray intercept plots for three field angles, longitudinal plots of spherical aberration and field curvature, and a plot of distortion. A discussion of the salient features of each design accompanies the sample designs, and comments (in some cases quite extensive) regarding the design approach are given for each class of lens. The Formulary, intended as a convenient reference, concludes the book.

The design of the telescope objective is covered in Chap. 6, beginning with the classic forms and continuing with several possible modifications which can be used to improve the aberration correction. These are treated in considerable detail because they represent techniques which are generally applicable to all types of designs. For similar reasons, Chap. 8 deals with the basic principles of airspaced anastigmats in a rather extended treatment. The complexities of the interrelationships involved in the Cooke triplet anastigmat are important to understand, as are the (almost universal) relationships between the vertex length of an ordinary anastigmat lens and its capabilities as regards speed and angular coverage.

Chapter

2

Automatic Lens Design: Managing the Lens Design Program

2.1 The Merit Function

What is usually referred to as *automatic lens design* is, of course, nothing of the sort. The computer programs which are so described are actually optimization programs which drive an optical design to a local optimum, as defined by a *merit function* (which is not a true merit function, but actually a defect function). In spite of the preceding disclaimers, we will use these commonly accepted terms in the discussions which follow.

Broadly speaking, the merit function can be described as a combination or function of calculated characteristics, which is intended to completely describe, with a single number, the value or quality of a given lens design. This is obviously an exceedingly difficult thing to do. The typical merit function is the sum of the squares of many image defects; usually these image defects are evaluated for three locations in the field of view (unless the system covers a very large or a very small angular field). The squares of the defects are used so that a negative value of one defect does not offset a positive value of some other defect.

The defects may be of many different kinds; usually most are related to the quality of the image. However, any characteristic which can be calculated may be assigned a target value and its departure from that target regarded as a defect. Some less elaborate programs utilize the third-order (Seidel) aberrations; these provide a rapid and efficient way of adjusting a design. These cannot be regarded as optimizing the image quality, but they do work well in correcting ordinary lenses. Another type of merit function traces a large number of

rays from an object point. The radial distance of the image plane intersection of the ray from the centroid of all the ray intersections is then the image defect. Thus the merit function is effectively the sum of the root-mean-square (rms) spot sizes for several field angles. This type of merit function, while inefficient in that it requires many rays to be traced, has the advantage that it is both versatile and in some ways relatively foolproof. Some merit functions calculate the values of the classical aberrations, and convert (or weight) them into their equivalent wavefront deformations. (See Formulary Sec. F-12 for the conversion factors for several common aberrations.) This approach is very efficient as regards computing time, but requires careful design of the merit function. Still another type of merit function uses the variance of the wavefront to define the defect items. The merit function used in the various David Grey programs is of this type, and is certainly one of the best of the commercially available merit functions in producing a good balance of the aberrations.

Characteristics which do not relate to image quality can also be controlled by the lens design program. Specific construction parameters, such as radii, thicknesses, spaces, and the like, as well as focal length, working distance, magnification, numerical aperture, required clear apertures, etc., can be controlled. Some programs include such items in the merit function along with the image defects. There are two drawbacks which somewhat offset the neat simplicity of this approach. One is that if the first-order characteristics which are targeted are not initially close to the target values, the program may correct the image aberrations without controlling these first-order characteristics; the result may be, for example, a well-corrected lens with the wrong focal length or numerical aperture. The program often finds this to be a local optimum and is unable to move away from it. The other drawback is that the inclusion of these items in the merit function has the effect of slowing the process of improving the image quality. An alternative approach is to use a system of constraints outside the merit function. Note also that many of these items can be controlled by features which are included in almost all programs, namely *angle-solves* and *height-solves*. These algebraically solve for a radius or space to produce a desired ray slope or height.

In any case, the merit function is a summation of suitably weighted defect items which, it is hoped, describes in a single number the worth of the system. The smaller the value of the merit function, the better the lens. The numerical value of the merit function depends on the construction of the optical system; it is a function of the construction parameters which are designated as variables. Without getting into the details of the mathematics involved, we can realize that the merit function is an n-dimensional space, where n is the number of the vari-

able constructional parameters in the optical system. The task of the design program is to find a location in this space (i.e., a lens prescription or a solution vector) which minimizes the size of the merit function. In general, for a lens of reasonable complexity there will be many such locations in a typical merit function space. The automatic design program will simply drive the lens design to the nearest minimum in the merit function.

2.2 Optimization

The lens design program typically operates this way: Each variable parameter is changed (one at a time) by a small increment whose size is chosen as a compromise between a large value (to get good numerical accuracy) and a small value (to get the *local* differential). The change produced in every item in the merit function is calculated. The result is a matrix of the partial derivatives of the defect items with respect to the parameters. Since there are usually many more defect items than variable parameters, the solution is a classical least-squares solution. It is based on the assumption that the relationships between the defect items and the variable parameters are linear. Since this is usually a false assumption, an ordinary least-squares solution will often produce an unrealizable lens or one which may in fact be worse than the starting design. The *damped least-squares* solution, in effect, adds the weighted squares of the parameter changes to the merit function, heavily penalizing any large changes and thus limiting the size of the changes in the solution. The mathematics of this process are described in Spencer, "A Flexible Automatic Lens Correction Program," *Applied Optics,* vol. 2, 1963, pp. 1257–1264, and by Smith in W. Driscoll (ed.), *Handbook of Optics,* McGraw-Hill, New York, 1978.

If the changes are small, the nonlinearity will not ruin the process, and the solution, although an approximate one, will be an improvement on the starting design. Continued repetition of the process will eventually drive the design to the nearest local optimum.

One can visualize the situation by assuming that there are only two variable parameters. Then the merit function space can be compared to a landscape where latitude and longitude correspond to the variables and the elevation represents the value of the merit function. Thus the starting lens design is represented by a particular location in the landscape and the optimization routine will move the lens design downhill until a minimum elevation is found. Since there may be many depressions in the terrain of the landscape, this optimum may not be the best there is; it is a local optimum and there can be no assurance (except in very simple systems) that we have found a global

optimum in the merit function. This simple topological analogy helps to understand the dominant limitations of the optimization process: the program finds the nearest minimum in the merit function, and that minimum is uniquely determined by the design coordinates at which the process is begun. The landscape analogy is easy for the human mind to comprehend; when it is extended to a 10- or 20-dimension space, one can realize only that it is apt to be an extremely complex neighborhood.

2.3 Local Minima

Figure 2.1 shows a contour map of a hypothetical two-variable merit function, with three significant local minima at points *A, B,* and *C*; there are also three other minima at *D, E,* and *F*. It is immediately apparent that if we begin an optimization at point *Z*, the minimum at point *B* is the only one which the routine can find. A start at *Y* on the ridge at the lower left will go to the minimum at *C*. However, a start

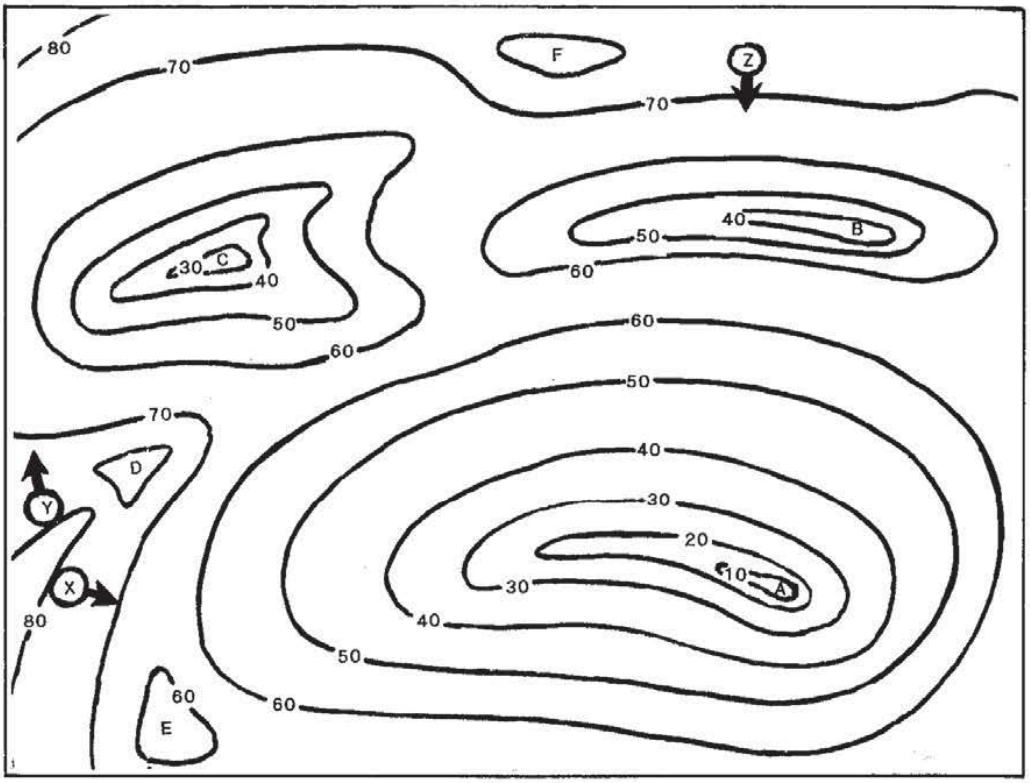


Figure 2.1 Topography of a hypothetical two-variable merit function, with three significant minima (*A, B, C*) and three trivial minima (*D, E, F*). The minimum to which a design program will go depends on the point at which the optimization process is started. Starting points *X, Y,* and *Z* each lead to a different design minimum; other starting points can lead to one of the trivial minima.

at *X*, which is only a short distance away from *Y*, will find the best minimum of the three, at point *A*. If we had even a vague knowledge of the topography of the merit function, we could easily choose a starting point in the lower right quadrant of the map which would guarantee finding point *A*. Note also that a modest change in any of the three starting points could cause the program to stagnate in one of the trivial minima at *D*, *E*, or *F*. It is this sort of minimum from which one can escape by "jolting" the design, as described below.

The fact that the automatic design program is severely limited and can find only the nearest optimum emphasizes the need for a knowledge of lens design, in order that one can select a starting design form which is close to a good optimum. This is the only way that an automatic program can systematically find a good design. If the program is started out near a poor local optimum, the result is a poor design.

The mathematics of the damped least-squares solution involves the inversion of a matrix. In spite of the damping action, the process can be slowed or aborted by either of the following conditions: (1) A variable which does not change (or which produces only a very small change in) the merit function items. (2) Two variables which have the same, nearly the same, or scaled effects on the items of the merit function. Fortunately, these conditions are rarely met exactly, and they can be easily avoided.

If the program settles into an unsatisfactory optimum (such as those at *D*, *E*, and *F* in Fig. 2.1) it can often be jolted out of it by manually introducing a significant change in one or more parameters. The trick is to make a change which is in the direction of a better design form. (Again, a knowledge of lens designs is virtually a necessity.) Sometimes simply freezing a variable to a desirable form can be sufficient to force a move into a better neighborhood. The difficulty is that too big a change may cause rays to miss surfaces or to encounter total internal reflection, and the optimization process may break down. Conversely, too small a change may not be sufficient to allow the design to escape from a poor local optimum. Also, one should remember that if the program is one which adjusts (optimizes) the damping factor, the factor is usually made quite small near an optimum, because the program is taking small steps and the situation looks quite linear; after the system is jolted, it is probably in a highly nonlinear region and a big damping factor may be needed to prevent a breakdown. A manual increase of the damping factor can often avoid this problem.

Another often-encountered problem is a design which persists in moving to an obviously undesirable form (when you *know* that there is a much better, very different one—the one that you want). Freezing the form of one part of the lens for a few cycles of optimization will often allow the rest of the lens to settle into the neighborhood of the

desired optimum. For example, if one were to try to convert a Cooke triplet into a split front crown form, the process might produce either a form which is like the original triplet with a narrow airspaced crack in the front crown, or a form with rather wild meniscus elements. A technique which will usually avoid these unfortunate local optima in this case is to freeze the front element to a plano-convex form by fixing the second surface to a plane for a few cycles of optimization. Again, one must know which lens forms are the good ones.

2.4 Types of Merit Functions

Many programs allow the user to define the merit function. This can be a valuable feature because it is almost impossible to design a truly *universal* merit function. As an example, consider the design of a simple Fraunhofer telescope objective: a merit function which controls the spherical and chromatic aberrations of the axial marginal ray and the coma of the oblique ray bundle (plus the focal length) is all that is necessary. If the design complexity is increased by allowing the airspace to vary and/or adding another element, the merit function may then profitably include entries which will control zonal spherical, spherochromatism, and/or fifth-order coma. But as long as the lens is thin and in contact with the aperture stop, it would be foolish to include in the merit function entries to control field curvature and astigmatism. There is simply no way that a thin stop-in-contact lens can have any control over the inherent large negative astigmatism; the presence of a target for this aberration in the merit function will simply slow down the solution process. It would be ridiculous to use a merit function of the type required for a photographic objective to design an ordinary telescope objective. (Indeed, an attempt to correct the field curvature may lead to a compromise design with a severely undercorrected axial spherical aberration which, in combination with coma, may fool the computer program into thinking that it has found a useful optimum.)

There are many design tasks in this category, where the requirements are effectively limited in number and a simple, equally limited merit function is clearly the best choice. In such cases, it is usually obvious that some specific state of correction will yield the best results; there is no need to *balance* the correction of one aberration against another.

More often, however, the situation is not so simple; compromises and balances are required and a more complex, suitably weighted merit function is necessary. This can be a delicate and somewhat tricky matter. For example, in the design of a lens with a significant aperture and field, there is almost always a (poor) local optimum in

which (1) the spherical aberration is left quite undercorrected, (2) a compromise focus is chosen well inside the paraxial focus, (3) the Petzval field is made inward-curving, and (4) overcorrected oblique spherical aberration is introduced to "balance" the design. A program which relies on the rms spot radius for its merit function is very likely to fall into this trap. A better design usually results if the spherical (both axial and oblique) aberrations are corrected, the Petzval curvature is reduced, and a small amount of overcorrected astigmatism is introduced. When one recognizes this sort of situation, it is a simple matter to adjust the weighting of the appropriate targets in the merit function to force the design into a form with the type of aberration balance which is desired. Another way to avoid this problem is to force the system to be evaluated/designed at the paraxial focus rather than at a compromise focus, i.e., to not allow defocusing. As can be seen, the design of a general-purpose merit function which will optimally balance a wide variety of applications is not a simple matter.

Although it is not always necessary, there are occasions when it is helpful to begin the design process by controlling only the first-order properties (image size, image location, spatial limitations, etc.). Then one proceeds to control the chromatic and perhaps the Petzval aberrations. (Things may even go better if the first-order and the chromatic are fairly completely worked out by hand before submitting the system to an automatic design process.) The next step in the sequence is to correct the primary aberrations (spherical, coma, astigmatism, and distortion), either directly or by using the Seidel coefficients, and finally proceed to balancing and correcting the higher-order residuals. This sort of ordered approach is sometimes useful (or even necessary) when one is exploring terra incognita, and, of course, it requires a user-defined merit function if it is to be implemented.

2.5 Stagnation

Sometimes the automatic design process will stagnate and the convergence toward a solution will become so slow as to be imperceptible. This can result from being in a very flat and broad optimum in the merit function. It can also result from an ill-designed merit function. Often first-order properties which are specified in the merit function are the cause of the problem. It is only too easy to require contradictory or redundant characteristics. This is especially true for zoom lenses or multiconfiguration systems, which can be confusingly complex. When stagnation occurs, or convergence is slower than you know it should be, it is wise to stop and examine the merit function for problems. Look critically at every item in the merit function and consider what it is intended to be doing and what it *actually* does. Eliminate

redundancies and try to make each entry in the merit function explicitly control its intended characteristic. Stagnation may also result from a starting design which is so far from a solution that differential changes to the variables have a negligible effect.

2.6 Generalized Simulated Annealing

The discussions above have centered on the standard damped least-squares program, or its equivalent. There have been several versions of *random search* programs proposed in the past. The most recent of these is quite sophisticated and is called *generalized simulated annealing*. In this, the computer randomly selects the lens dimensions (within a limited range and according to some probability distribution) and evaluates the resulting lens prescription. If the new version is better than the old, it is unconditionally accepted. If it is worse, it *may* be accepted, on the basis of random chance, weighted by a probability function which reduces the chance of acceptance in proportion to the amount that the lens is worse than the original form. This sort of approach obviously allows the program an easy escape from the local minima described with Fig. 2.1, but it equally obviously requires a very large number of trials before a random chance can find a good combination of dimensions for the lens. Nonetheless, it does work, but not rapidly. Perhaps as computers increase in speed, a program of this sort will displace or supplement the damped least-squares as the routine of choice for automatic lens design.

2.7 Considerations about Variables for Optimization

The potential variables for use in optimization include: the surface curvatures, conic constants, and asphericities; the surface spacings; and the refractive characteristics of the materials involved. Occasionally tilts and decentrations are also included as variables.

Materials

Although the material characteristics are not continuous variables, for optical glasses at least, the index and dispersion (or V value) can be varied within the boundaries of the glass map (Fig. 2.2) as if they were. The real glass nearest the optimized values can be substituted for the optimized glass to achieve nearly the same resultant design after another cycle or two of optimization with the real glass. Note that this is *not* true for partial dispersions, since there are relatively few glasses with partial dispersions unusual enough to be useful in the

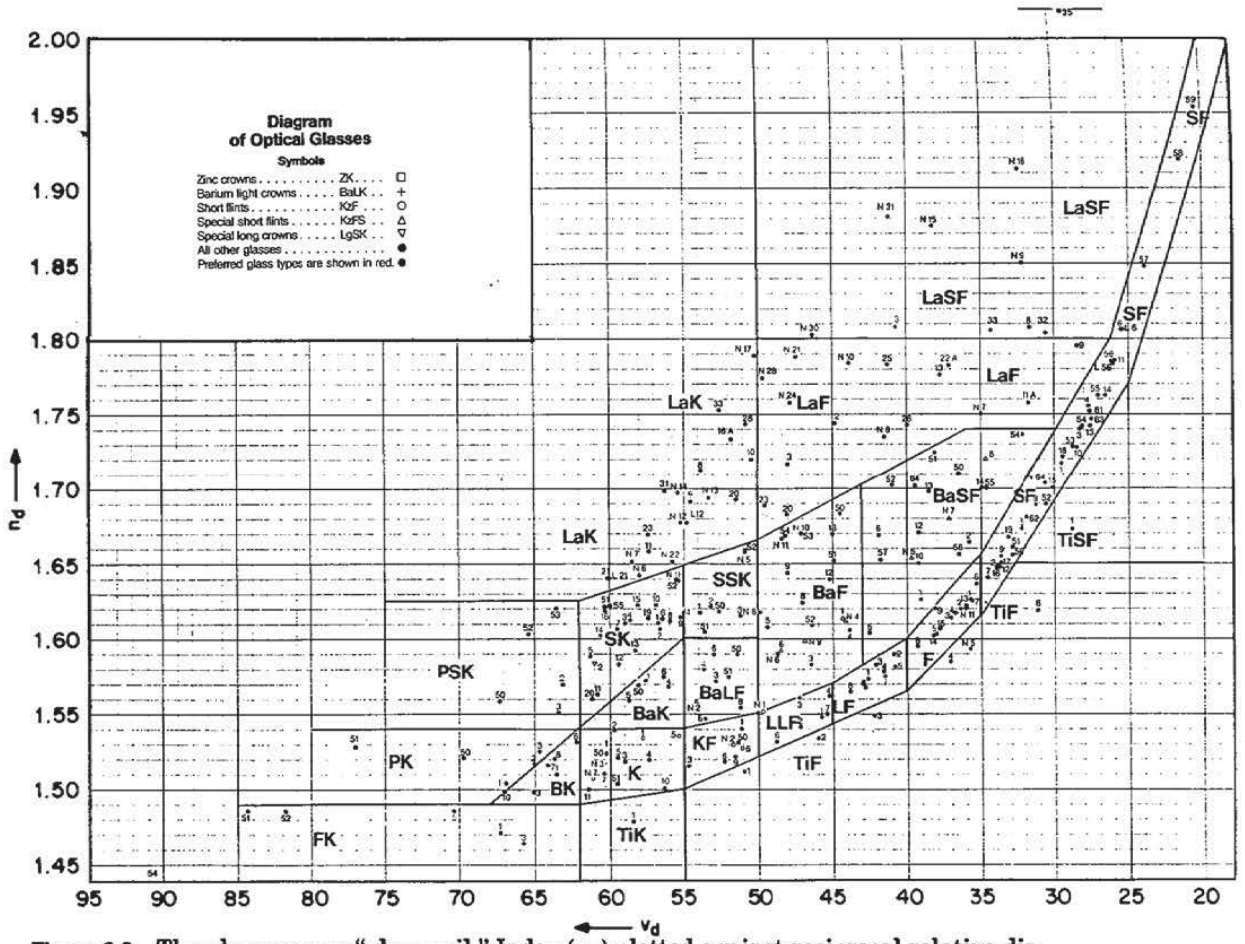


Figure 2.2 The glass map or "glass veil." Index (n_d) plotted against reciprocal relative dispersion (Abbe V value). The glass types are indicated by the letters in each area. The glass line is made up of the glasses of types K, KF, LLF, F, and SF, which are strung along the bottom of the veil. (Note that K stands for *kron*, German for crown, and S stands for *schwer*, or heavy or dense.) (*Courtesy of Schott Glass Technologies, Inc., Duryea, Pa.*)

correction of secondary spectrum. Obviously, for applications outside the spectral regions where optical glass is usable, one cannot treat the refractive characteristics as variables, since the available materials tend to be few and far between.

In many of the simpler types of designs it is essential to allow the glass characteristics to vary. In the Cooke triplet for example, the relationship between the V values of the crown and flint elements determines the overall length of the lens. As described in Sec. 8.3, the length of a triplet (and that of most anastigmats) determines the amount of higher-order spherical aberration and astigmatism; these in turn determine the aperture and field coverage capabilities of the lens. If these types of lenses are to be optimized to suit the application at hand, the glass characteristics *must* be allowed to vary.

Some optimization programs have difficulty with the bounds of the glass map; if this is a problem, the optimization process is often facilitated by starting the variable glass well away from the boundary, so that it can find its best value before encountering the boundary problem.

It is often better to vary the flint glasses than to vary the crowns. This is because the crowns usually tend to go to the upper left corner (high index, high V value) of the glass map. Flints head for the lower right corner, and are then, of course, constrained to lie on the glass line. The glasses along the glass line are numerous, inexpensive, and almost universally well behaved. On the other hand, the crown glasses in the upper left corner include in their number many which are expensive and/or easily attacked by the environment. Thus one might be willing to accept the computer's choice of a glass along the glass line, but would prefer to make a more discriminating selection from among the others.

Curvatures

In general, one would expect to want to make use of every available variable. This is almost always true regarding the curvatures, all of which, unless there is a reason to constrain the shape of an element, are usually allowed to vary.

Airspaces

Ordinarily, airspaces may be regarded in the same light as curvatures, since they are continuously variable and are very effective variables.

Defocusing

Although the distance by which the design image plane departs from the paraxial focus is usually an airspace, and can be regarded as a variable,

its effects can be insidious. If the image surface is allowed to depart from the paraxial focus from the beginning of the optimization process, an unfortunate lens may result. In some lenses, and with some optimization merit functions, the tendency is to produce a lens with:

1. The image plane well inside the paraxial focus
2. A large undercorrected spherical aberration
3. A strongly inward-curving field
4. A heavy overcorrecting oblique spherical

Although this combination occupies a local optimum in the merit function, this is usually not the best state of correction. It fools the optimization program because the undercorrected spherical causes the best axial focus to lie to the left of the paraxial focus and the overcorrected oblique spherical causes the best off-axis focus to lie to the right of the inward-curving field; the net result is that, to the program, the field seems flat. One can usually avoid this pitfall by not allowing any defocus in the early stages of the optimization and/or putting a heavy penalty on the defocusing. Note that for *some* lenses, such as non-diffraction-limited systems used with detectors, the correction described above may in fact be a good one. See the comments on aberration balance in Sec. 3.8.

Thickness

Element thicknesses must be regarded quite differently than airspaces. They must of course be bounded by the necessity for a practical edge thickness for the positive elements and a reasonable center thickness for the negative elements. In many designs, element thickness is an insignificant and ineffective variable (and one whose effects are easily duplicated by an adjacent airspace). In this circumstance one can arbitrarily select a thickness on the basis of economy or ease of fabrication. The elements in such designs are typically quite thin; see, for example, a telescope objective or an ordinary Cooke triplet.

There are, however, many systems in which the element thickness is not only an effective variable, but one which is essential to the success of the design type. The older meniscus lenses (Protar, Dagor, etc.) and the double-Gauss forms depend on the separation of the concave and convex surfaces of their thick meniscus components to control the Petzval curvature and, in many instances, the higher-order aberrations. In lenses of this type it is absolutely essential that these glass thicknesses be allowed to vary.

One must be wary of and skeptical toward a thickness variable which is very weakly effective. Occasionally an optimization program will produce a design with an overly thick element, where the large

thickness produces only a very small improvement (which is not worth the added cost of producing the thick element). This occurs because the optimization routine will seek out any improvement that it can get, no matter how small, and without concern as to the cost. It is wise to test the value of a thick element if there is any doubt about its utility. This is readily accomplished with another optimization run which fixes the thickness in question to a smaller value. Very often the performance of the thin version will not be noticeably different from that of the "optimum" thicker version. Although most significant with respect to lens thickness, this same rationale obviously applies to airspaces as well.

Aspheric surfaces

Surface asphericity can be an extremely effective (if sometimes expensive) variable, but it is one that often requires a bit of finesse. On occasion, one may be ill-advised to begin an optimization with the conic constant and all the aspheric deformation coefficients used simultaneously as variables. The conic constant and the fourth-order deformation coefficient both affect the third-order aberrations in exactly the same way. Thus they are at least partially redundant, but more significantly, identical variables have an undesirable effect on the mathematics of the optimization process. It is often advisable to vary one or the other, but not both. A safe practice is to vary only the conic constant (or the fourth-order term) at first, and then add the higher-order terms (sixth, eighth, tenth) one at a time, as necessary. The tenth-order term is, in many systems, totally unnecessary, adding little or nothing to the quality of the system; in fact, the eighth-order term is often something that can be done without.

A surface defined by a tenth-order polynomial can cause the spherical aberration to be corrected exactly to zero at four ray heights. If there are only four axial rays in the merit function, their ray intercept errors may all be brought to zero; the danger is that, between these rays, the residual aberration may be unacceptably large. A tenth-order surface can be a rather extreme shape. Thus the use of an aspheric surface sometimes calls for more rays in the merit function than one might otherwise expect to need. With a program which allows wavefront deformation or optical path difference (OPD) targets in the merit function, the severity of this problem can be lessened.

2.8 How to Increase the Speed or Field of a System and Avoid Ray Failure Problems

Very often, the lens designer is faced with the necessity of increasing the speed (i.e., relative aperture, numerical aperture, etc.) and/or the

field of view of an existing optical system. There are two common reasons why this may be desirable. One may want to adapt an existing design (such as those in this book) to an application which requires a larger aperture or wider field than that for which the original lens has been configured. The other common situation is simply the creation of an entirely new system with a relatively large aperture and/or field. In either case, the difficulty which can arise is that the rays which are needed to design the system may not be able to get through the initial lens prescription.

There are two reasons that a ray may not be able to get through. One reason is that the height of the ray at a surface may be greater than the radius of the surface; the ray simply misses the surface entirely and its path obviously cannot be calculated any further. The second reason is that the ray may encounter total internal reflection (TIR) in passing from a higher index to a lower; again, the ray path cannot be calculated. Each of these conditions represents a boundary which, if crossed, causes failure of the ray trace. Note also that, as these boundaries are approached, the situation rapidly becomes very unstable. This is because, close to the boundary, the angle of incidence (for the case of the ray approaching the value of the radius) or the angle of refraction (for the case of the ray approaching TIR) is very near to 90°. Near this angle, Snell's law of refraction ($n \sin I = n' \sin I'$) becomes *very* nonlinear, producing a highly unstable situation which often explodes as the lens construction parameters are incremented in the course of the optimization.

A good way of dealing with this situation is simply to back off from the aperture and/or field requirement that is causing the problem. Many design programs have the capability to easily adjust or scale the aperture and field angle. If a change is made to smaller values of aperture or field, the rays will no longer be so near to the failure boundary. If the lens is now optimized, the program is very likely to adjust the lens parameters so as to reduce the angles of incidence, because this is usually a factor which causes the aberrations to be reduced. The optimizing changes can thus be expected to pull the problem situation in the system further away from the ray failure boundary, *if this is possible.*

After the optimization has relaxed the problem, the field and/or the aperture can usually be adjusted (scaled) to a moderately larger value without again encountering the failure boundary. Depending on just how sensitive the system is to the problem, an increase of about 10 to 50 percent may be appropriate. Now another cycle of optimization will strongly tend to again reduce the troublesome angles of incidence.

This process of adjusting (scaling) the field or the aperture to larger values and then optimizing is continued until the desired aperture or

field is attained without ray failures. This works well, *provided* that this desired result is possible for the lens configuration which is under study. It may be necessary to choose another configuration, usually one with more elements. When this is necessary, a drawing of the lens and rays (of the last design form which has successfully passed all the rays) will usually indicate which rays and which surfaces are causing the problem. One simply looks for angles of incidence or refraction which are large (and often near 90°). Then the offending element can be split into two (or more) elements which are shaped to reduce these angles. The scale-and-optimize process can now be repeated with a much improved chance of success. Note that, in general, the examination of the critical ray paths (typically those of the marginal rays) for large angles of incidence or refraction is a technique which will often indicate the source of a design problem.

2.9 Test Plate Fits, Melt Fits, and Thickness Fits

When the deleterious effects of fabrication tolerances become too large to bear, a technique commonly used to reduce these effects is to fit the lens design to the known values for the radius tooling and/or to the measured glass indices. The former is called a *test plate* (or *test glass* or *tooling*) fit; the latter is called a *melt fit.*

The *test plate fit* is begun by first obtaining a list of the available test plate radii from the shop which is scheduled to fabricate the lens. It is wise to ascertain that the radius values of the list are not just nominal values, but are based on accurate and recent measurements of the test plates, since there is often a significant difference between the two.

The fit is carried out as follows: A surface is selected at which to begin the process. This selection is based on one of the following criteria: (1) the surface most sensitive to change,* (2) the surface with the shortest radius, (3) the strongest surface [i.e., with the largest surface power $(n' - n)/r$], or (4) the surface which shows the largest curvature difference from the nearest available test plate radius. Very often all four of these criteria will indicate the same surface; if not, the choice of which one to use is almost a matter of taste. The nearest radius on the test plate list is substituted for the selected surface and the lens is reoptimized, allowing all the variable parameters to change

*Note that the relative sensitivity of any dimension of the system can be determined quite easily by making an incremental change in the dimension and noting the effect it produces on the merit function. This, of course, assumes a merit function which accurately represents the quality of the image.

(except of course, the radius which has been set to the test plate value).

Another surface is then chosen and fixed to the nearest test plate radius; the lens is reoptimized again. This process is repeated until all the surfaces have been fitted to test plates. It is usually wise to avoid fitting both surfaces of a singlet (or all surfaces of a component) until all components have at least one fitted surface each. This allows the unfitted surface to vary and adjust for power, chromatic, and Petzval aberrations and the like for as long as possible.

With a reasonably complete test plate list, all radii can usually be fitted without significantly degrading the image quality (i.e., the merit function). In fact, the merit function is often slightly improved by this process, because of the additional cycles of optimization which have been performed. If the test plate list is limited or has a gap in it, it may not be possible to fit all the design radii to test plates without degrading the performance. In such a case, one must either fabricate new tooling and test plates, or seek a new vendor for the parts.

A *melt fit* reoptimizes the lens design by using measured data for the material indices instead of the nominal values from the glass catalog. This measured data comes in the form of what is called a *melt sheet* provided by the glass manufacturer or supplier. For noncritical applications, this data is usually sufficient. A worthwhile elaboration of the process is to determine the difference between the measured index values and the catalog values. These differences are then plotted against wavelength. This plot should be a smooth, relatively level line. A data point which does not plot smoothly is suspect; the measured data may well be in error. Next, a smoothly drawn curve through the points is used to determine improved values for the index differences. These differences are then applied to the catalog values to arrive at better values for the measured melt data, and the improved values are used in the melt fit reoptimization. The smoothed curve of differences can also be used to determine the index for wavelengths which are not included in the melt sheet.

An ordinary melt sheet will list the indices for the wavelengths of d, e, C, F, and g light. These are usually not individually measured. Instead, the index is measured for d light and the index difference between C and F light is measured. These two measurements are fed into a computer program which uses the known characteristics of the glass type to calculate the indices for e, C, F, and g light. This, while not ideal, is adequate for many, even most, applications. However, for some critical applications, and for all applications in which an attempt has been made to reduce or correct the secondary spectrum, it is usually quite unsatisfactory. For an additional charge, the glass manufacturer can provide what is usually called a *precision* melt sheet, for

which the indices have been measured individually, and to a greater precision. Index values for wavelengths specific to the application can also be measured. It is, of course, wise to subject even this data to the difference test and smoothing process described in the preceding paragraph.

We called the last of these fits a *thickness fit* in the heading for this section. This is a process which is carried out during the final assembly of the lens. In sum, one reoptimizes the lens by varying the airspaces of the system, using accurately measured data for the radii, thicknesses, and indices of the fabricated elements. If the melt data and the test plate data are available, this just amounts to adding the measured element thicknesses and reoptimizing.

As can be seen, all of the above procedures are designed to almost completely eliminate the effects of the fabrication tolerances on the performance of the system. What is left, instead of the tolerances, is the uncertainty or inaccuracy of the various measurements on which the fits are based. The effect is usually quite modest and, therefore, acceptable. However, these uncertainties are, in fact, the exact equivalent of tolerances in determining the performance of the fabricated system.

2.10 Spectral Weighting

In any ray-tracing process, the index of refraction of the material is, of necessity, that corresponding to a single specific wavelength. Most lens *design* programs allow the use of only three wavelengths to represent the spectral bandpass for which the system is to be designed. Several *analysis* programs allow the use of five or ten suitably weighted wavelengths in calculating such things as MTF, point spread functions, radial energy distributions, and the like.

In either case, the question becomes, "What wavelengths should I use, and how should they be weighted?" For visual systems and just three wavelengths, the classical answer is to use d (or e), C, and F light, with weightings typically set at 1.0, 0.5, and 0.5 respectively. For other applications, this approximately corresponds to using the central wavelength and the wavelengths 25 percent from the extreme edges of the passband. If the results of an image analysis need not be especially precise, three wavelengths may be sufficient. For calculations done in the midst of the design process, this is often good enough to enable a judgment as to the relative merit of, or the rate of improvement between, two stages in the design process.

However, what should one do *in general?* To immediately dispose of the obvious, it is apparent that the more wavelengths that are used, the more accurate the results can be. That said, how do we select the

wavelengths to be used? There are three obvious choices. In order to assess the full effects of the chromatic aberrations, one would like to include the extreme long and short wavelengths. One would also probably like to consider dividing the spectral weighting function into increments of equal power or response, so that each wavelength would represent an equally weighted sector. But if this is done, the extremes of the spectral passband are not included. Another possible option is to choose wavelengths which are evenly distributed across the spectral band, and to weight them according to the spectral response function. One might choose an even distribution on a wavelength scale, or what might be a bit better, an even distribution on a wave number (reciprocal wavelength) scale.

From this it should be apparent that most wavelength and weighting choices are based on some sort of compromise. Often the outer two wavelengths are chosen to be fairly close to the ends of the passband and the intermediate wavelengths and weights are a compromise between an even power distribution and an even wavelength spacing. If there are peaks or bumps in the spectral response function, wavelengths are often selected to be at or near the peaks. Note that, to a limited extent, one's choices will partially control the design process: heavier weighting at the ends of the spectrum will obviously emphasize both primary and secondary chromatic aberration, spherochromatism, and the like; heavy weights on the central wavelengths will emphasize the monochromatic aberrations at the expense of the chromatic.

2.11 How to Get Started

Experienced designers are often asked questions such as "How do you know where to start?" or "How did you decide that a (name of a design type) could meet the specs?" or "Why did you shift the components around?"

The answer to these questions is almost always "Experience," which probably means that the full answer is different for every problem. It would be foolhardy to pretend to be able to give a complete, definitive answer or set of answers to such questions, but there are a few guidelines which should be reasonably dependable.

Figure 2.3 is, in effect, a compilation of some of that experience. If the system to be designed roughly corresponds to a photographic objective or to one of the other types indicated, the figure can be used as an easy guide to the selection of an appropriate form. In this plot the areas corresponding to various combinations of field and aperture are labeled to indicate the type of system which is commonly used there. Obviously the boundary of each area is fuzzy and ill-defined. In a pre-

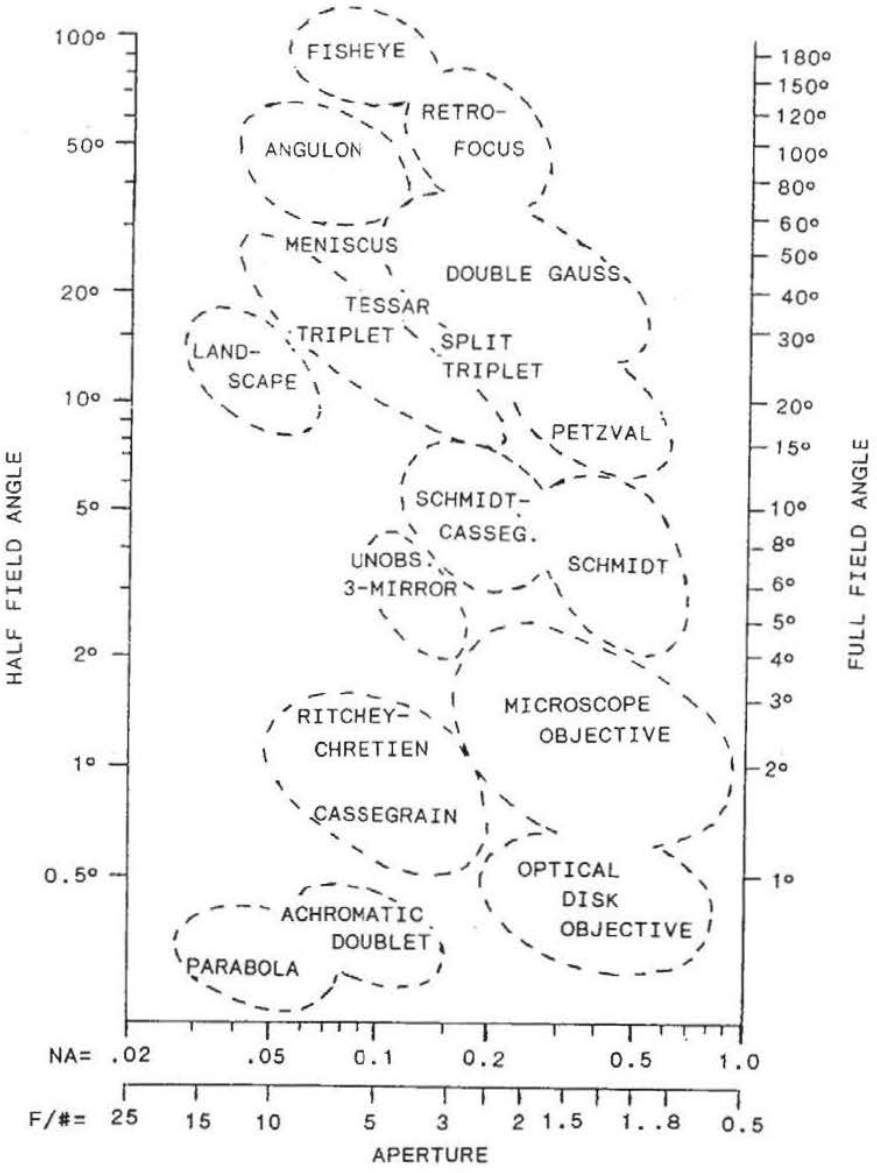


Figure 2.3 Map showing the design types which are commonly used for various combinations of aperture and field of view.

sentation of this type there is also an implied level of performance within each area, which is presumably typical of each particular lens type. In general, the performance (image quality, resolution, or whatever) is better when a given lens type is optimized for a smaller field and/or smaller aperture, that is, for a combination closer to the origin of the plot. Thus one can select the design type from Fig 2.3 corresponding to the field-aperture combination required with reasonable assurance that it is an appropriate choice. But should the performance

prove inadequate for the application at hand, one can move up to the form which is above and to the right on the plot, which should have a greater performance capability. Of course, if the required level of performance is known in advance to be relatively high, one would select the more capable type to begin with. The same selection approach can be applied to the components of a more complex system. There are many design types and applications which are not represented on this necessarily abbreviated chart. However, many of these are presented in the lens designs included in this book; the reader may wish to mark up Fig. 2.3 to add the types which are of particular interest.

For a complex system, the approach must start on a more fundamental level. The first step is to collect and tabulate the requirements to be met; these may include such things as:

Resolution or performance (versus diffraction limit)

Wavelength

Fields of view

Image size

Aperture, numerical aperture (NA), and *f* number

Vignetting or illumination uniformity

Focal lengths or magnifications

Space limitations

The next step is to make a first-order layout of the system which will satisfy the requirements. The first-order layout is simply an arrangement of component powers (or focal lengths) and spacings which will produce an image in the required location, in the required orientation, and of the required size. At this stage, no consideration is given to the design type of each component; one is concerned only with its power, aperture, and field as first-order, i.e., paraxial, characteristics. For systems (or portions of systems) which consist of two components, the equations of Sec. F.7 can be extremely useful. For more complex systems, the component by component ray-tracing equations of Sec. F.6 may be used. A general approach is to trace rays which will define the required characteristics. The ray-trace results (ray heights, ray slopes, intersection lengths, etc.) can be expressed as equations with the component powers and spacings as unknowns to be solved for. An important facet of this stage is that one should try to find a layout which minimizes the component powers, or minimizes (or equalizes) the "work" (ray height times component power). Doing this will almost always produce a system with less aberration residuals, one

which is less expensive to fabricate and less sensitive to fabrication and alignment errors.

Sometimes one can leap directly from the first-order layout to choosing the component design types, and thence to the optimization stage. However, if this is not the case, the next step is usually to analyze and/or correct the chromatic aberrations. Equations F.9.6 and F.10.7 can be used for the whole system, or the components can be individually achromatized. To this end, the element powers for a thin achromatic doublet are given by

$$\phi_A = \frac{V_A}{(V_A - V_B)F} \tag{2.1}$$

$$\phi_B = \frac{V_B}{(V_B - V_A)F} = \frac{1}{F} - \phi_A \tag{2.2}$$

and the element surface curvatures are determined from

$$C_1 - C_2 = \frac{\phi}{n - 1} = \frac{1}{r_1} - \frac{1}{r_2} \tag{2.3}$$

At this point a sketch of the system is often helpful. Simply make a scale drawing, showing each element as either a plano-convex or an equi-convex form (or plano-concave or equi-concave for negative elements). If the elements look too fat, they should be split into two or more elements. Be sure that the element diameters are sized properly for the rays that they must pass. At this stage the system should begin to look like a good lens.

The next step is to give some consideration to the Petzval curvature. Choosing suitable anastigmat types for the components is one way to handle this. Another is to use a field flattener in an appropriate location (i.e., near an image) in the system. And, of course, the usual device of configuring the system or component with separated positive and negative elements or surfaces to reduce the Petzval sum can always be utilized if a new design must be created from scratch.

Often many of these steps can be handled conveniently and expeditiously by the automatic design program. The first-order layout can be done with zero-thickness plano-convex or plano-concave elements, allowing the spaces and the curvatures of the curved surfaces (but not the plano surfaces) to vary. The merit function is a simple one, configured to define the required first-order characteristics. Note well the comments in Sec. 2.5 regarding stagnation and contradictory or redundant first-order entries in the merit function.

Since the chromatic aberrations and the Petzval curvature depend on element power and not on element shape, the lens design program

can also be used to find a layout which is a preliminary solution with the chromatic and Petzval adjusted to desired, reasonable values, which typically should be small and negative.

Sometimes it is useful as a next step to allow the elements to bend, and to correct the third-order aberrations. This can be done by putting an angle-solve on the second surface of each element so that the axial ray slope is maintained. More often than not, however, the next step will skip over the third-order and go directly to a full-dress thick lens optimization run.

And, of course, in the best of all worlds, one simply sets up what seems to be a likely layout and proceeds directly to the automatic lens design program, which promptly turns out an excellent design. "Experience!" Lots of luck!

Chapter

3

Improving a Design

3.1 Standard Improvement Techniques

There are several classic design modification techniques which can be reliably used to improve an existing lens design. They are:

1. Split an element into two (or more) elements
2. Compound a singlet into a doublet (or triplet)
3. Raise the index of the positive singlets
4. Lower the index of the negative singlets
5. Raise the index of the elements in general
6. Aspherize a surface (or surfaces)
7. Split a cemented doublet
8. Use unusual partial dispersion glasses to reduce secondary spectrum (see Chap. 6, "Telescope Objectives")

The simple, straightforward application of these techniques is no guarantee of improvement in a lens, in that they do not automatically correct the defects that they are intended to address. In general, these changes tend to reduce the aberration contributions of the modified components; in order to take full advantage of this, the aberrations of the balance of the system must be reduced as well. The operative principle is this: if large amounts of aberrations are corrected or balanced by equally large amounts of aberrations of opposite sign, then the residual aberrations also tend to be large. Conversely, if the balancing aberrations are both small, then the residuals tend to be correspondingly small.

3.2 Glass Changes: Index and *V* Value

The refractive characteristics of the materials used in a lens are obviously significant and important to the design. In general, for a posi-

tive element, the higher the index the better. The higher index reduces the inward Petzval curvature which plagues most lenses. It also tends to reduce most of the other aberrations as well. As an example, see Fig. 3.1, which clearly indicates the effect of higher index in reducing the spherical aberration of a single element. This sort of reduction is primarily a result of the fact that the surface curvature required to produce a given element power is inversely proportional to

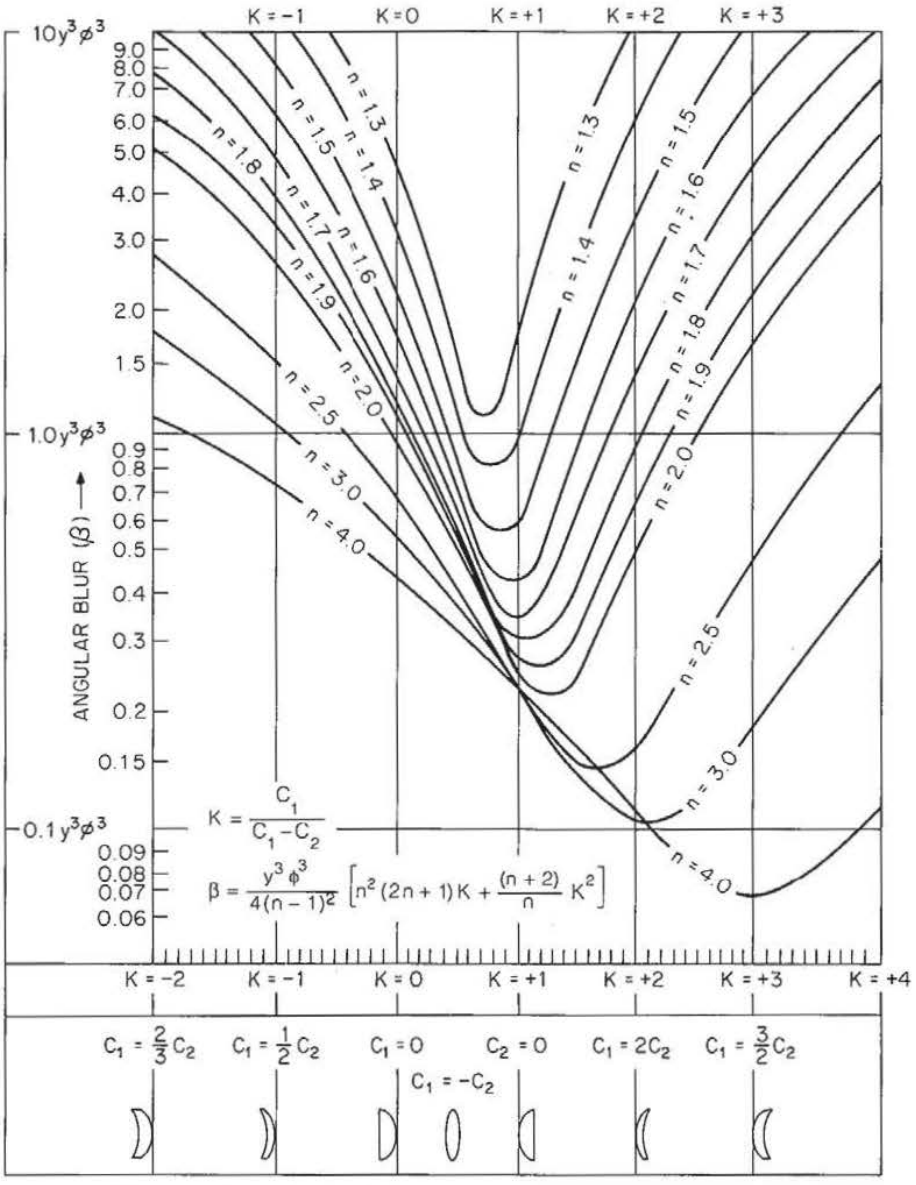


Figure 3.1 The angular spherical aberration blur of a single lens element as a function of lens shape, for various values of the index of refraction; ϕ is the element power and y is the semiaperture. The angular blur can be converted to longitudinal spherical aberration by $LA = 2B/y\phi^2$, or to transverse aberration by $TA = -2B/\phi$. (The object is at infinity.)

$(n - 1)$. The improvement also results from the reduction of the angles of incidence at the surfaces of the element.

In a negative element, the situation is less clear. From the standpoint of the Petzval correction, a low index would increase the overcorrecting contribution of a negative element. This can help to offset the (inward) undercorrection which is a major problem in most lenses. On the other hand, a higher index would reduce the surface curvatures and have a generally desirable effect on the overall state of correction. The situation is usually resolved with the negative elements made from a glass along the glass line boundary of the glass map (Fig. 2.2).

A high V value for the positive element and a low V value for the negative element of an achromatic doublet reduce the element powers; this is ordinarily desirable. In lenses (such as the Cooke triplet) where the relative V values of separated elements control the element spacing or the system length, this desideratum may be overridden by other concerns.

Note that, as usual, when you are dealing with components of negative focal length, many of the considerations outlined above are reversed. In a negative achromatic doublet, the negative element is often made of crown glass and the positive is made of flint. Here a high-index (flint) positive element will reduce the inward Petzval curvature, as will a low-index (crown) negative element.

3.3 Splitting Elements

Splitting an element into two (or more) approximately equal parts whose total power is equal to the power of the original element can reduce the aberration contribution by a significant factor. The reason that this reduces aberrations is that it allows the angles of incidence to be reduced; the nonlinearity of Snell's law means that smaller angles introduce less aberration than do large ones. This technique is often used in high-speed lenses to reduce the zonal spherical residual and in wide-angle lenses to control astigmatism, distortion, and coma.

Figure 3.2 shows the thin lens third-order spherical aberration for spherical-surfaced positive elements which are shaped (or bent) to minimize the undercorrected spherical. The upper plot shows the spherical as a function of the index of refraction for a single element with a distant object. The curve labeled $i = 2$ shows the spherical for two elements whose total power is equal to that of the single element. The best split is 50-50—i.e., the split elements have equal power; this minimizes the spherical. (The same is true for a split into more than two elements, i.e., three, four, etc., as shown in the curves labeled $i = 3$ and $i = 4$.) The improvement produced by splitting an element in

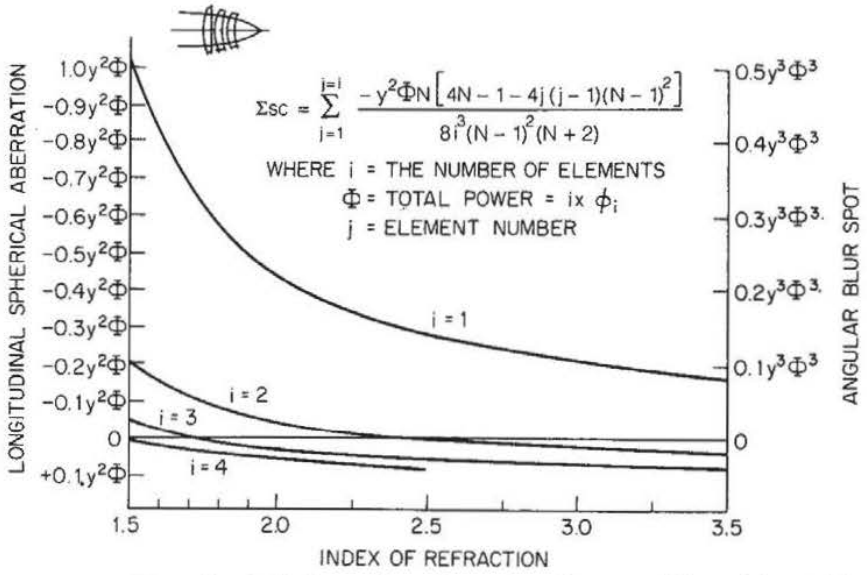


Figure 3.2 The spherical aberration of one, two, three, and four thin positive elements, each bent for minimum spherical aberration, plotted as a function of the index of refraction, and showing the reduction in the amount of aberration produced by splitting a single element into two or more elements (of the same total power). Each plot is labeled with *i*, the number of elements in the set. (The object is at infinity.)

two can be seen to be a factor of about 5 for lenses of index equal to 1.5. The higher the index, the greater the reduction; for an index of 1.8, the factor is about 7. At an index of 2.5 or higher, the spherical can be brought to zero or even overcorrected with just two positive elements. Most other aberrations are similarly affected by splitting, although it should be obvious that neither Petzval nor chromatic is changed by splitting.

In high-speed lenses this technique is frequently used to reduce the residual zonal spherical; the positive elements are split. This illustrates the basic idea. If the residual zonal spherical is negative (undercorrected), one splits a positive element; in the rare event that the zonal is positive (overcorrected), one would split a negative element. A similar philosophy can be applied for troublesome residuals of the other aberrations as well.

The choice of which element to split is often less apparent. The logical candidate would obviously seem to be the element which contributes most heavily to the problem aberration. (An examination of the third- and fifth-order surface contributions can often locate the source of the aberration.) However, other considerations often become significant. For example, in the Cooke triplet, the rear element is the prime candidate for the split, and such a split is quite effective in reducing

the zonal spherical, as Figs. 14.1 and 14.2 will attest. But the better choice for the split is the front element, not because it does a better job of reducing the zonal spherical, but because the resulting lens is better corrected for the other aberrations. Figure 14.3 shows the simple split-front triplet. This is the ancestor of the Ernostar family of lenses; Figs. 14.10 through 14.15 illustrate designs which can be considered as descendants from the split-front triplet. Although they have been largely superseded by the more powerful double-Gauss form, they are nonetheless excellent design types.

Many retrofocus and wide-angle lenses which use strong outer meniscus negative elements illustrate the use of this technique for the control of coma, astigmatism, and distortion by splitting these negative elements.

The implementation of this technique with an automatic design program is often far from easy. For example, if one decides to split one of the crowns of a Cooke triplet and simply replaces one crown with two, after the computer optimization has run its course, the resultant lens may look like an ordinary triplet with a narrow cracklike airspace in the split element (a *cracked crown triplet*). The performance of the lens is the same as the original triplet; the split has not improved a thing. This is because the original lens was in a local optimum of the merit function. Aberrations other than the zonal spherical dominated the design; this caused the program to return the lens to its original design configuration. What is necessary in this situation is to force the split elements into a configuration which will accomplish the desired result.

Consider the split-front triplet. There are two ways to get to a design like Fig. 14.3. One approach is to make the lens so fast that the zonal spherical is by far the single dominant aberration in the merit function. Then the program will probably choose a form which reduces the zonal spherical; the lens shapes in Fig. 14.3 are a likely result. A difficulty with this approach may be that you aren't interested in a very fast lens, or if you are, the rays may miss the surfaces of the initial design completely, or encounter TIR. The alternative approach is to constrain the front elements to a configuration in which the spherical is minimized. Simply fixing the first element to a plano-convex form (by not allowing the plano surface curvature to vary) or holding the second to an aplanatic meniscus shape is usually sufficient to obtain a stable design which is enough different from the cracked crown triplet. When this has been accomplished the constraint can be released and the automatic design routine allowed to find what is (one hopes) a new and better local optimum. The problem here is that this approach presupposes a knowledge of the configuration which will

produce a good result. Obviously a knowledge of both aberration theory and of successful design forms is a useful tool to the designer.

3.4 Separating a Cemented Doublet

Airspacing a cemented doublet can provide two additional degrees of freedom: two bendings instead of one, plus an airspace. While this technique does not have the inherent aberration reduction capability that many other modifications possess, the extra variables may indirectly make a design improvement possible. A difficulty in implementing this is that the refraction at the cemented surface is apt to become much more abrupt when it is split into two glass-air interfaces than when it was a cemented surface; in fact rays may encounter TIR if a simple split is attempted without a concomitant reduction of the angle of incidence. Manual intervention in the form of adjusting the radii to reduce the angle of incidence is often necessary.

3.5 Compounding an Element

Compounding a singlet to a doublet can be viewed in two different ways:

1. As a way of simulating a desirable but nonexistent glass type
2. As a way of introducing a cemented interface into the element in order to control the ray paths

Note that in almost all examples of Tessar-type lenses (and other types which utilize compounded elements), the doublets have positive elements with high index and high V values, while the negative element of the doublet has both a lower index and V value. See Chap. 12 for examples.

The longitudinal axial chromatic of a singlet is given by $LA_{ch} = -f/V$. Thus a fully achromatized lens (with $LA_{ch} = 0.0$) has the chromatic characteristic of a lens made from a material with a V value of infinity; a partially achromatized doublet acts like a singlet with a very high V value.

The Petzval radius of a singlet is given by $\rho = -nf$, where n is its index. An *old achromat* with a low-index crown and a higher-index flint has a shorter Petzval radius than a singlet of the crown glass. For example, an achromat of BK7 (517:649) and SF1 (717:295) has a Petzval radius $\rho = -1.37f$; in other words, in regard to Petzval field curvature, it behaves like a singlet with an index of 1.37. A *new achromat* has a high-index crown and a lower-index flint. A new

achromat of SSKN5 (658:509) and LF5 (581:409) has a Petzval radius $\rho = -2.19f$ and a Petzval curvature which is characteristic of a singlet with an index of 2.19.

Thus an achromatized (or partially achromatized) doublet with a high-index positive element and a low-index negative element has many of the characteristics of a lens made of a high-index, high-V-value crown glass. (Note that for a negative focal length doublet, the reverse is true.) Both conditions are usually to be desired, in order to flatten the Petzval field and to achieve achromatism.

Figure 3.3 shows a singlet, an old achromat, and a new achromat, each with the same focal length. The equivalent V value of each achromat is, of course, equal to infinity. The Petzval radius for each is given in the figure caption.

The cemented interface of the doublet can be used for specific control of specific rays. In a lens such as the Tessar, where the doublet is located well away from the aperture stop, the upper and lower rim rays of the oblique fan have very different angles of incidence at the cemented surface. In Fig. 3.4 it can be seen that the angle of incidence at this surface is much larger for the upper ray than for the lower. In this type of lens the cemented surface is typically a convergent one, and the (trigonometric) nonlinear characteristic of Snell's law means that the upper ray is, in this case, refracted downward more than it would be were the refraction linear with angle. Thus the upper ray is deviated in such a way as to reduce any positive coma of this ray. This

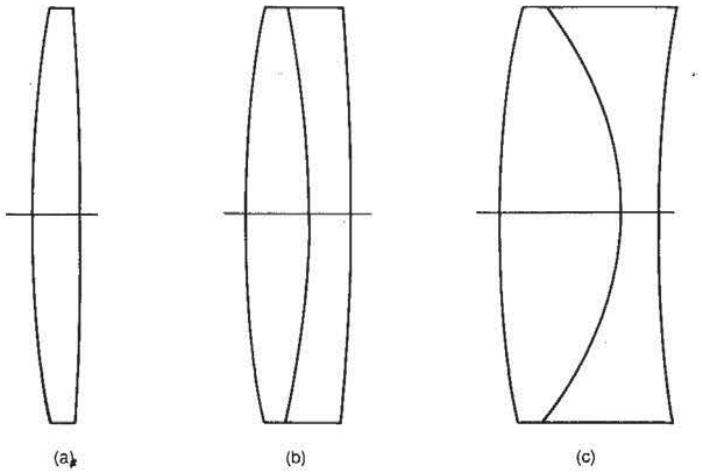


Figure 3.3 Three lenses, each with the same focal length f. (*a*) A singlet of BK7 (517:642) glass; Petzval radius equals $-1.52f$. (*b*) An old achromat of BK7 (517:642) and SF1 (717:295) glasses; Petzval radius equals $-1.37f$. (*c*) A new achromat of SSKN5 (658:509) and LF5 (581:409); Petzval radius is $-2.19f$.

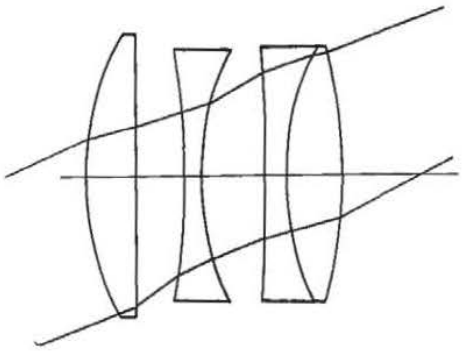


Figure 3.4 The upper and lower rim rays have significantly different angles of incidence at the cemented interface in the rear doublet of this Tessar design. Properly handled, this difference can be used to modify the correction of the coma-type aberrations.

illustrates the manner in which a cemented surface can be used for an asymmetrical effect on an oblique beam.

The Merté surface

A strongly curved, collective cemented surface with a small index break (to the order of 0.06) has an effect which can be used to reduce the undercorrected zonal spherical aberration. The central doublet of the Hektor lens shown in Fig. 3.5 illustrates this principle. The ce-

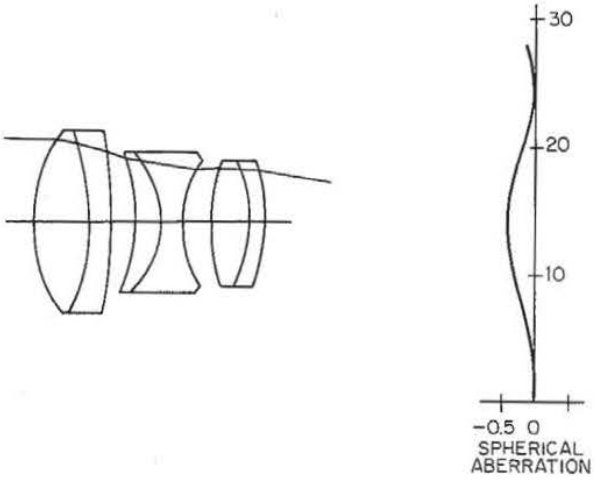


Figure 3.5 The cemented surface in the center doublet of this Hektor lens is what is called a Merté surface. The index break $(n' - n)$ across the surface is small, but at the margin of the aperture the angle of incidence for the axial ray becomes quite large. This combination produces an undercorrecting seventh-order spherical aberration which, as the plot shows, dominates the spherical aberration at the margin of the aperture, causing the marginal spherical to be negative rather than the usual positive value.

mented surface is a collective one (in that $[n' - n]/r$ is positive) and contributes undercorrected spherical aberration. For rays near the axis, the spherical aberration contribution of the surface is modest. However, when the ray intersection height increases and the angle of incidence becomes large, as shown in Fig. 3.5, the trigonometric nonlinearity of Snell's law causes the amount of ray deviation to be disproportionately increased. This causes the undercorrection from this surface to dominate the spherical aberration. The result is a spherical aberration characteristic like that shown in Fig. 3.5. The spherical in the central part of the aperture appears quite typical: the undercorrected third-order dominates close to the axis and the overcorrecting fifth-order causes the plot to curve back as the ray height increases. However, toward the edge of the aperture the undercorrection of the Merté surface becomes dominant and the aberration plot reverses direction again. The net result is the equivalent of a reduced zonal spherical aberration.

It is rare to see as extreme an example of the Merté surface as that illustrated in the Hektor of Fig. 3.5. Such a surface is very sensitive to fabrication errors and is thus expensive to make. It is also often best used close to the aperture stop because, if it is located away from the stop, the asymmetrical effects described two paragraphs above can become quite undesirable. However, it is well worth noting that an ordinary collective cemented surface has a tendency to behave as a mild Merté surface and to reduce the spherical zonal, at least somewhat.

3.6 Vignetting and Its Uses

Vignetting, which is simply the mechanical limitation or obstruction of an oblique beam, is usually regarded primarily as something which reduces the off-axis illumination in the image. However, vignetting often plays an essential role in determining the off-axis image quality as well as the illumination. Of course there are many applications for which vignetting cannot be tolerated; the illumination must be as uniform as possible across the entire field of view. The complexity of the lens design, therefore, must be sufficient to produce the required image quality at full aperture over the full field.

But for many applications, vignetting is, in fact, quite tolerable. In commercial applications the clear apertures may well be established so as to be just sufficient to pass the full aperture rays for the axial image. It is not at all unusual for vignetting to exceed 50 percent at the edge of the field. For a camera lens, this vignetting will, of course, completely disappear when the iris of the lens is stopped down to an aperture below the vignetting level. Since camera lenses are most of-

ten used at less than full aperture, the vignetting is not as significant as it is in a lens which is always used at full aperture, such as a microscope or projection lens.

The *benefit* of vignetting is that it cuts off the upper and/or lower rim rays of the oblique tangential fan. Since these are ordinarily the most poorly behaved rays, the image quality may well be improved by their elimination. Most lenses which cover a significant field are afflicted with oblique spherical aberration, a fifth-order aberration which looks like third-order spherical aberration, but which varies as the square of the field angle. And since its magnitude is different for sagittal than for tangential rays, it can be seen to have characteristics of both astigmatism and spherical aberration. Oblique spherical aberration usually causes the rays at the edge of the oblique bundle to show strongly overcorrected spherical aberration; vignetting is a simple way to block these aberrant rays from the image.

Another factor favoring the use of vignetting is that it results from lens elements with small diameters. In general, one can count on a smaller-diameter lens being less costly to fabricate.

For a camera lens, one must be certain that the iris diaphragm is located centrally in the oblique beam so that, when the iris is closed down, the central rays of the beam are the ones which are passed. These are usually the best-corrected rays of the oblique beam. Also this location assures that the vignetting will be eliminated at the largest possible aperture.

3.7 Eliminating a Weak Element; the Concentric Problem

Occasionally an automatic design program will produce a design with an element of very low power. Frequently this means that the element can be removed from the design without adversely affecting the quality of the design. Often a straightforward removal will not work; the design process may simply "blow up." An approach which usually works (if anything will) is to add the thickness and the surface curvatures of the element to the merit function with target values of zero, allowing them to continue as variable parameters. Sometimes targeting the difference between the two curvatures is also useful. Usually, if the element isn't necessary to the design, a few cycles of optimization, possibly with gradually increasing weights on these targets, will change the element to a very thin, nearly plane parallel plate, which can then be removed without severe trauma to the design. If your design program will not accept curvatures and thicknesses as targets, an alternative technique is to remove the curvatures and the thickness as variables and to gradually weaken the curvatures and reduce

the thickness (by hand) while continuing to optimize with the other variables.

An unfortunate form of the "weak" element is a fairly strongly bent meniscus, which the computer uses for a relatively important design function, such as the correction of spherical aberration or the reduction of the Petzval curvature. It is rarely possible to eliminate such an element because it is an integral part of the design. The unfortunate aspect of this situation is that, if the surfaces of the meniscus element are concentric or nearly so, the customary centering process used in optical manufacture is impossible or impractical, and the element is costly to fabricate. This situation can be ameliorated by forcing the centers of curvature of the surfaces apart by a distance sufficient to allow the use of ordinary centering techniques. Again, including the required center-to-center spacing in the merit function and reoptimizing will usually modify the offending element to a more manufacturable form without any significant damage to the system performance.

3.8 Balancing Aberrations

The optimum balance of the aberrations is not always the same in every case; the best balance varies with the application and depends on the size of the residual aberrations. In general, for well-corrected lenses, the aberrations should be balanced so as to minimize the OPD, i.e., the wavefront variance, but there are significant exceptions.

Spherical aberration

If a lens is well-corrected and the high-order residual spherical aberration is small, so that the OPD is to the order of a half-wave or less, then the best correction is almost always that with the marginal spherical corrected to zero, as illustrated in Fig. 3.6*b*. However, when the zonal spherical is large, there are two situations where one may want to depart from complete correction of the marginal spherical.

If the lens will always be used at full aperture (as a projection lens, for example), and if the spherical aberration residual is large (say to the order of a wave or so), the diffraction effects will be small when compared to the aberration blur; then the spherical aberration should be corrected to minimize the size of the blur spot rather than to minimize the OPD. This will produce the best contrast for an image with relatively coarse details, i.e., for a resolution well below the diffraction limit. As an example, at a speed of *f*/1.6, a 16-mm projection lens has a diffraction cutoff frequency of about 1100 line pairs per millimeter (lpm). But its performance is considered quite good if it resolves 100

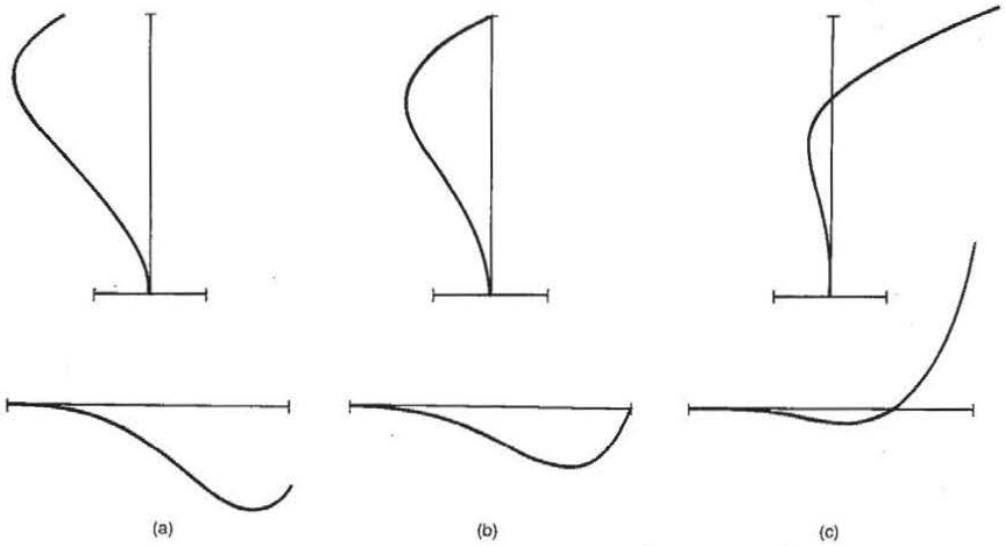


Figure 3.6 Three states of correction of spherical aberration are shown. Each has the same amount of fifth-order spherical, but different amounts of third-order. (*a*) Spherical aberration balanced to give the smallest possible size blur spot. This correction may be optimum when the aberration is large and the required level of resolution is low compared to the diffraction limit. (*b*) Spherical aberration balanced for minimum OPD. This is optimum when the system is diffraction-limited. (*c*) Spherical aberration balanced to minimize the focus shift as the lens aperture is stopped down. This correction is used in camera lenses when the residual spherical is large. The upper row is longitudinal spherical versus ray height; the lower row is transverse ray intercept plots.

lpm, an order of magnitude less than the diffraction limit. Such a lens can advantageously be corrected for the minimum diameter geometrical blur spot. This state of correction occurs (for third- and fifth-order spherical) when $LA_z = 1.5LA_m$, or $TA_z = 1.05TA_m$; the result is a high-contrast, but low-resolution, image. This correction is illustrated in Fig. 3.6*a*. See also the comments on defocusing in Secs. 2.4 and 2.7.

For a lens which is used at varying apertures, as is a typical camera lens, it is important that the best focus position not shift as the size of the aperture stop is changed. If the spherical aberration is corrected at the margin of the aperture, or corrected as described in the paragraph above, the position of the best focus will shift as the aperture is changed. The best focus will move toward the paraxial focus as the aperture is reduced. The state of correction which is often used in such a case is overcorrection of the marginal spherical, as shown on Fig. 3.6*c* (assuming an undercorrected zonal residual). The result is a design in which the focus is quite stable as the lens is stopped down. The *resolution* is better than it would be otherwise, but, at full aperture, the *contrast* in the image is quite low. This works out reasonably well in a high-speed camera lens because camera lenses are only infrequently used at full aperture. Typically, photographs are taken with the lens

stopped down well below the full aperture, and, when the camera is stopped down, this state of correction yields a much better photograph.

The three correction states shown in Fig. 3.6 also indicate the manner in which the spherical aberration is changed when the third-order aberration is changed. This is a typical situation often encountered in lens design: the fifth- (and higher-) order aberrations are relatively stable and difficult to change, but the third order is easily modified (by bending an element, for example). In the figure, all three illustrations have exactly the same amount of fifth-order spherical; the difference is solely in the amount of third-order. Note that, in the (upper) longitudinal plots, the change from one illustration to the next varies as y^2, whereas in the (lower) transverse plots the differences vary as y^3.

Chromatic aberration and spherochromatism

Here the question is how to balance the spherochromatism, which typically causes the spherical aberration at short (blue) wavelengths to be overcorrected and that at the long (red) wavelengths to be undercorrected. If the aberration is small (diffraction-limited), the best correction is probably with the chromatic aberration corrected at the 0.7 zone of the aperture. This means that the central half of the aperture area is undercorrected for color and the outer half of the aperture is overcorrected, as shown in Fig. 3.7*a*. But if the amount of the aberration is large, the spherical overcorrection of the blue marginal ray causes a blue flare and a low contrast in the image. In these circumstances the correction zone can advantageously be moved to (or toward) the marginal zone, as shown in Fig. 3.7*b*. This will probably reduce the resolution somewhat, because it increases the size of the core of the image blur, but it improves the contrast significantly and yields a more pleasing image, free of the blue flare and haze. This state of correction is accomplished by increasing the undercorrection of the chromatic aberration of the paraxial rays.

Astigmatism and Petzval field curvature

In a typical anastigmat lens the fifth-order astigmatism tends to become significantly undercorrected (i.e., negative) as the field angle is increased. In order to minimize the astigmatism over the full field, the third-order astigmatism is made enough overcorrected to balance the undercorrected fifth-order astigmatism. The result is the typical field curvature correction with the sagittal focal surface located inside the tangential focal surface in the central part of the field because of the overcorrected third-order astigmatism, and the reverse arrangement

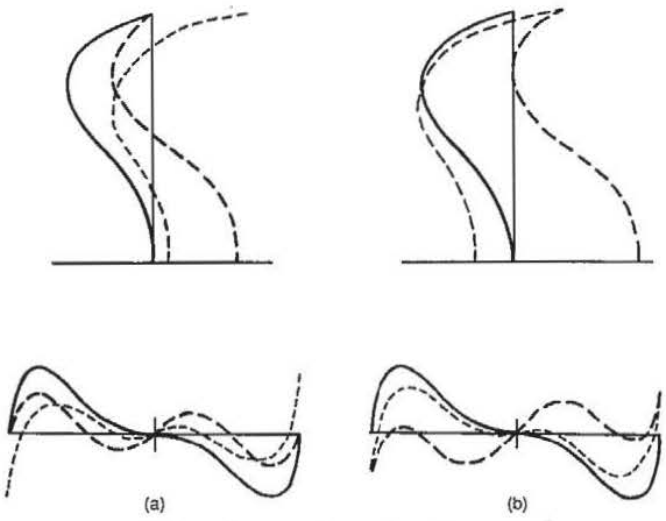


Figure 3.7 Spherochromatism. (*a*) Chromatic aberration balanced so that the outer half of the aperture is overcorrected and the inner half is undercorrected. This may be best if the amount of aberration is small. (*b*) Chromatic aberration balanced so that it is corrected at the margin. If the aberration is large, this correction eliminates the blue flare which can result from the type of correction in (*a*). Note that the state of correction is more easily perceived in the upper, longitudinal aberration plots, whereas the effect on the blur spot size and flare is much more apparent in the lower transverse ray intercept plots.

in the extreme outer portions. The field angle at which the s and t fields cross (i.e., where the astigmatism is zero) is called the *node*. Usually the two fields separate very rapidly outside the node, and the image quality quickly deteriorates, often suddenly. The Petzval curvature is usually made somewhat negative, so that both fields are slightly inward-curving and the effective field is as flat as possible. A typical state of correction is shown in Fig. 3.8.

Note that a field correction with the s and t focal surfaces spaced equally on either side of the focal plane (so that the compromise "smallest circle of confusion" focal surface is flat) is definitely *not* the best state of correction. In considering the correction of the field curvature, one should bear in mind that the oblique spherical aberration (a fifth-order aberration which varies as the cube of the aperture and the square of the field angle) typically goes overcorrected with increasing field angle. In addition, the oblique spherical is usually more significant for the tangential fan of rays than for the sagittal. Thus the effective field curvature for the full fan of rays is usually more backward-curving than the X_s and X_t field curves indicate. These curves indicate the imagery of a very small bundle of rays close to the principal ray, and do not take the oblique spherical of the full aperture

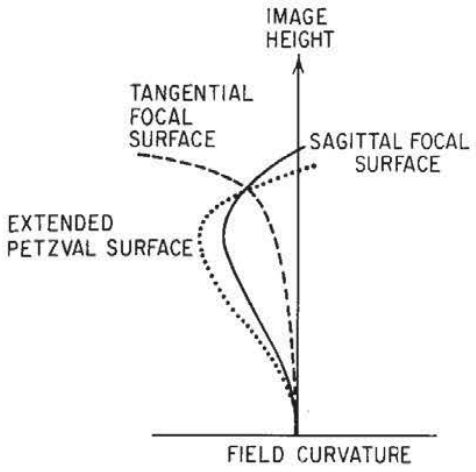


Figure 3.8 This is the typical balance of astigmatism and Petzval curvature in the presence of undercorrecting fifth-order astigmatism and overcorrecting high-order Petzval curvature. This is achieved by leaving the third-order Petzval slightly inward-curving and overcorrecting the third-order astigmatism by a small amount. This is the usual aberration balance for most anastigmats.

into account. Thus, for most designs, the astigmatism and field curvature are usually arranged somewhere between the state at which the s and t curves are superimposed (i.e., zero astigmatism) and that at which the t field is approximately flat. Often a through-focus MTF plot which includes both on-axis and off-axis plots will indicate quite clearly the effective field curvature, which is, of course, more informative than the X_s and X_t curves.

Note well that these discussions have assumed the type of higher-order residual aberrations which one ordinarily finds: overcorrected fifth-order spherical aberration, undercorrected fifth-order astigmatism, and overcorrected spherochromatism for the shorter wavelengths. Although rare, the reverse is sometimes encountered. In such circumstances the obvious move is to apply the above advice in reverse.

As an additional consideration, note that the undercorrection of either the chromatic aberration or the Petzval curvature has the usually desirable side effect of reducing the power of the elements of the lens system. Thus a secondary benefit of this undercorrection is the reduction of residual aberrations in general, because a lower-power element produces less aberration, which means that there is less higher-order residual aberration left when the aberrations are balanced out.

3.9 The Symmetrical Principle

When an optical system has mirror symmetry about the aperture stop (or a pupil), as shown in Fig. 3.9, the system is free of coma, distortion, and lateral color. This results from the fact that the components on one side of the stop have aberrations which exactly cancel the aberrations from the components on the other side of the stop. Obviously, to have mirror symmetry, the system must work at unit magnification,

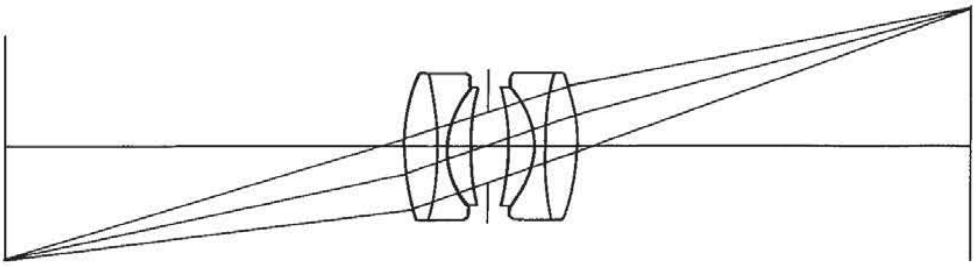


Figure 3.9 A fully (left to right) symmetrical system is completely free of coma, distortion, and lateral color, because the aberration in one half of the system exactly cancels out the aberration in the other half.

with equal object and image distances. For the symmetry to be *absolutely* complete, the object and image surfaces must be identical in shape; this then would imply separately curved sagittal and tangential surfaces at both object and image. However, the third-order coma, distortion, and lateral color are completely removed by symmetry, even with flat object and image surfaces.

Of course, most systems do not operate at unit magnification, and therefore a symmetrical construction of the lens will not completely eliminate these aberrations. However, even for a lens with an infinitely distant object, these aberrations are markedly reduced by symmetry, or even by an approximately symmetrical construction. This is why so many optical systems which cover a significant angular field display a rough symmetry of construction. Consider the Cooke triplet: it has outer crown elements which are similar, but not identical in shape, and the center flint, while not equi-concave, is bi-concave, and, except for slow-speed triplets, the airspaces are quite similar in size. The benefit of this is that the higher-order residuals of coma, distortion, and lateral color are markedly reduced by this symmetry. This is especially true for wide-angle lenses when good distortion correction is important.

3.10 Aspheric Surfaces

Many designs can be improved by the use of one or more aspheric surfaces. Except for the case of a molded or diamond-turned element, an aspheric surface is many times more expensive to fabricate than a spherical surface. A conic aspheric is easier to test than a general aspheric and is therefore somewhat less costly. For many systems, e.g., mirror objectives, aspheric surfaces are essential to the design and cannot be avoided.

One technique for introducing an aspheric into an optical system is to first vary only the conic constant. (Note that the conic constant and

the fourth-order deformation term have exactly the same effect on the third-order aberrations. Thus, allowing both to vary in an automatic design program may cause a slowing of the convergence or, in extreme cases, a failure of the process. Occasionally the difference between the effect of the conic and the fourth-order term on the fifth- and higher-order aberrations may be useful in a design, but more often than not the two are redundant.) If the effect of varying the conic constant alone is inadequate, one can then allow the sixth-order term to vary, then the eighth-order, etc. Some designs have aspherics specified to the tenth-order term when just the sixth or eighth would suffice. It is a good idea to calculate the surface deformation caused by the highest-order term used; if it is a fraction of a wave at the edge of the surface aperture, its utility may well be totally imaginary.

Occasionally one encounters a design specification or print in which the aspheric is specified by a tabulation of sagittal heights instead of an equation. The optimization program can be used to fit the constants of the standard aspheric surface equation to the tabulated data. The specification table is entered in the merit function as the sag of the intersections of (collimated) rays at the appropriate heights. The surface coefficients are allowed to vary, and the result is a least-squares fit to the sag table.

The equations of Sec. F.11 indicate the effects of a conic or a fourth-order aspheric term on the third-order aberrations. Several points are worthy of note. The conic has no effect on the Petzval curvature or on axial or lateral chromatic. Further, if the conic is located at the aperture stop or at a pupil, then the principal ray height, y_p, is zero and the conic has no effect on third-order coma, astigmatism, or distortion; it can only affect third-order spherical. In the Schmidt camera the aspheric surface is located at the stop because the coma and astigmatism are already zero, because the stop is at the center of curvature of the spherical mirror; the purpose of the aspheric is to change *only* the spherical aberration. Conversely, if the purpose of an aspheric is to affect the coma, astigmatism, or distortion, then it must be located a significant distance from the aperture stop.

It is also worth noting that the primary effect of the conic, or fourth-order, deformation term is on the third-order aberrations. The primary effect of the sixth-order deformation term is on the fifth-order aberrations, etc., etc.

Chapter

4

Evaluation: How Good Is This Design?

4.1 The Uses of a Preliminary Evaluation

At some point in the process of designing an optical system, the designer must decide whether the design is good enough for the application at hand. With modern computing power, it is not a difficult matter to calculate the MTF or the point spread function (PSF), and to accurately include the effects of diffraction in the calculations. The process does consume a finite amount of time, however (which, on a slow computer, may be a significant amount), and it is useful to be able to make a reasonable estimate of the system performance from a more limited amount of data. A good estimate can avoid wasting time and computer paper in evaluating a clearly deficient design, or it can signal an appropriate point at which to conduct a full-dress evaluation.

4.2 OPD versus Measures of Performance

The distribution of illumination in the point spread function, particularly in the diffraction pattern of a reasonably well-corrected lens, is often used as a measure of image quality. The *Strehl ratio* (or *Strehl definition*) is the ratio of the illumination at the center of an (aberrated) point image to the illumination at the center of the point image formed by an aberration-free system. Figure 4.1 illustrates the concept. Another measure of image quality uses the percentage of the total energy in the point image which is contained within the diameter of the Airy disk. This diameter remains relatively constant in size for small amounts of aberrations. The table of Fig. 4.2 gives the relationships between the wavefront deformation (or OPD), the Strehl ratio, and the energy distribution.

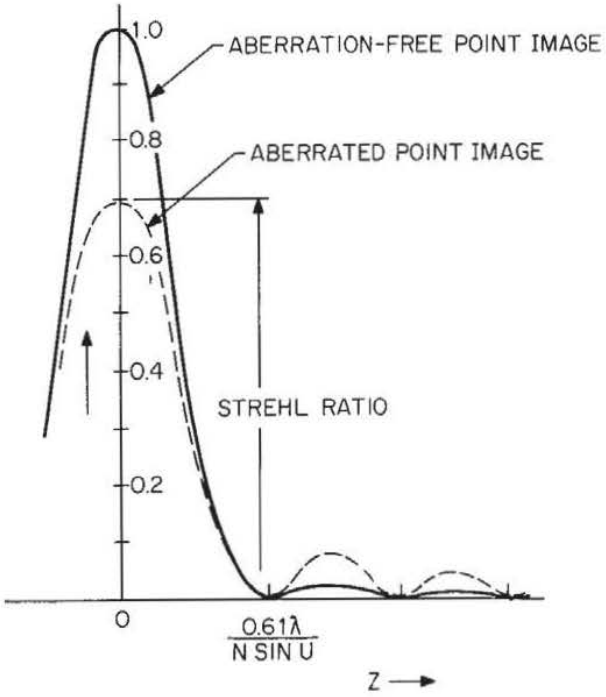


Figure 4.1 The Strehl ratio is the illumination at the center of the diffraction pattern of an aberrated image, relative to that of an aberration-free image.

Relation of Image Quality Measures to OPD

P-V OPD	RMS OPD	Strehl Ratio	% energy in Airy Disk	% energy in Rings
0.0	0.0	1.00	84	16
0.25RL = λ/16	0.018λ	0.99	83	17
0.5RL = λ/8	0.036λ	0.95	80	20
1.0RL = λ/4	0.07λ	0.80	68	32
2.0RL = λ/2	0.14λ	0.4*	40	60
3.0RL = 0.75λ	0.21λ	0.1*	20	80
4.0RL = λ	0.29λ	0.0*	10	90

*The smaller values of the Strehl ratio do not correlate well with image quality.

Figure 4.2 Tabulation of the Strehl ratio and the energy distribution as a function of the wavefront deformation. RL means the Rayleigh limit of one-quarter-wavelength peak-to-valley OPD.

Another commonly utilized measure of performance is the modulation transfer function, which describes the image modulation or contrast as a function of the spatial frequency of the object or image. The MTF of a perfect, aberration-free system is given by

$$\mathrm{MTF}(v) = \frac{2}{\pi}(\phi - \cos\phi \sin\phi) \tag{4.1}$$

where

$$\phi = \arccos\left(\frac{\lambda v}{2\mathrm{NA}}\right) \tag{4.2}$$

This is plotted as curve *A* in Figs. 4.3 and 4.4. Figure 4.3 shows the effect on the MTF of defocusing an otherwise aberration-free lens. The spatial frequency in these plots is normalized to the cutoff frequency

$$v_0 = \frac{2\mathrm{NA}}{\lambda} = \frac{1}{\lambda(f\,\mathrm{number})} \tag{4.3}$$

Figure 4.4 shows the effect of simple third-order spherical aberration on the MTF. Note that, although the curves of Figs. 4.3 and 4.4 are not identical, they *are* quite similar. This similarity of effect is the basis for the common rule of thumb that a given amount of OPD will de-

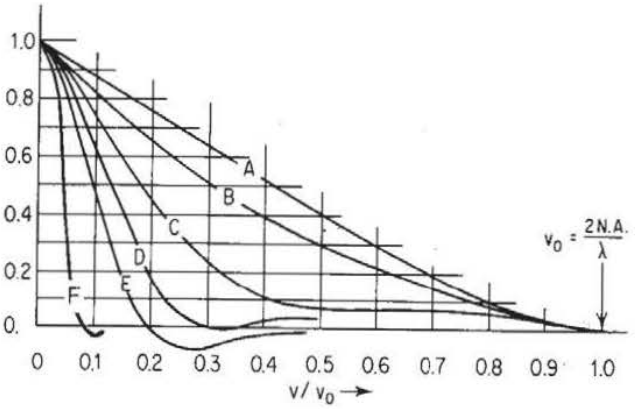


Figure 4.3 The effect of defocusing on the modulation transfer function of an aberration-free system.
(*A*) In focus OPD = zero
(*B*) Defocus = $\lambda/2n \sin^2 U$ OPD = $\lambda/4$
(*C*) Defocus = $\lambda/n \sin^2 U$ OPD = $\lambda/2$
(*D*) Defocus = $3\lambda/2n \sin^2 U$ OPD = $3\lambda/4$
(*E*) Defocus = $2\lambda/n \sin^2 U$ OPD = λ
(*F*) Defocus = $4\lambda/n \sin^2 U$ OPD = 2λ
(Curves are based on diffraction effects—not on a geometric calculation.)

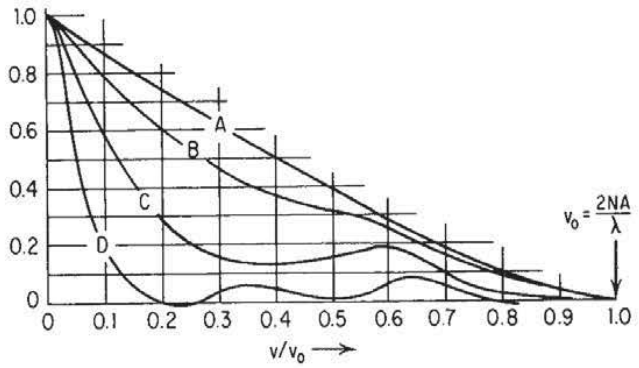


Figure 4.4 The effect of third-order spherical aberration on the modulation transfer function.
(*A*) LA_M = zero OPD = 0
(*B*) $LA_M = 4\lambda/n \sin^2 U$ OPD = $\lambda/4$
(*C*) $LA_M = 8\lambda/n \sin^2 U$ OPD = $\lambda/2$
(*D*) $LA_M = 16\lambda/n \sin^2 U$ OPD = λ

grade the image by the same amount regardless of what type of aberration produced the OPD.

When the OPD is large (say more than one or two waves), the following geometrical approximation (derived from the geometric defocusing expression) can be used to calculate the MTF with reasonable accuracy:

$$\text{MTF}(v) = \frac{J_1[8\pi n\text{OPD}(v/v_0)]}{4\pi n\text{OPD}(v/v_0)} \tag{4.4}$$

where v_0 is the cutoff frequency (Eq. 4.3), n is the index of the image medium, OPD is the peak-to-valley wavefront deformation in waves, and

$$J_1[x] = \frac{x}{2} - \frac{(x/2)^3}{1^2 2} + \frac{(x/2)^5}{1^2 2^2 3} - \cdots$$

The relationships between the basic aberrations and the OPD are given in Sec. F.12, as are the relationships between rms OPD and peak-to-valley OPD and between rms OPD and the Strehl ratio.

A convenient relationship to remember is that a quarter-wave of OPD corresponds to a transverse spherical aberration (either marginal or zonal) of about

$$\text{TA} = \frac{4\lambda}{\text{NA}} \tag{4.5}$$

This is a useful way to make a quick and dirty evaluation from just the ray intercept plots.

4.3 Blur Spot Size versus Certain Aberrations

Many times, the system characteristic of interest is the size of the blur produced as the image of a point source. There are a few simple relationships which are useful in this regard:
Third-order spherical at best focus (three/fourths of the way from paraxial to marginal focus):

$$B = 0.5\,\mathrm{LA}_m \tan U_m = 0.5\mathrm{TA}_m \tag{4.6}$$

Third- and fifth-order spherical (with the marginal spherical corrected, focused at 0.42 LA_z from the paraxial focus):

$$B = 0.84\,\mathrm{LA}_z \tan U_m = 0.59\,\mathrm{TA}_z \tag{4.7}$$

Third- and fifth-order spherical (corrected so that $\mathrm{LA}_z = 1.5\,\mathrm{LA}_m$, or $\mathrm{TA}_m = 1.06\,\mathrm{TA}_z$, and focused at 0.83 LA_z from the paraxial focus; this correction yields the smallest-diameter blur spot for a given amount of fifth-order spherical):

$$\begin{aligned} \mathrm{B} &= 0.5\,\mathrm{LA}_m \tan U_m = 0.5\mathrm{TA}_m \\ &= 0.33\,\mathrm{LA}_z \tan U_m = 0.47\mathrm{TA}_z \end{aligned} \tag{4.8}$$

Note that the above are based on the idea of the smallest spot containing 100 percent of the energy in the image of a point. For many applications this concept is valid and useful, but for best image quality there is usually another focus or correction at which the image has a smaller, brighter core and a larger flare; this is usually judged to be better for definition and pictorial purposes.

The effect of a large amount of defocusing on a well-corrected image is to produce a uniformly illuminated blur disk with a diameter of

$$B = 2(\text{defocus}) \tan U_m \approx \frac{\text{defocus}}{f\,\text{number}} \tag{4.9}$$

Astigmatism and field curvature can be evaluated by applying Eq. 4.9 separately in the sagittal and tangential meridians.

Although ordinary axial chromatic is also defocusing, the blur it produces is not uniformly illuminated, but has the energy centrally concentrated. At the midway focus point, the diameter of the blur containing 100 percent of the energy is

$$B = \mathrm{LA}_{\mathrm{ch}} \tan U_m = \mathrm{TA}_{\mathrm{ch}} \tag{4.10}$$

However, the central concentration leads to a situation where 75 to 90 percent of the energy is in a spot only half this size, and 40 to 60 percent is in a spot one-quarter as large. (The smaller percentages apply

for a uniform spectral response distribution and the larger for a triangular spectral distribution.)

For third-order coma, the blur is the typical comet shape, and has a height equal to the tangential coma and a width (in the sagittal direction) two-thirds this size. Note, however, that some 50 to 60 percent of the energy in the coma patch is in the point of the figure, whose size equals one-third the tangential coma.

4.4 MTF—The Modulation Transfer Function

The interpretation of an MTF plot is often problematical; it is not the easiest thing in the world to decide how good an image is on the basis of an examination of its MTF plot.

The limiting resolution is easily determined if the system sensor can be characterized by an aerial image modulation (AIM) curve. This is a plot of the threshold, or minimum, modulation required in the image for the sensor to produce a response. When plotted against spatial frequency, the intersection of the AIM curve and the MTF plot clearly indicates the limiting resolution, as shown in Fig. 4.5.

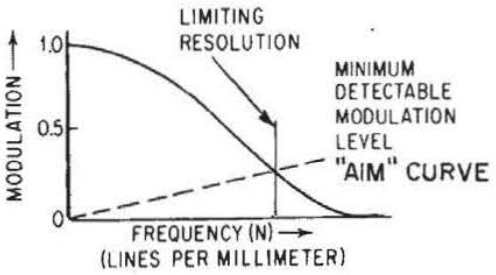


Figure 4.5 The intersection of the AIM curve and the MTF curve indicates the limiting resolution of the system.

A criterion for *excellent* performance (one which is often used as a design goal for top-of-the-line professional motion picture camera lenses) is to look for a 50 percent MTF at 50 lpm. Another criterion which has been presented for commercial 35-mm camera lenses is 20 percent MTF at 30 lpm over 90 percent of the field. Both criteria are applied at full aperture. These will give some idea of the range of the MTF values which are more or less standard for this type of work.

Chapter

5

Lens Design Data

5.1 About the Sample Lenses

One of the features of this volume is a fairly large set of lens designs, their prescriptions, and their aberration plots. This set is not intended to be a complete or extensive *collection* of all or even most of the published lens designs. We happily leave that to others. The set is intended to be a *selection* of lens designs which will serve two primary purposes. These are to serve as a set of suitable starting designs and to serve as a set of designs which illustrate to the reader the principles and techniques of successful lens designs.

The designs in this book were drawn from many sources. A significant number are from the original version of OPTICS TOOLBOX.* (All of the designs in this text, plus many others which were considered but not chosen for inclusion, have been incorporated in the *Warren J. Smith Lens Library.*†) Many are derived from the patent literature, or books which include patent references. Some of the designs are from the technical literature, such as journals, proceedings, or other books about lenses. Some have never been published previously.

In most cases the published designs have been modified to some extent. For the majority of the designs, we have specified the optical glass as one of those from the Schott (Schott Glass Technologies, Inc.) catalog. We have chosen what we feel is the nearest Schott glass to that indicated in our source for the lens data. Occasionally this may constitute a significant change, but we have attempted to stay as close to the original data as possible. In a few designs, non-Schott glasses have been used.

*OPTICS TOOLBOX is a product of Genesee Optics Software, Inc.

†Warren J. Smith Lens Library is a trademark of Genesee Optics Software, Inc., and is incorporated in their optical design software products.

The aperture and field which are indicated for any given lens design are more a matter of taste than anything else. What constitutes an acceptable level of aberration depends mostly on the application to which the lens is to be put. Thus the values for field and aperture which accompany each design in this book have often been selected somewhat arbitrarily to yield a level of correction which we thought reasonable.

The choice of the clear apertures for the lens elements is equally arbitrary. Obviously, the clear aperture of an element cannot be so large that the edge thickness at that diameter becomes negative or impractically thin. We have selected what seemed to be reasonable values for the clear apertures, based on both edge thickness considerations and the choice of a vignetting factor which allows a reasonably sized oblique beam through the lens and also trims the oblique beam to eliminate the worst-behaved rays.

5.2 Lens Prescriptions, Drawings, and Aberration Plots

The lens design data and the associated graphics for this book have been produced by computer. While data input errors and other glitches are always possible, by producing the lens data table, the lens drawing, and the aberration plots all from the same lens data file, we hope to prevent most of the errors which have afflicted some other efforts of this type. The design examples have all been scaled to focal lengths which are within a few percent of 100 units in order that comparisons can be easily made. So that the details of the aberration correction will be readily apparent, the computer was programmed to select the scale of each aberration plot to make the plot fill the available space. This does, of course, have the disadvantage that each lens and its aberrations may be plotted to a different scale. In order to minimize this disadvantage we have limited the scales used to decimal factors of 2, 5, and 10.

Lens prescription

A sample is shown in Fig. 5.1. The lens construction data are tabulated in a quite straightforward way. The columns are headed *radius, thickness, mat'l, index, V-no,* and *sa* (for semiaperture); the meanings should be apparent. The radius value follows the usual sign convention that a positive radius has its center of curvature to the right of the surface. Plano surfaces (i.e., with infinite radius) are indicated by a blank entry in the radius column. The thickness and material following a surface are presented on the same line as the surface radius, and have the same number.

F/4.5 25.2deg TRIPLET US 1,987,878/1935 SCHNEIDER

radius	thickness	mat'l	index	V-no	sa
26.160	4.916	LAK12	1.678	55.2	11.7
1201.700	3.988	air			11.7
-83.460	1.038	SF2	1.648	33.8	10.2
25.670	4.000	air			10.2
	6.925	air			9.2
302.610	2.567	LAK22	1.651	55.9	10.3
-54.790	81.433	air			10.3

```
EFL   = 98.56                 = EFFECTIVE FOCAL LENGTH
BFL   = 81.43                 = BACK FOCAL LENGTH
NA    = -0.1127 (F/4.4)       = NUMERICAL APERTURE (F-NUMBER)
GIH   = 46.33 (HFOV=25.17)    = IMAGE HEIGHT (HALF FIELD IN DEGREES)
PTZ/F = -2.831                = (PETZVAL RADIUS)/EFL
VL    = 23.43                 = VERTEX LENGTH
OD    infinite conjugate      = OBJECT DISTANCE
```

Figure 5.1 Sample lens prescription.

With few exceptions, the material names are those of Schott Glass Technologies, Inc. The index and V number values correspond to the wavelengths given with the ray intercept plot (e.g., see Fig. 5.3); for most lenses we have used the *d, F,* and *C* lines. The location of the aperture stop is indicated by a blank in the radius column with air on both sides of the surface. Aspheric surfaces are specified by the conic constant kappa and/or the aspheric deformation coefficients. The equation for the surface is

$$x = \frac{cy^2}{1 + [1 - (1 + \kappa)c^2y^2]^{1/2}} + ADy^4 + AEy^6 + AFy^8 + AGy^{10} \qquad (5.1)$$

The data below the prescription tabulation has the following meanings:

EFL	Effective focal length
BFL	Back focal length (the distance from the last surface to the paraxial focal point)
NA	Numerical aperture (the corresponding *f* number is in parentheses)
GIH	Gaussian (paraxial) image height (half-field in degrees is in parentheses)
PTZ/*F*	Petzval radius as a fraction of EFL
VL	Vertex length from first to last surface
OD	Object distance

Lens drawing

A sample lens drawing is shown in Fig. 5.2. The scale of the lens drawing is indicated by the dimensioned length of the line immediately below the lens sketch. The two rays in the sketch are the marginal and principal rays corresponding to the aperture and field angle which are

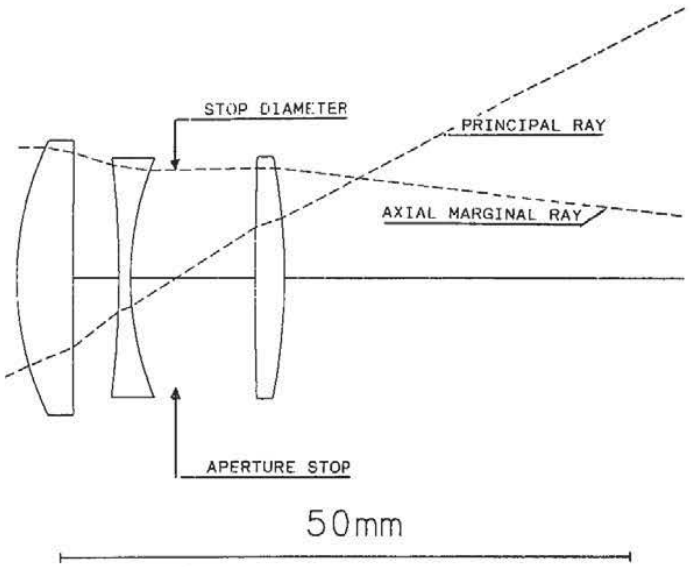


Figure 5.2 Sample lens drawing.

tabulated with the lens data. The aperture stop location is indicated by the point at which the principal ray crosses the optical axis. The lens elements are drawn to the clear apertures given in the prescription table as *sa*, the semiaperture.

Additional rays can easily be added to the lens drawing if desired, by using a technique which is exact only for paraxial rays, but which is often accurate enough for use in estimating or drawing ray paths. A ray may be scaled by simply multiplying the heights at which it strikes the surfaces by a scaling constant. Also, rays may be added by adding their intersection heights together. In each case the result is a reasonable approximation to the path of another ray. Obviously, two rays can be scaled and then added. Thus any desired third ray can be drawn by determining its intersection heights from

$$Y_3 = AY_1 + BY_2 \tag{5.2}$$

where A and B are scaling factors and Y_1 and Y_2 are the ray heights of the rays in the lens drawing. If one defines the desired third ray by its intersection with any two surfaces (which may include the object or image surface), then a simultaneous solution for A and B may be found from the two equations which result when the appropriate values of Y_1, Y_2, and Y_3 are substituted into the equation above.

Aberration plots

A sample aberration plot is shown in Fig. 5.3. The aberration plots include both tangential and sagittal ray intercept plots (sometimes

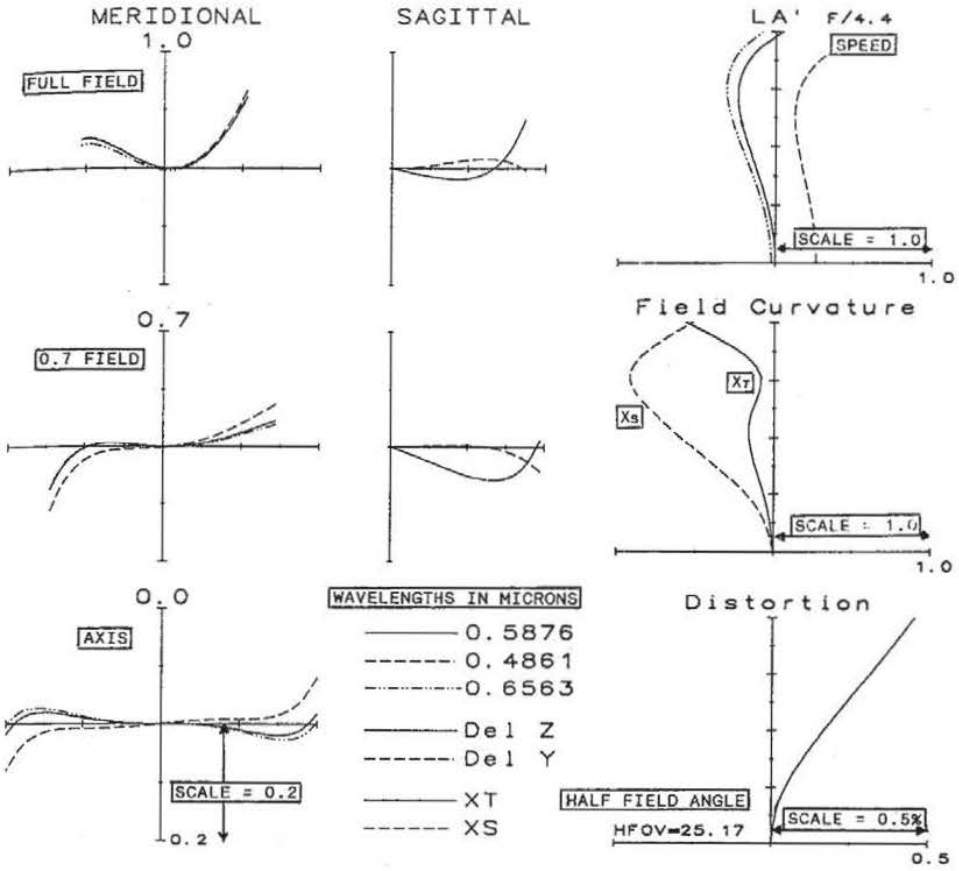


Figure 5.3 Sample aberration plot.

called *H*-tan *U* curves) for the axis, 0.7 field, and full field. The ray displacements are plotted vertically, as a function of the position of the ray in the aperture. The vertical scale is given at the lower end of the vertical bar for the axial plot; the number given is the half-length (i.e., from the origin to the end) of the vertical line in the plot. The horizontal scale is proportional to the tangent of the ray slope angle. Following the usual convention, the upper ray of the ray fan is plotted to the right. In the sagittal plots, the solid line is the transverse aberration in the *z*, or sagittal, direction and the dashed line is the ray displacement in the *y* direction (which is sagittal coma).

In addition to the ray intercept plots (which are, in general, probably the most broadly useful presentation of the aberration characteristics of a design), two aberrations are also presented as longitudinal plots. The longitudinal representations of spherical aberration and field curvature have been the classical, conventional presentation for decades, despite the fact that they give a very incomplete picture of the state of correction of the lens. However, a longitudinal plot of the spherical aberration in three wavelengths does allow a much clearer

understanding of the spherochromatism, as well as the secondary spectrum. The scale factor for this plot is the number given at the right end of the horizontal axis; the number is the half length of the horizontal line. The vertical dimension of the plot is the height of the ray at the pupil; the f number is given at the top of the plot. This is the f number of the imaging cone and is equal to 1/2NA. The longitudinal field curvature plots yield an excellent picture of the correction of the Petzval curvature and the astigmatism. The scale for X_s and X_t is given at the right end of the horizontal axis; again the number is the half length of the horizontal line. The solid line is X_t and the dashed line is X_s. The vertical scale is the fraction of the gaussian image height (GIH); the half field angle is given at the left side of the distortion plot. The scale for distortion is in percent, and the number is the half-length of the horizontal line.

5.3 Estimating the Potential of a Design

It is relatively easy to estimate the effects of a modest redesign on the aberration plots of an existing design by applying a knowledge of third-order aberration theory. This is because the third-order aberrations of a lens are easily adjusted by changing the spaces or the shapes of the elements, whereas the amount of higher-order aberration tends to be quite stable and resistant to change.

Equations 5.3 and 5.4 are a power series expansion of the relationships between the ray intersection with the image plane (y', z') as a function of the object height h and the ray position in the pupil (defined in polar coordinates s and θ), as shown in Fig. 5.4.

$$\begin{aligned} y' = {} & A_1 s\cos\theta + A_2 h + B_1 s^3\cos\theta + B_2 s^2 h(2+\cos 2\theta) \\ & + (3B_3 + B_4) s h^2 \cos\theta + B_5 h^3 + C_1 s^5 \cos\theta + (C_2 + C_3\cos 2\theta) s^4 h \\ & + (C_4 + C_6\cos^2\theta) s^3 h^2 \cos\theta + (C_7 + C_8\cos 2\theta) s^2 h^3 + C_{10} s h^4 \cos\theta \\ & + C_{12} h^5 + D_1 s^7 \cos\theta + \cdots \end{aligned} \tag{5.3}$$

$$\begin{aligned} z' = {} & A_1 s\sin\theta + B_1 s^3 \sin\theta + B_2 s^2 h \sin 2\theta \\ & + (B_3 + B_4) s h^2 \sin\theta + C_1 s^5 \sin\theta + C_3 s^4 h \sin 2\theta \\ & + (C_5 + C_6\cos^2\theta) s^3 h^2 \sin\theta + C_9 s^2 h^3 \sin 2\theta + C_{11} s h^4 \sin\theta \\ & + D_1 s^7 \sin\theta + \cdots \end{aligned} \tag{5.4}$$

Notice that, in the A terms, the exponents of s and h are unity. In the B terms, the exponents total 3, as in s^3, s^2h, sh^2, and h^3. In the C

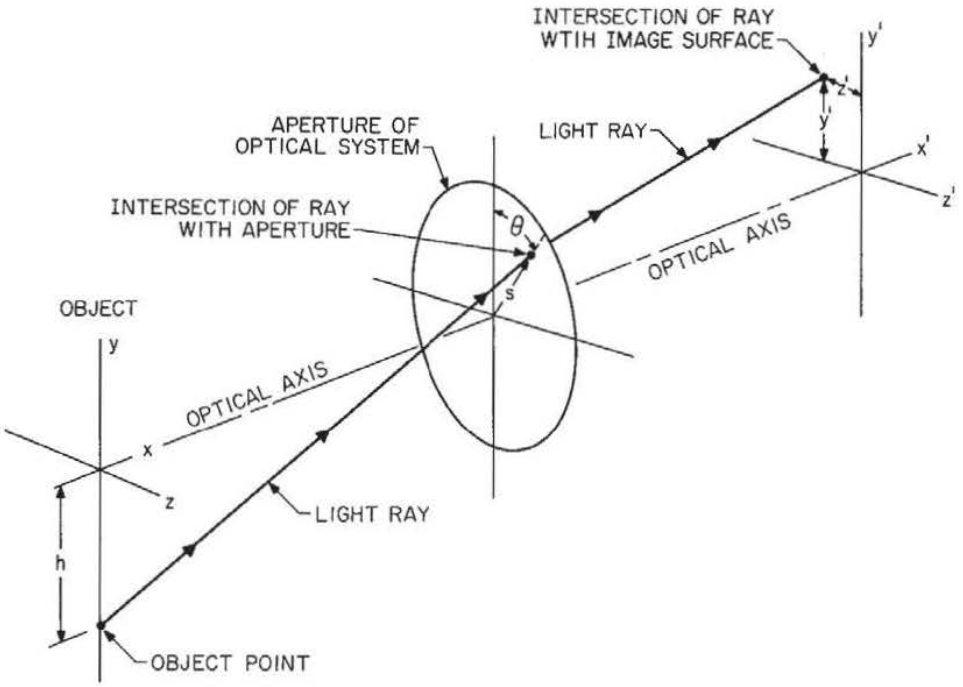


Figure 5.4 A ray from the point $y = h$, $z = 0.0$ in the object passes through the aperture of the optical system at a point defined by its polar coordinates (s, θ), and intersects the image surface at point y', z'.

terms, the exponents total 5, and in the D terms, 7. These are referred to as the first-order, third-order, and fifth-order terms, etc. There are 2 first-order terms, 5 third-order, 9 fifth-order, and $[(n + 3)(n + 5)/8 - 1]$ nth-order terms. In an axially symmetrical system there are no even-order terms; only odd-order terms may exist (unless we depart from symmetry as, for example, by tilting a surface or introducing a toroidal or other nonsymmetrical surface).

It is apparent that the A terms relate to the paraxial (or first-order) imagery. A_2 is simply the magnification (h'/h), and A_1 is a measure of the distance from the paraxial focus to our "image plane." All the other terms in Eqs. 5.3 and 5.4 are called *transverse aberrations.* They represent the distance by which the ray misses the ideal image point as described by the paraxial imaging equations.

The B terms are called the *third-order,* or *Seidel,* or *primary* aberrations. B_1 is spherical aberration, B_2 is coma, B_3 is astigmatism, B_4 is Petzval, and B_5 is distortion. Similarly, the C terms are called the *fifth-order* or *secondary* aberrations. C_1 is fifth-order spherical aberration; C_2 and C_3 are linear coma; C_4, C_5, and C_6 are oblique spherical aberration; C_7, C_8, and C_9 are elliptical coma; C_{10} and C_{11} are Petzval and astigmatism; and C_{12} is distortion.

The 14 terms in D are the seventh-order or tertiary aberrations; D_1 is the seventh-order spherical aberration. A similar expression for OPD, the wavefront deformation, is given in Sec. 5.4 (Eqs. 5.5 and 5.6).

The terms with the B coefficients are the transverse third-order aberrations. The longitudinal aberrations are equal to the transverse aberrations divided by $-u_k'$; since u_k' is a direct function of the ray height s, one can convert these into longitudinal aberrations by reducing the exponent of s by 1 (and adjusting the coefficients).

Thus one can assume that if the spherical aberration is adjusted so as to correct it at a given aperture, the change in the longitudinal aberration will be proportional to the square of the ray heights. For example, if the spherical at ray height Y_1 is to be changed by $d\text{LA}_1$, then at ray height Y_2 it will change by $d\text{LA}_2 = d\text{LA}_1(Y_2/Y_1)^2$. The change of the transverse spherical will vary as a cubic function, so that the transverse change $d\text{TA}_2 = d\text{TA}_1(Y_2/Y_1)^3$. Figure 3.6 shows the effect of changing the third-order spherical (assuming a constant fifth-order spherical).

If the axial chromatic is changed, the ray intercept plots for the different colors will simply rotate with respect to each other. The secondary spectrum and spherochromatism will change very little, so that one can readily estimate the ray intercept plots which will result from a simple change in the axial chromatic. Figure 3.7 shows two different balances of axial chromatic and spherochromatism. Similarly, a lateral chromatic change will change the relative heights of the different color ray plots, and the amount of the height change will be proportioned to the image height.

The change in the longitudinal astigmatism and field curvature is proportional to the square of the image height or field angle. Thus if the field curvature is changed by dX_1 at image height H_1, the change at H_2 is $dX_2 = dX_1(H_2/H_1)^2$. The change in the tangential field curvature X_t is three times the change in the sagittal field curvature X_s if the change is produced by changing the amount of the astigmatism. However, if the change results from a change in the Petzval curvature, both X_s and X_t are shifted by the same distance. Note that the slope of the ray intercept plot ($dH'/d \tan U$) at the principal ray is equal to the tangential field curvature (X_t) (or X_s for the sagittal plot).

Changes in the third-order coma produce a parabolic-shaped change in the ray intercept plot. If the plot is raised by an amount dH at the ends of the plot, it will be raised by $(0.7)^2\, dH = 0.5\, dH$ at the 0.7 zones of the aperture. The amount of the coma change for other field angles will vary directly with the field angle or image height.

The change in percent distortion will vary with the square of the field angle. The change in the lateral color varies directly with the field angle.

Thus one can look at the aberration plots for a given design and, by applying the techniques outlined above, easily visualize what they will look like after an adjustment has been made to fit the design to the application at hand.

5.4 Scaling a Design, Its Aberrations, and Its MTF

A lens prescription can be scaled to any desired focal length simply by multiplying all of its dimensions by the same constant. All of the *linear* aberration measures will then be scaled by the same factor. Note however, that percent distortion, chromatic difference of magnification (CDM), the numerical aperture or f number, aberrations expressed as angular aberrations, and any other *angular* characteristics remain completely unchanged by scaling.

The exact *diffraction* MTF cannot be scaled with the lens data. The diffraction MTF, since it includes diffraction effects which depend on wavelength, will not scale because the wavelength is not (ordinarily) scaled with the lens. A *geometric* MTF can be scaled by dividing the spatial frequency ordinate of the MTF plot by the scaling factor. Of course, because it neglects diffraction, the geometric MTF is quite inaccurate unless the aberrations are very large (and the MTF is correspondingly poor).

A diffraction MTF can be scaled *very* approximately as follows: Determine the OPD which corresponds to the MTF value of the lens for several spatial frequencies. This can be done by comparing the MTF plot for the lens to Figs. 4.3 and 4.4, which relate the MTF to OPD. Then multiply the OPD by the scaling factor and, again using Figs 4.3 and 4.4, determine the MTF corresponding to these scaled OPD values. Obviously the accuracy of this procedure depends on how well the simple relationships of Figs. 4.3 and 4.4 represent the usually complex mix of aberrations in a real lens.

In the event that a proposed change of aperture or field is expected to produce a change in the amount of the aberrations, one can attempt to scale the MTF as affected by aberration. This is done by determining the type of aberration which most severely limits the MTF, then scaling the OPD according to the way that this aberration scales with aperture or field, in a manner analogous to that described in Sec. 5.3. In general, OPD as a function of aperture varies as one higher exponent of the aperture than does the corresponding transverse aberration. For example, the OPD for third-order transverse spherical (which varies as Y^3) varies as the fourth power of the ray height. In a form analogous to Eqs. 5.3 and 5.4, which indicate a power series expansion of the transverse aberrations as a function of aperture and

field, Eq. 5.5 gives the relationship for OPD. As in Sec. 5.3, the terms of the equation refer to Fig. 5.4.

$$\begin{aligned}\text{OPD} = {} & A'_1 s^2 + A'_2 sh\cos\theta + B'_1 s^4 + B'_2 s^3 h\cos\theta \\ & + B'_3 s^2 h^2\cos^2\theta + B'_4 s^2 h^2 + B'_5 s h^3\cos\theta + C'_1 s^6 + C'_2 s^5 h\cos\theta \\ & + C'_4 s^4 h^2 + C'_5 s^4 h^2\cos^2\theta + C'_7 s^3 h^3\cos\theta + C'_8 s^3 h^3\cos^3\theta \\ & + C'_{10} s^2 h^4 + C'_{11} s^2 h^4\cos^2\theta + C'_{12} s h^5\cos\theta + D'_1 s^8 + \cdots \end{aligned} \quad (5.5)$$

Note that although the constants here correspond to those in Eqs. 5.3 and 5.4, they are not numerically the same. However, the expressions are related by

$$y' = \text{TA}_y = \frac{l}{N}\frac{\partial \text{OPD}}{\partial y} \quad \text{and} \quad z' = \text{TA}_z = \frac{l}{N}\frac{\partial \text{OPD}}{\partial z} \quad (5.6)$$

where l is the pupil-to-image distance and N is the image space index. Note that the exponent of the semiaperture term s is larger by 1 in the wavefront expression than in the ray-intercept equations.

5.5 Notes on the Interpretation of Ray Intercept Plots

When the image plane intersection heights of a fan of meridional rays are plotted against the slope of the rays as they emerge from the lens, the resultant curve is called a ray intercept curve, an $H' - \tan U'$ curve, or sometimes (erroneously) a rim ray curve. The shape of the intercept curve not only indicates the amount of spreading or blurring of the image directly, but also can serve to indicate which aberrations are present.

In Fig. 5.5 an oblique fan of rays from a distant object point is brought to a perfect focus at point P. If the reference plane passes through P, it is apparent that the $H' - \tan U'$ curve will be a straight horizontal line. However, if the reference plane is behind P (as shown) then the ray intercept curve becomes a tilted straight line since the height, H', decreases as $\tan U'$ decreases. Thus it is apparent that shifting the reference plane (or focusing the system) is equivalent to a rotation of the $H' - \tan U'$ coordinates. A valuable feature of this type of aberration representation is that one can immediately assess the effects of refocusing the optical system by a simple rotation of the abscissa of the figure. Notice that the slope of the line $(\Delta H'/\Delta \tan U')$ is equal to the distance δ from the reference plane to the point of focus, so that for an oblique ray fan the tangential field curvature is equal to the slope of the ray intercept curve.

Figure 5.6 shows a number of intercept curves, each labeled with

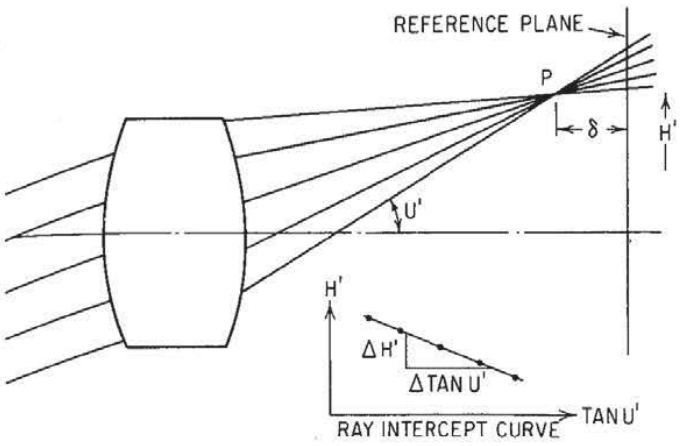


Figure 5.5 The ray intercept curve (H' versus $-\tan U'$) of an image point which does not lie in the reference plane is a tilted straight line. The slope of the line ($dH'/d \tan U'$) is mathematically identical to the distance from the reference plane to the point P. Note that this distance is equal to X_t, the tangential field curvature (if the reference plane is the paraxial focal plane).

the aberration represented. The generation of these curves can be readily understood by sketching the ray paths for each aberration and then plotting the intersection height and slope angle for each ray as a point of the curve. Distortion is not shown in Fig. 5.6; it would be represented as a vertical displacement of the curve from the paraxial image height h'. Lateral color would be represented by curves for two colors which were vertically displaced from each other. The ray intercept curves of Fig. 5.6 are generated by tracing a fan of meridional or tangential rays from an object point and plotting their intersection heights versus their slopes. The imagery in the other meridian can be examined by tracing a fan of rays in the sagittal plane (normal to the meridional plane) and plotting their z-coordinate intersection points against their slopes in the sagittal plane (i.e., the ray slope relative to the principal ray lying in the meridional plane).

It is apparent that the ray intercept curves which are "odd" functions, that is, the curves which have a rotational or point symmetry about the origin, can be represented mathematically by an equation of the form

$$y = a + bx + cx^3 + dx^5 + \cdots$$

or

$$H' = a + b \tan U' + c \tan^3 U' + d \tan^5 U' + \cdots \qquad (5.7)$$

All the ray intercept curves for *axial* image points are of this type. Since the curve for an axial image must have $H' = 0$ when $\tan U' = 0$,

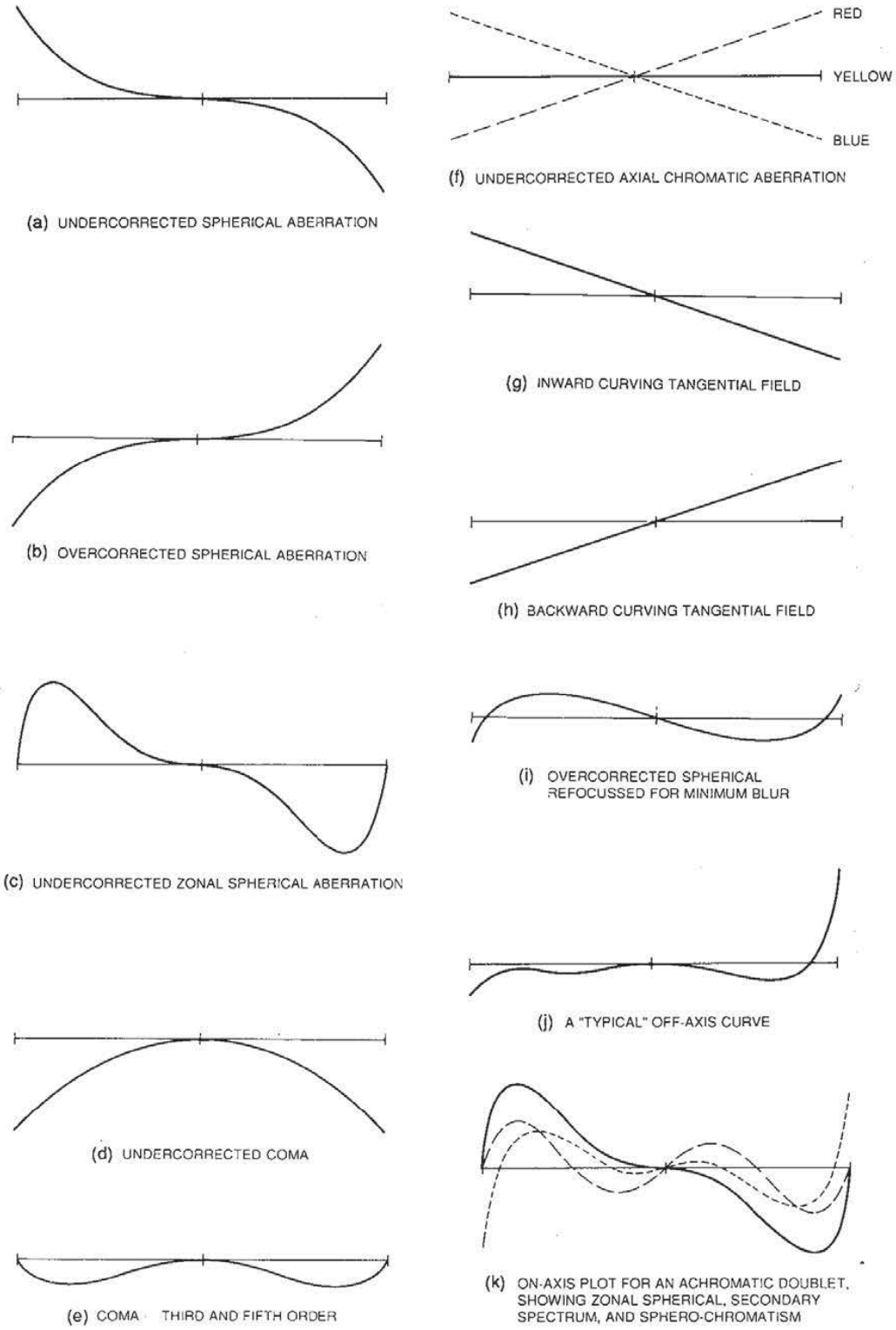


Figure 5.6 Sample ray intercept plots for various aberrations. The ordinate for each curve is the height at which the ray intersects the (paraxial) image plane; usually H is plotted relative to the principal ray height, which is set to zero. The abscissa is tan U, the final slope of the ray with respect to the optical axis. Note that, regardless of the sign convention for the ray slope, it is conventional to plot the ray through the top of the lens at the right of the figure, and that the curves for image points above the axis are usually shown. Observance of these conventions makes it much easier to interpret the plots.

it is apparent that the constant a must be a zero. It is also apparent that the constant b for this case represents the amount the reference plane is displaced from the paraxial image plane. Thus the curve for lateral spherical aberration plotted with respect to the paraxial focus can be expressed by the equation

$$TA' = c \tan^3 U' + d \tan^5 U' + e \tan^7 U' + \cdots \qquad (5.8)$$

It is, of course, possible to represent the curve by a power series expansion in terms of the final angle U', or sin U', or the ray height at the lens (Y), or even the initial slope of the ray at the object (U_0) instead of tan U'. The constants will, of course, be different for each.

For simple uncorrected lenses, the first term of Eq. 5.8 is usually adequate to describe the aberration. For the great majority of corrected lenses the first two terms are dominant; in a few cases three terms (and rarely four) are necessary to satisfactorily represent the aberration. As examples, Figs. 5.6*a* and *b* can be represented by $TA' = c \tan^3 U'$, and this type of aberration is called *third-order spherical*. Figure 5.6*c* however, would require two terms of the expansion to represent it adequately; thus $TA' = c \tan^3 U' + d \tan^5 U'$. The amount of aberration represented by the second term is called the *fifth-order aberration*. Similarly, the aberration represented by the third term of Eq. 5.8 is called the *seventh-order aberration*. The fifth-, seventh-, ninth-, etc., order aberrations are collectively referred to as *higher-order aberrations*.

The ray intercept plot is subject to a number of interesting interpretations. It is immediately apparent that the top-to-bottom extent of the plot gives the size of the image blur. Also, a rotation of the horizontal (abscissa) lines of the graph is equivalent to a refocusing of the image and can be used to determine the effect of refocusing on the size of the blur.

Figure 5.5 shows that the ray intercept plot for a defocused images is a sloping line. If we consider the slope of the curve at any point on an H–tan U ray intercept plot, the slope is equal to the defocus of a small-diameter bundle of rays centered about the ray represented by that point. In other words, this would represent the focus of the rays passing through a pinhole aperture which was so positioned as to pass the rays at that part of the H–tan U plot. Similarly, since shifting an aperture stop along the axis is, for an oblique bundle of rays, the equivalent of selecting one part or another of the ray intercept plot, one can understand why shifting the stop can change the field curvature.

The OPD (optical path difference) or wavefront aberration can be derived from an H–tan U ray intercept plot. The area under the curve between two points is equal to the OPD between the two rays which

correspond to the two points. Ordinarily, the reference ray for OPD is either the optical axis or the principal ray (for an oblique bundle). Thus the OPD for a given ray is usually the area under the ray intercept plot between the center point and the ray.

Mathematically speaking, then, the OPD is the integral of the H–tan U plot and the defocus is the first derivative. The coma is related to the curvature or second derivative of the plot, as a glance at Fig. 5.6*d* will show.

It should be apparent that a ray intercept plot for a given object point can be considered as a power series expansion of the form

$$H' = h + a + bx + cx^2 + dx^3 + ex^4 + fx^5 + \cdots \tag{5.9}$$

where h is the paraxial image height, a is the distortion, and x is the aperture variable (e.g., tan U'). Then the art of interpreting a ray intercept plot becomes analogous to decomposing the plot into its various terms. For example, cx^2 and ex^4 represent third- and fifth-order coma, while dx^3 and fx^5 are the third- and fifth-order spherical. The bx term is due to a defocusing from the paraxial focus and could be due to curvature of field. Note that the constants a, b, c, etc., will be different for points of differing distances from the axis.

Chapter

10

Telephoto Lenses

10.1 The Basic Telephoto

The arrangement shown in Fig. 10.1, with a positive component followed by a negative component, can produce a compact system with an effective focal length F which is longer than the overall length L of the lens. The ratio of L/F is called the *telephoto ratio,* and a lens for which this ratio is less than unity is classified as a telephoto lens. The smaller the ratio, the more difficult the lens is to design. Note that many camera lenses which are sold as telephoto lenses are simply long-focal-length lenses and are not true telephotos.

Many of the comments in Chap. 9 regarding retrofocus or reverse telephoto lenses are equally applicable to the telephoto lens. Equations 9.1 and 9.2 may also be applied to the telephoto. The usual Petzval problem is with a backward-curving field, just as with the retrofocus, and the same glass choices are appropriate for the telephoto. Since the system is unsymmetrical, each component must be individually achromatized if both axial and lateral color are to be cor-

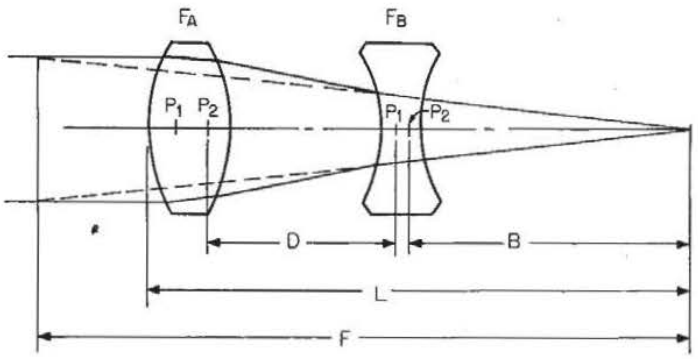


Figure 10.1 The basic power arrangement for a telephoto lens yields a compact lens with an overall length which is less than its effective focal length.

rected. The aperture stop is usually at the front member or part way toward the rear. Since a telephoto lens usually covers only a relatively small angular field, coma, distortion, and lateral color (which in many lenses are reduced by an approximate symmetry about the stop) are not as troublesome as they would be with a wider field.

10.2 Close-up or Macro lenses

The correction of a long-focal-length unsymmetrical lens is usually quite sensitive to a change in object distance, and, for most telephoto lenses, the image quality deteriorates severely when they are focused on nearby objects. Note that this effect varies inversely with the object distance expressed in focal length units; i.e., for a given design type, the image quality may remain acceptable as long as the object distance exceeds some number of focal lengths. Thus, for a given object distance, this effect is more of a problem for a long-focal-length lens than for a short. Since retrofocus lenses tend to have short focal lengths, this problem is somewhat less frequently encountered, in spite of their asymmetry.

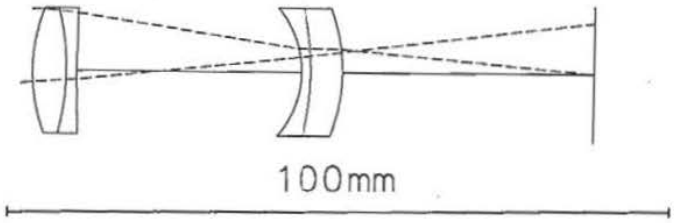
Many newer telephoto lenses and the specialized close-focusing lenses (called *macro* lenses) utilize a floating component or separately moving elements to maintain the aberration correction when the lens is focused at a close distance. For many lenses, the spherical aberration and the astigmatism become undercorrected at close conjugates. Thus a relative motion of the elements to increase the marginal ray height on a negative (or overcorrecting) element/component can be used to stabilize the spherical. The astigmatism can be controlled by a motion which increases the height of the chief ray on a component which contributes overcorrected astigmatism, or which reduces it on an undercorrecting one.

The design of such a system is carried out just like the design of a zoom lens. Two (or more) configurations are set up, one with a long (perhaps infinite) object conjugate distance and the other with a short one. The computer then uses the same lens elements with different spacings for each configuration and optimizes the merit function for both configurations simultaneously.

Figures 19.3 to 19.6, and 20.5 show nontelephoto designs with macro features.

10.3 Sample Telephoto Designs

Figures 10.2 and 10.3 show two very basic telephoto lenses; each consists of just two cemented or closely airspaced achromatic doublets, about as simple a construction as possible. Figure 10.2 covers less



KINGSLAKE TELEPHOTO MODIFIED BY HOPKINS

radius	thickness	mat'l	index	V-no	sa
24.607	5.080	BK7	1.517	64.2	9.2
-36.347	1.600	F2	1.620	36.4	9.2
212.138	12.300	air			9.0
	21.699	air			6.7
-14.123	1.520	BK7	1.517	64.2	9.4
-38.904	4.800	F2	1.620	36.4	9.4
-25.814	37.934	air			9.4

EFL = 101.6
BFL = 37.93
NA = -0.0893 (F/5.6)
GIH = 7.44
PTZ/F = -19.38
VL = 47.00
OD infinite conjugate

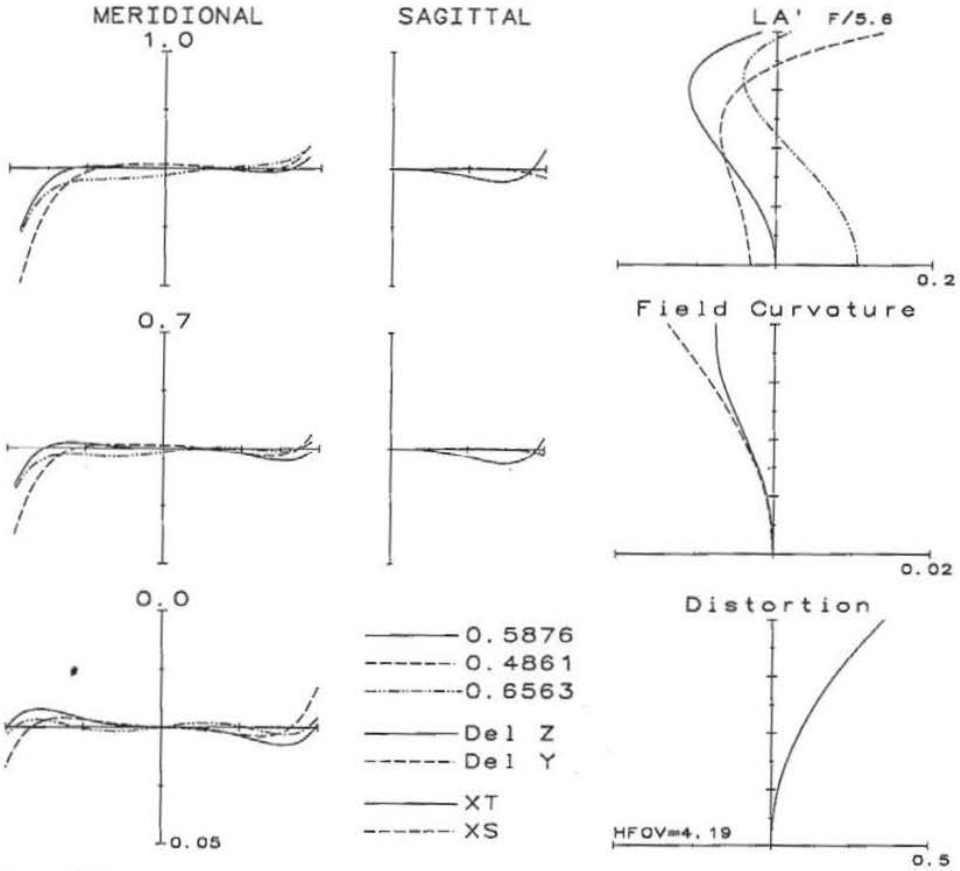
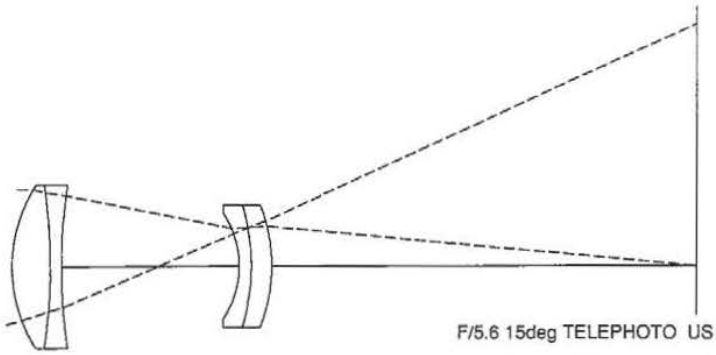


Figure 10.2



50mm

F/5.6 15deg TELEPHOTO US 2,390,387

radius	thickness	mat'l	index	V-no	sa
17.000	4.500	BK7	1.517	64.2	9.3
-62.570	0.006	air			9.3
-61.860	1.300	SF2	1.648	33.8	9.3
60.450	11.450	air			9.3
	9.540	air			6.1
-10.610	1.400	KF9	1.523	51.5	6.0
-24.760	0.011	air			7.0
-24.250	2.400	SF1	1.717	29.5	7.0
-18.570	50.137	air			7.0

EFL = 98.52
BFL = 50.14
NA = -0.0888 (F/5.6)
GIH = 26.60 (HFOV=15.11)
PTZ/F = 16.55
VL = 30.61
OD infinite conjugate

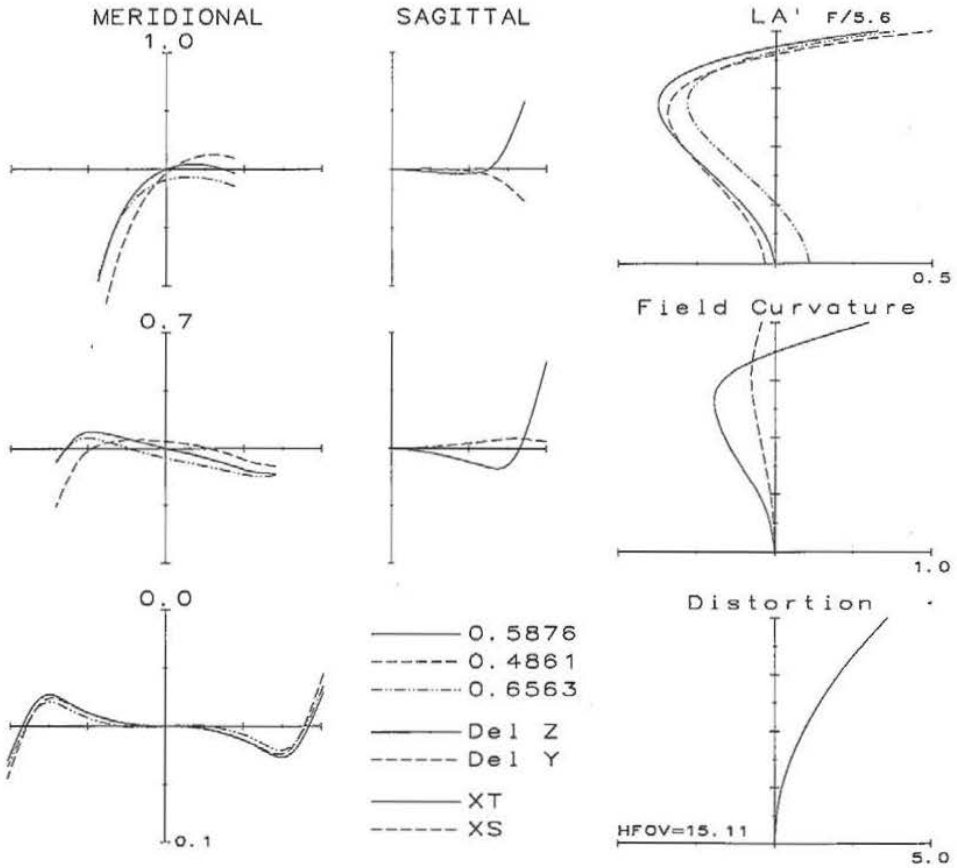


Figure 10.3

than 9° at *f*/5.6 and has a telephoto ratio of 0.85; it uses BK7 and F2 glasses and was done as an illustrative design exercise. Figure 10.3, at the same speed, covers 3 times as large a field with a telephoto ratio of 0.81. It utilizes heavier flint glasses (higher index and lower V value) to achieve a modestly improved performance.

Figure 10.4 covers a 30° field at *f*/4.5 with excellent distortion correction, illustrating the benefits derived from the added degrees of freedom gained by splitting the cemented doublets into widely airspaced components. The large telephoto ratio of 0.91 and the modestly high-index glasses are also helpful.

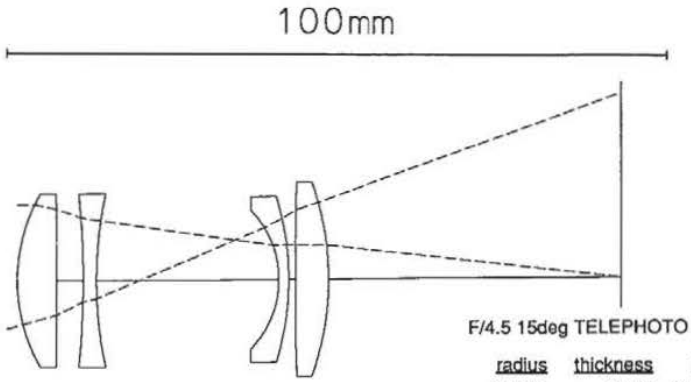
Figure 10.5 illustrates the use of unusual partial dispersion glasses (as described in Chap. 6) to reduce the secondary spectrum. The term *superachromat* implies that at least four wavelengths are brought to a common focus, whereas the term *apochromat* indicates that three wavelengths are corrected. Notice, however, that the spherochromatism and zonal spherical aberration in this lens are much larger than the axial chromatic aberration; these are the aberrations which will determine the limiting performance of this lens.

Figures 10.6, 10.7, and 10.8 each have five elements and illustrate some of the different ways that the inherent capabilities of this configuration can be utilized. In Figs. 10.6 and 10.8, the crown element of the front doublet is split into two elements to reduce the zonal spherical aberration (among others). Figure 10.6 is the result of a classroom exercise which specified a 200-mm *f*/5.0 lens with a telephoto ratio of 0.80 for a 35-mm camera. It uses quite ordinary glasses and achieves an excellent level of performance. Figure 10.7 uses high-index glass and a different arrangement to get to a speed of *f*/4.0, but falls a bit short in performance and telephoto ratio (at 0.91). Figure 10.8 uses unusual partial dispersion glasses and breaks the contact in the front doublet, to achieve what is (potentially) a high level of correction, although the telephoto ratio is only a modest 0.95.

Figure 10.9 uses seven elements to produce a well-corrected *f*/5.6, 6° field lens with an extremely short telephoto ratio of 0.66. Notice the overcorrected Petzval field, with $\rho/f = +2.1$; this is one reason that small telephoto ratios are troublesome.

With a telephoto ratio of 1.06, Fig. 10.10 doesn't really qualify as a true telephoto, but at a speed of *f*/1.8 and a field of 18°, it is an interesting lens, even if it is difficult to classify.

Figures 10.11 and 10.12 show an internal-focusing telephoto with a modest ratio of 0.92. The front component is fixed and the lens is focused for close-ups by moving the rear component toward the image plane. This could be considered as a sort of macro-style lens.



F/4.5 15deg TELEPHOTO JOSA 1950 KAZAMAKI & KONDO

radius	thickness	mat'l	index	V-no	sa
24.840	6.000	SSKN5	1.658	50.9	12.7
-671.820	4.000	air			12.7
-123.500	2.000	SF6	1.805	25.4	12.7
58.170	6.530	air			12.7
	21.000	air			7.9
-14.290	1.500	FK3	1.465	65.8	10.2
-39.560	1.000	air			12.2
500.000	5.000	BAF10	1.670	47.1	14.3
-41.470	44.437	air			14.3

EFL = 100
BFL = 44.44
NA = -0.1111 (F/4.5)
GIH = 27.00 (HFOV=15.11)
PTZ/F = -6.285
VL = 47.03
OD infinite conjugate

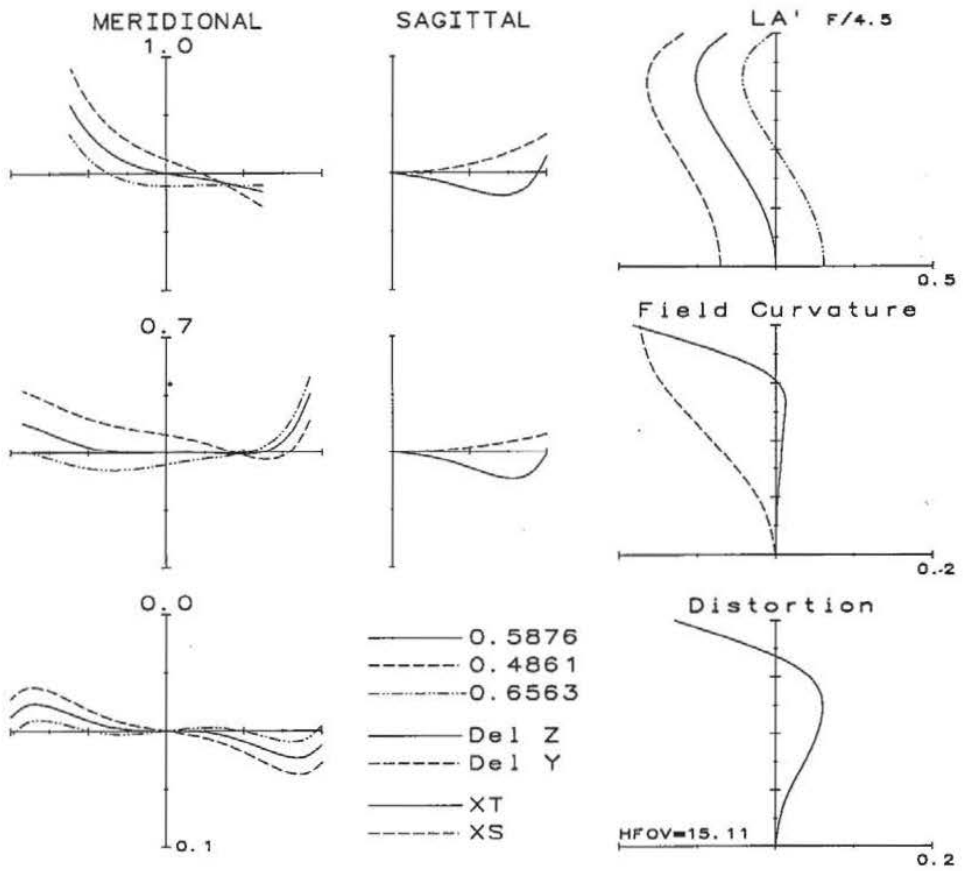
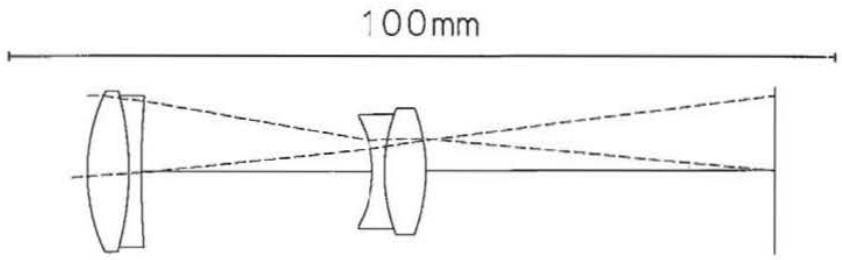


Figure 10.4



SIGLER; SUPER ACHROMAT; TELEPHOTO EFL=254

radius	thickness	mat'l	index	V-no	sa
21.851	5.008	PK51	1.529	77.0	9.5
-34.546	1.502	KZFS9	1.599	46.9	8.9
108.705	1.127	air			8.3
	26.965	air			8.1
-12.852	1.502	KZFS1	1.613	44.3	6.3
19.813	5.008	BASF5	1.603	42.5	6.7
-20.378	42.174	air			7.4

EFL = 100
BFL = 42.17
NA = -0.0898 (F/5.6)
GIH = 8.75
PTZ/F = -40.38
VL = 41.11
OD infinite conjugate

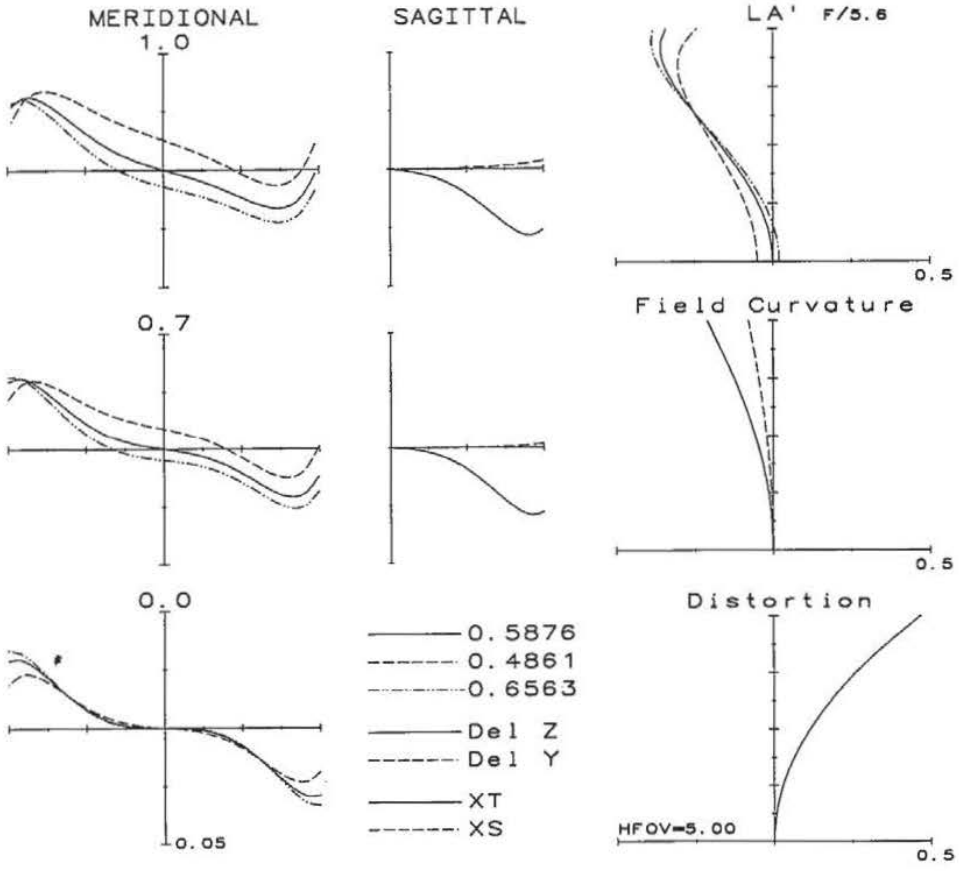
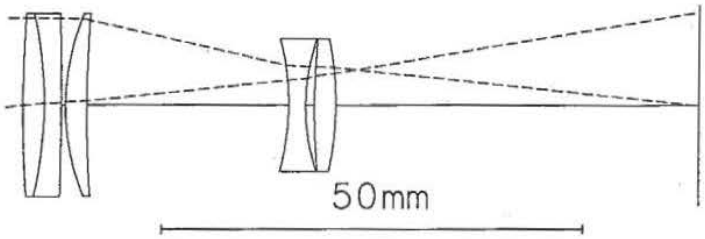


Figure 10.5



F/5 6deg TELEPHOTO

radius	thickness	mat'l	index	V-no	sa
149.035	2.500	SK4	1.613	58.6	10.5
-46.003	2.000	SF14	1.762	26.5	10.5
-477.921	0.500	air			10.5
26.522	2.500	SK4	1.613	58.6	10.5
132.322	24.060	air			10.5
-28.605	2.000	SK4	1.613	58.6	7.6
22.989	1.050	air			7.6
82.834	2.500	F5	1.603	38.0	7.6
-36.911	42.897	air			7.6

EFL = 100
BFL = 42.9
NA = -0.1000 (F/5.0)
GIH = 10.50 (HFOV=5.99)
PTZ/F = 7.68
VL = 37.11
OD infinite conjugate

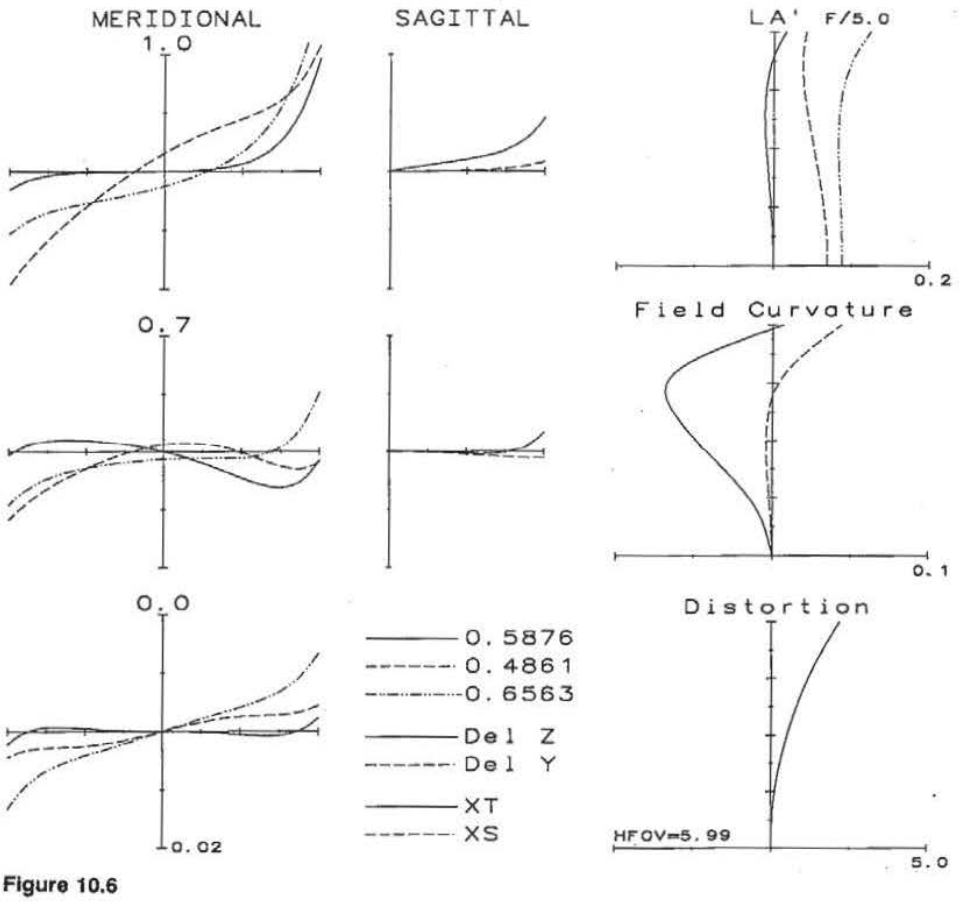
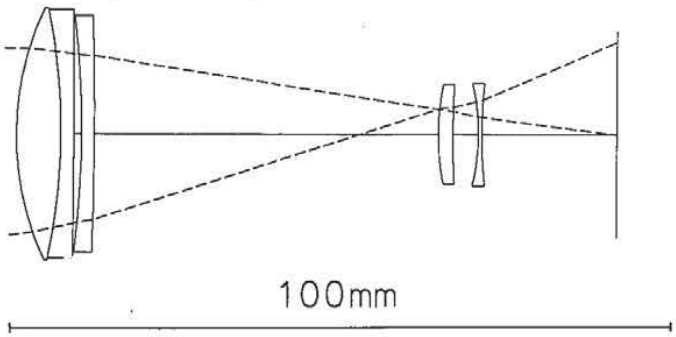


Figure 10.6



J. EGGERT ET AL; USP 3388956; F/4 15 DEG. TELEPHOTO #1

radius	thickness	mat'l	index	V-no	sa
42.156	6.676	LAKN6	1.642	58.0	18.4
-88.214	1.945	SF10	1.728	28.4	18.2
-1800.073	1.183	air			17.4
-137.288	1.945	SF9	1.654	33.7	17.3
-530.649	40.290	air			16.8
	11.913	air			5.4
34.202	2.234	LAF11	1.757	31.7	7.3
117.496	3.859	air			7.3
-31.069	0.526	BSF10	1.650	39.1	7.3
69.782	20.320	air			7.5

EFL = 99.99
BFL = 20.32
NA = -0.1256 (F/4.0)
GIH = 13.16 (HFOV=7.50)
PTZ/F = 4.272
VL = 70.57
OD infinite conjugate

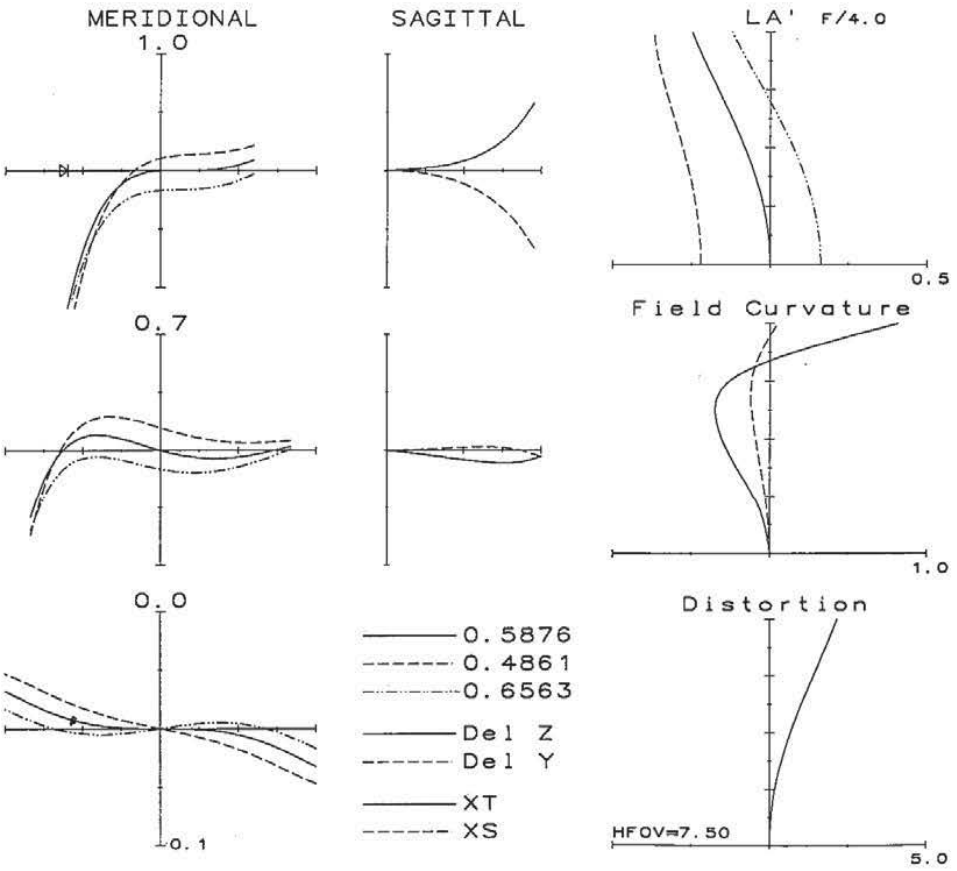
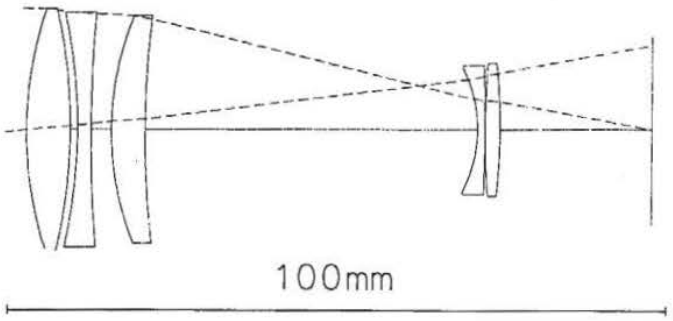


Figure 10.7



SEI MATSUI; USP 4338001; F/2.8 14 DEG. TELEPHOTO LENS #1

radius	thickness	mat'l	index	V-no	sa
54.585	6.667	FCD1	1.497	81.6	17.9
-77.813	1.111	air			17.7
-76.698	2.056	LAFN7	1.750	34.9	17.3
207.222	3.056	air			17.0
43.208	5.111	BED5	1.658	50.9	16.8
134.444	50.667	air			16.2
-19.462	1.111	K3	1.518	59.0	9.0
-305.556	0.056	air			9.5
121.887	2.222	TAF2	1.794	45.4	9.7
-89.277	22.862	air			9.8

EFL = 100
BFL = 22.86
NA = -0.1790 (F/2.8)
GIH = 12.28 (HFOV=7.00)
PTZ/F = -9.12
VL = 72.06
OD infinite conjugate

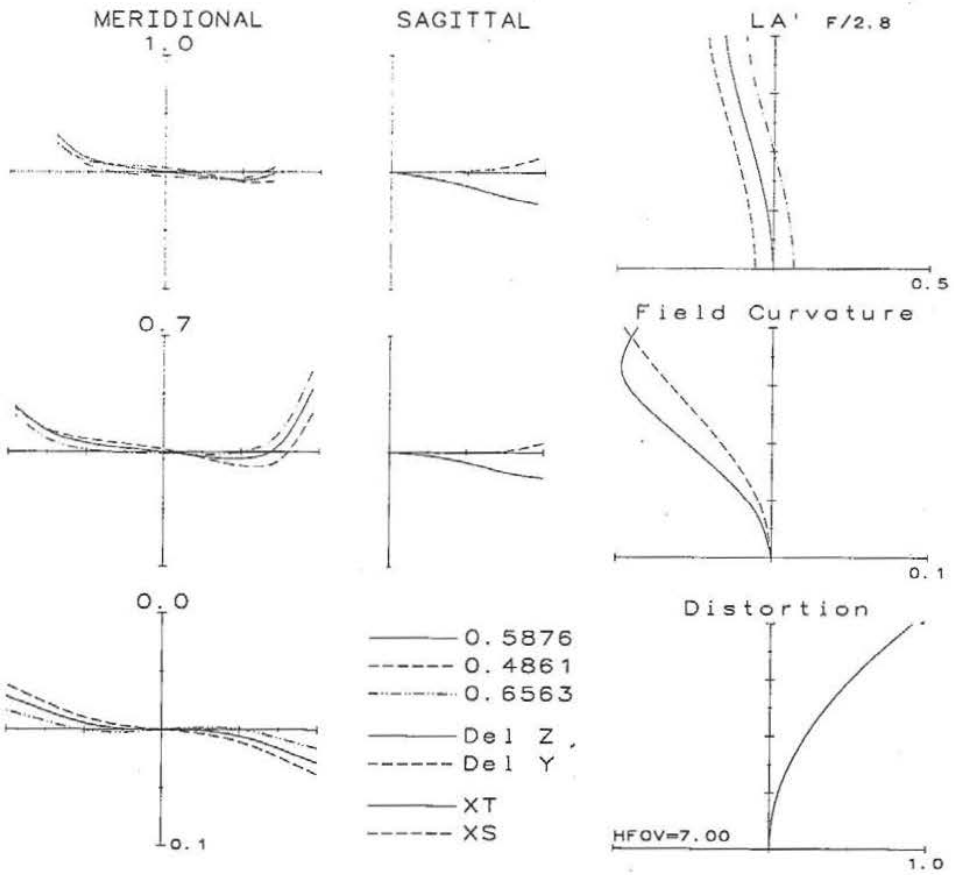
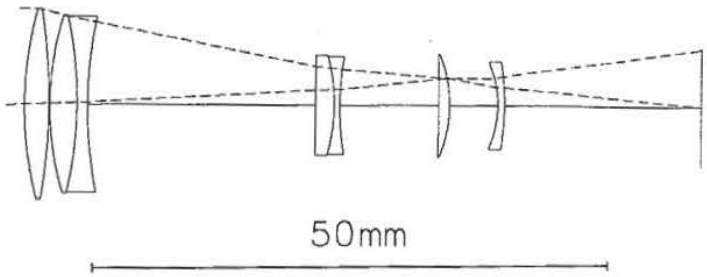


Figure 10.8



MELVYN H. KREITZER; USP 4359272; 390 MM F/5.6 6 DEG. TEL. #1

radius	thickness	mat'l	index	V-no	sa
33.072	2.386	C3	1.518	59.0	8.9
-53.387	0.077	air			8.9
27.825	2.657	C3	1.518	59.0	8.4
-35.934	1.025	LAF7	1.749	35.0	8.3
40.900	22.084	air			7.8
	1.794	FD110	1.785	25.7	4.7
-16.775	0.641	TAFD5	1.835	43.0	4.6
27.153	9.607	air			4.5
-120.757	1.035	CF6	1.517	52.2	4.8
-12.105	4.705	air			4.8
-9.386	0.641	TAF1	1.773	49.6	4.0
-24.331	18.960	air			4.1

EFL = 100
BFL = 18.96
NA = -0.0892 (F/5.6)
GIH = 5.24
PTZ/F = 2.097
VL = 46.65
OD infinite conjugate

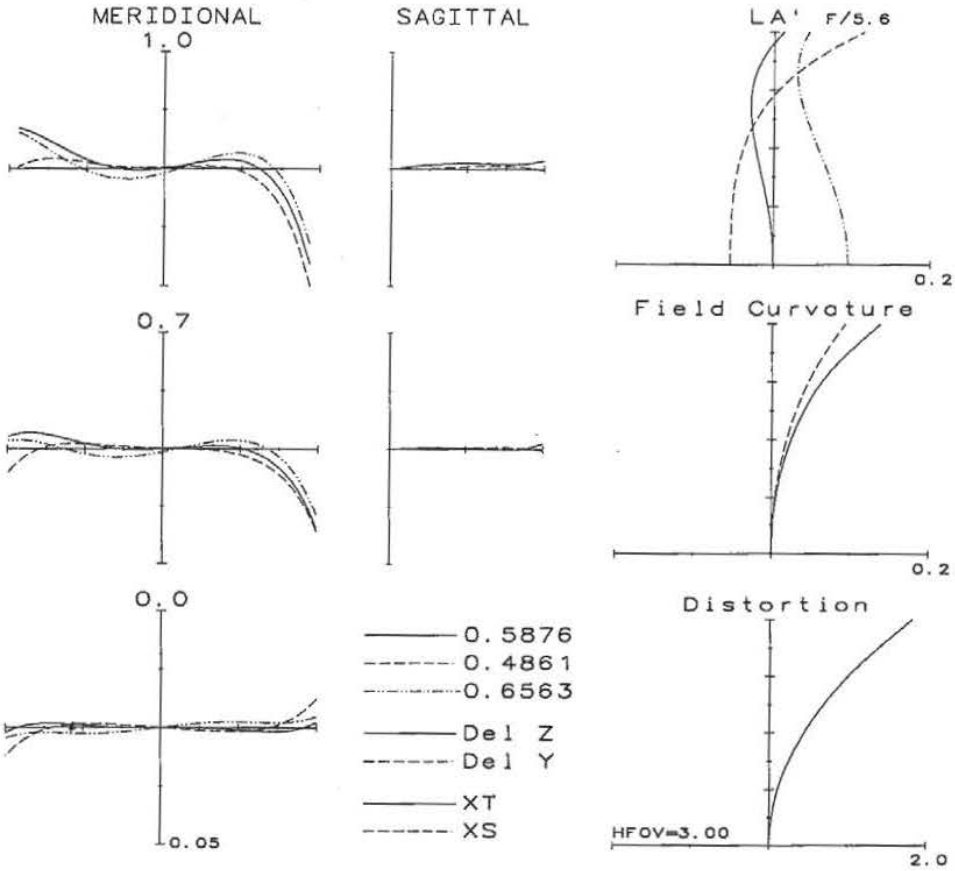
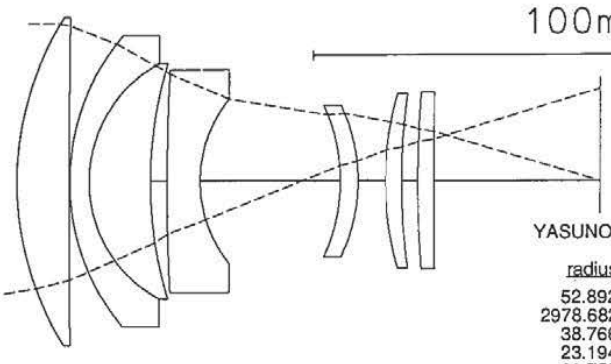


Figure 10.9



YASUNORI ARAI; USP 4447137; F/1.8 TELEPHOTO #2

radius	thickness	mat'l	index	V-no	sa
52.892	9.556	SK16	1.620	60.3	28.9
2978.682	0.074	air			28.9
38.766	3.571	BASF8	1.723	37.9	25.6
23.194	11.112	LAK14	1.697	55.5	20.7
69.755	3.000	air			19.7
341.551	6.052	SF5	1.673	32.1	19.6
22.422	18.520	air			14.5
	6.949	air			12.4
-30.641	3.237	BSF07	1.702	41.2	12.5
-29.652	5.089	air			13.3
53.327	3.000	LAF11	1.720	46.0	15.4
106.470	2.963	air			15.4
184.198	2.963	LAF01	1.700	48.1	15.6
1116.126	30.191	air			15.6

EFL = 100
BFL = 30.19
NA = -0.2779 (F/1.80)
GIH = 16.20 (HFOV=9.20)
PTZ/F = -8.83
VL = 76.09
OD infinite conjugate

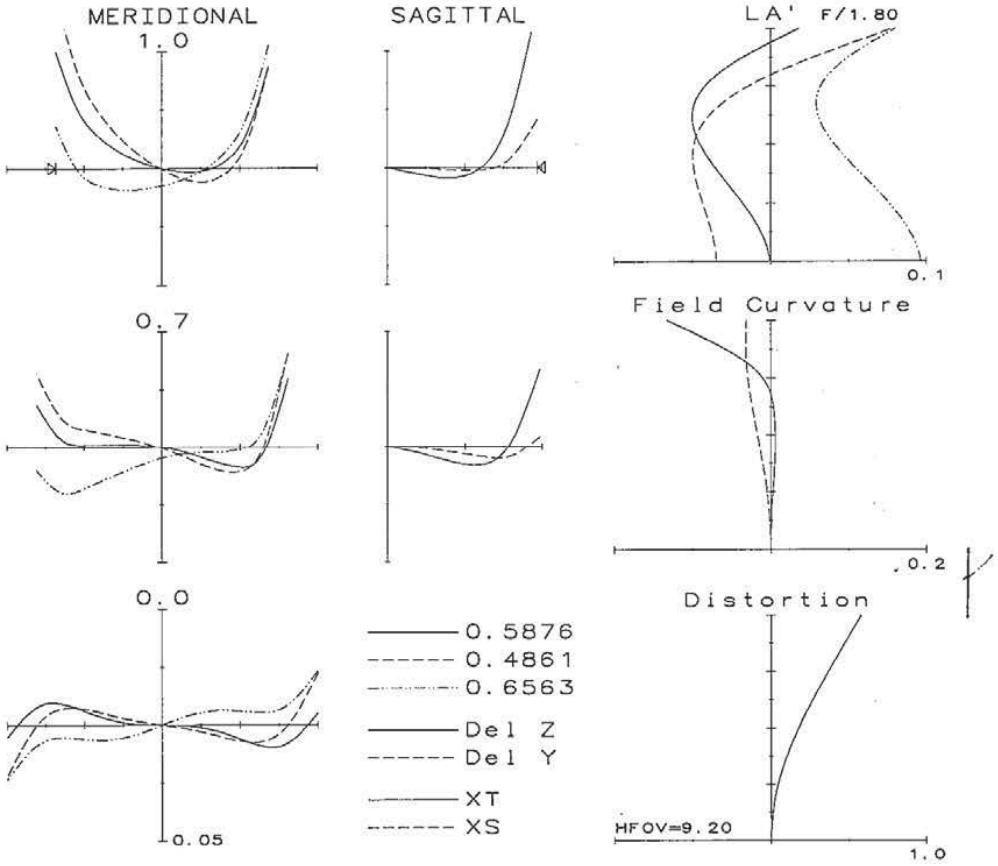
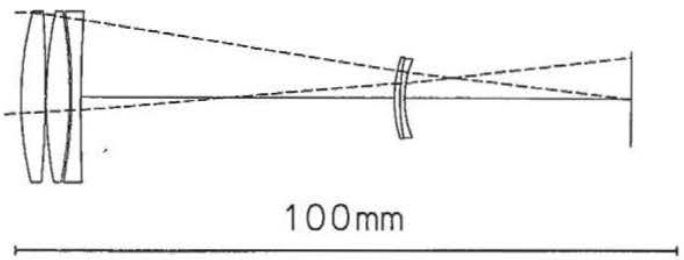


Figure 10.10



MOMIYAMA USP 4,037,935; F/4.5 (LONG EFL POSITION)

radius	thickness	mat'l	index	V-no	sa
48.796	3.645	CAF	1.434	94.9	12.6
-168.096	0.140	air			12.6
62.820	3.505	FK5	1.487	70.2	12.6
-70.733	0.303	air			12.6
-73.188	1.402	LASF3	1.806	40.9	12.6
250.187	22.431	air			12.6
	25.150	air			7.9
19.535	0.841	SF4	1.755	27.5	6.0
27.456	0.701	LSF16	1.772	49.6	5.8
14.524	34.173	air			5.7

EFL = 112.2
BFL = 34.17
NA = -0.1099 (F/4.5)
GIH = 6.06
PTZ/F = -4.07
VL = 58.12
OD Infinite conjugate

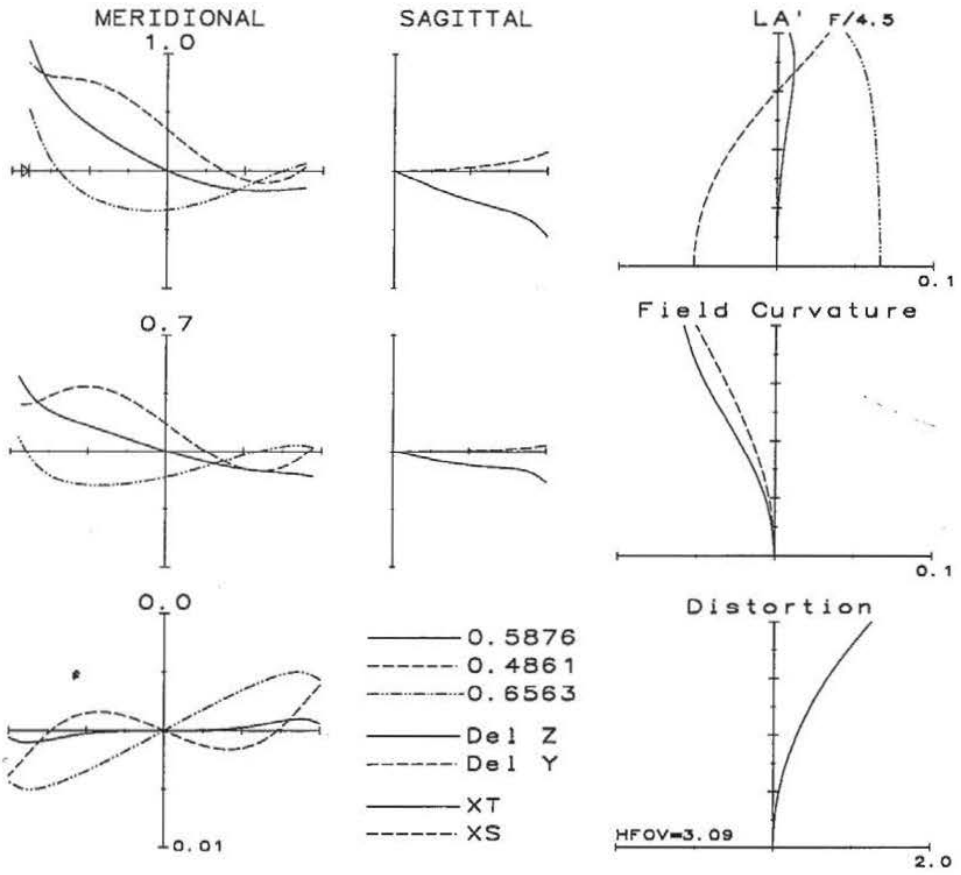


Figure 10.11

100mm

MOMIYAMA USP 4,037,935; F/4.5 (SHORT EFL POSITION)

radius	thickness	mat'l	index	V-no	sa
48.796	3.645	CAF	1.434	94.9	12.6
-168.096	0.140	air			12.6
62.820	3.505	FK5	1.487	70.2	12.6
-70.733	0.303	air			12.6
-73.188	1.402	LASF3	1.806	40.9	12.6
250.187	22.431	air			12.6
	32.440	air			7.9
19.535	0.841	SF4	1.755	27.5	6.0
27.456	0.701	LSF16	1.772	49.6	5.8
14.524	26.857	air			5.7

EFL = 100
BFL = 26.66
NA = -0.1139 (F/4.4)
GIH = 5.93
PTZ/F = -4.605
VL = 65.41
OD = 2243.07 (MAG = -0.048)

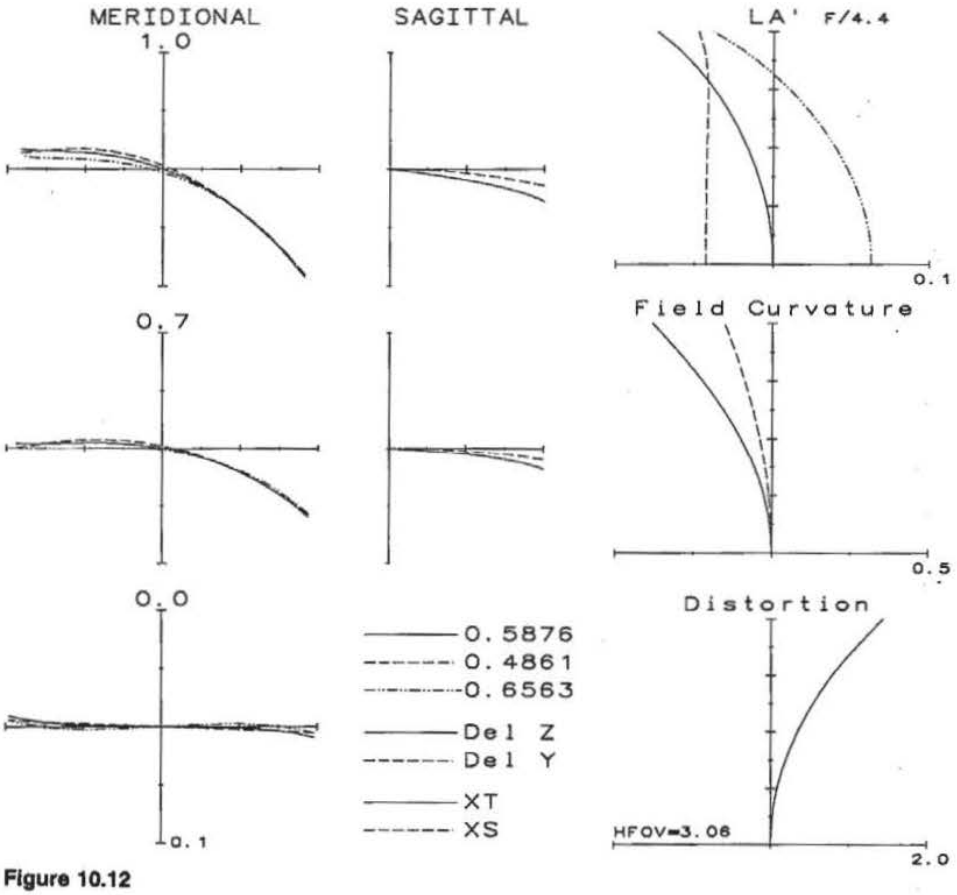


Figure 10.12