

Optical design and engineering: lessons learned

Pantazis Mouroulis*

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 91109

ABSTRACT

This paper contains some recommendations for the optics curricula that seek to serve students intent on a career in optical engineering, as well as suggestions for the young professional embarking on such a career. It also illustrates the suggestions with some examples of actual optical systems.

Keywords: Optical design, optical engineering, optics education, imaging spectrometry, grating design, achromatic lens, ray aiming.

1. INTRODUCTION

Having studied and then taught optics for a number of years, the author eventually found himself in the position of having to produce optical systems and designs for a living. There arose then an interesting opportunity for assessing how much of the theoretical knowledge was actually useful in practice, as well as what was missing from it that could have made a difference had it been taught. While one person's view and experience is undoubtedly limited, it is hoped that a few of the lessons learned will be of more general interest. Three themes are examined in this paper: 1) the incorporation of a 'systems' outlook as an inherent part of optical design and engineering, 2) the uses of optical design software, and 3) concepts found very useful or not so useful in optical design or engineering practice.

2. OPTICS IS NOT AN ISLAND

2.1 Some recommendations for optics curricula

It may come as a shock to the newly minted optical engineer to learn that optics is often a rather small part of the budget of an optical instrument and consequently receives only proportional attention. This situation is encountered in extreme form in the aerospace industry where the cost of an instrument can often be predicted by its mass and with no knowledge of its function or principle of operation; but it can be true even in commercial instrumentation. The final output of many, or even most systems is not an aerial image but a digitally stored data array; and the data quality must be preserved to the end, thus involving many disciplines outside of optics.

Accordingly, an optical engineering curriculum needs to instill a systems outlook. While it is necessary and useful to master the basics of individual areas within optics, there is usually no place in the curriculum where a complete system is analyzed. There are two distinct aspects to this outlook: 1) the interaction between optics and the other disciplines that go into making a complete system, and 2) the incorporation into the optical design of considerations that relate to the manufacturability and performance of the system in its intended environment.

Considering the second aspect, the student of optical design is often in danger of believing that a good design is one in which the aberrations have been reduced to very low levels. But this can be far from true. In the first place, a large amount of aberration may be permissible in certain systems where alternative considerations prevail. But even for aberration-limited systems, a good design is one which can be built for the money and time available, and maintain performance through its life. The designer must take into account manufacturability, cost, ease of alignment, and inherent stability (ruggedness, insensitivity to thermal and dynamic environment). Alignment is especially critical with off-axis mirror systems. A

* pantazis.mouroulis@jpl.nasa.gov, phone: (818) 393-2599, fax (818) 393-9567

beautiful design that is difficult to align is not so beautiful, unless accompanied by a feasible and not overly complicated alignment plan. The design is not complete until such plan is in place, since otherwise the design performance is not likely to be met.

A related topic is tolerancing. Though this is addressed in texts¹⁻³, it is not clear that an instructor would always find the time to address it in class. One possible way to insert a minimum appreciation is to let the students specify a cemented doublet for fabrication. This would make immediately obvious the necessity of tolerancing even simple systems, as well as teach some of the language of optical fabrication. Any other simple lens example could fulfill the same purpose.

If the emphasis is placed on the optical engineering rather than optical design aspect, then the example should be taken further and have a significant portion of a course dedicated to it. The course content would center around a notional instrument design which would then be taken to completion of at least the optical aspects: tolerancing, fabrication, alignment, assembly, and performance testing, the latter hopefully to include such matters as interferometric measurements and simulation of system errors for determining misalignments. Within this context, the concept of stray light and basic strategies for its amelioration can be at least briefly discussed. Stray light analysis includes overall system considerations naturally, as it forces one to realize that the main sources of stray light may be entirely external to the instrument that one spends so much time optimizing (for example, glint from an unrelated piece of hardware). But it also begins to bring in detector-related and system-related questions in terms of tolerable stray light amounts.

The theme of a notional instrument can also proceed with the incorporation of considerations that are beyond the narrowly defined boundaries of optics. The closest relative of optical design is opto-mechanics, where the optical engineer needs not only an appreciation of lens mounting techniques but also an understanding of the principles of minimizing thermal or mounting stresses and distortion. The principle of kinematic mounting should be discussed theoretically and demonstrated in the laboratory. In this way, stress minimization as well as the ease of precise positioning and re-positioning afforded by kinematic mounting can be immediately appreciated.

Next would come some considerations regarding systems in which the final detector is electronic, which probably form the great majority of modern systems. At least two possibilities exist, and either one can serve as a useful example from which there is a lot to learn: sampled imaging systems, or systems in which the significant aspect is the extraction of a signal from a single detector. In the first case, it is not unusual for the optical designer to produce, or be asked to produce, a diffraction-limited lens when in fact such a lens could lead to a substantially undersampled image. It is important to understand that sharper is not always better but also to go beyond the simple qualitative idea by having a basic procedure for optimizing a complete system, not just the lens in isolation.⁴

The second category of systems can be represented by spectrometers or polarimeters especially in a remote sensing context, but it can also be represented by modern scanning microscopes, laser scanners, and several other systems. These system designs can be taken through a signal-to-noise computation that incorporates the detector characteristics and various noise mechanisms. Such calculations are a natural extension of radiometry, which is more often than not a part of the curriculum at least in a minimum form.

While I recognize that many of these topics are normally taught in an optics curriculum, the difference is that they are taught separately. I believe that there is significant value in bringing them all together and showing how they all interact in order to build a successful instrument. In some sense, this approach would be providing the student with engineering experience before graduation. Overall, my suggestion is as follows: if there is a balance to be struck or a trade between teaching more detailed lens design and teaching more system design, the optical design curriculum should lean towards the latter. Experience suggests that the optical engineer who can understand systems trades and can design an overall balanced instrument will in general be more sought after than the lens designer who does not extend beyond the narrow confines of the trade. To be sure, there will always be a need for expert lens designers; but the proliferation and increasing user-friendliness of optical design software together with the vast lens databases available make it possible for many engineers to design quite acceptable lens systems for simple to moderate needs, which is the majority of cases.

2.2 The young engineer's survival kit

These recommendations for the entering optical engineer are based on practices that I have found particularly useful and effective. To this date, they have not let me down.

1) *Three words: simplify, simplify, simplify.* Resist the tendency to increase complexity if the system is not optimizing immediately. Too often the inexperienced designer will add elements or resort to aspheric or tilted surfaces in order to correct a system that could have optimized using perhaps a different merit function or starting point. And then, when the design appears satisfactory and one thinks it is finished, one should ask the question whether it could be made simpler. If a fully diffraction-limited solution has been obtained with four elements, it could mean that a satisfactory, near-diffraction limited solution could be obtained with three. A simpler design will generally be easier to align and will have a better chance of maintaining performance in the field.

2) *Design with cost in mind from the beginning.* Even if you are told that cost is not an issue, cost is still an issue. An inexpensive design will always be welcome, but also it will tend to be simpler than a more expensive one. Of course, this does not hold always, for example there may be cases where an aspheric singlet is preferable to a cheaper spherical doublet. But when other considerations are accounted for, the less expensive design is the better design.

3) *Eliminate the middleman* or at least strive to reduce the need for one. Attempt to understand the customer requirements and translate them into optical design metrics. In other words, do not start from an image quality specification derived from the customer requirements by someone else (or even by the customer), but rather from an understanding of the job that the system is required to perform – then derive an optical specification from that. I call this end-user oriented optical design.⁵

4) *Forget "first-cut" designs* even if you are asked to produce one. Insert complete system considerations into the design from the beginning. Resist the temptation to produce designs that need a lot of work before they satisfy the requirements. Take a broad view of system constraints and include them all into the first design attempt. For example, a system that satisfies the image quality requirements but is substantially larger than the available space is not a good first solution. In fact, it is not a solution at all, and it does not necessarily show the way to a design that would satisfy all constraints.

5) *Question "impossible" specifications.* Probe relentlessly into the customer's "impossible" requirements. At least half of the time the customer has no good reason, and a simpler/different solution is possible or even optimum. Of course, seemingly impossible problems are the most rewarding to solve, but one should attempt the impossible only for a good reason. If a simple solution suffices, it will become obvious with hindsight; and the complicated system that solved the non-existent problem will be to no one's credit.

6) *Think on the customer's behalf,* and try to solve the problems that s/he does not know s/he has. Do not be satisfied with merely meeting the minimum specifications. If those are met and the system does not fulfill its intended function nobody will be satisfied. The reduction of measurement requirements to a set of specifications for the design is a very demanding problem, so it is not always safe to assume that it has been performed correctly by others.

7) *Try to keep up with developments by reading the current literature.* Optical engineers, perhaps no more so than any other engineers, tend to think that engineering is mostly a matter of employing sound practices while only the researchers need to keep up with the research literature. And yet, while the fraction of groundbreaking papers is probably decreasing in proportion to their increasing specialization, there is still a lot to learn that can make the difference when a clever solution is called upon. Often the difference between an average engineer or designer and a very good one is the speed with which the latter can come up with a solution to a difficult problem. And with the exception of the inventive geniuses, that speed is proportional to the store of relevant knowledge.

2.3 Example: Some considerations in the design of pushbroom imaging spectrometers

Some of the above points can be illustrated with the Offner spectrometer design,⁶ which has become the preferred solution for many pushbroom imaging spectrometer systems. The insight of replacing the convex secondary with a diffraction grating has produced a set of spectrometer designs that can satisfy a wide variety of applications. The original Offner relay⁷ (Fig. 1) is a marvel of simplicity: two spherical, concentric mirrors with symmetry about the stop. Diffraction from the grating breaks that symmetry and therefore creates some complication. It is often found advantageous to split the large concave mirror into two separate ones; it has been shown that a size reduction of as much as 50% can be achieved with a three-mirror design over a two-mirror version with comparable performance.⁸ This results in a system as shown in Fig. 1(b). Now, in optimizing this design, the designer will normally allow all curvatures and separations as variables, thus retaining at most only the common axis of rotation from the original Offner relay symmetry. Yet, if one thinks ahead to alignment, it becomes evident that retaining the concentricity of the two concave mirrors is a great advantage, as it allows them to be co-aligned in front of an interferometer with great accuracy and speed, thereby leaving only grating positioning as a variable, and removing all degrees of freedom associated with the concave mirrors. Demanding concentricity of the two concave mirrors turns out to have a negligible impact on design performance, as the optimum solution is close to the concentric condition anyway. However, interferometric positioning reduces tolerances greatly and allows a better overall result. This procedure has been successfully demonstrated,⁹ yet to this date most Offner designs appear not to have taken advantage of it; the result is substandard performance. The author has seen designs that reduce the symmetry even further by making use of tilts and decenters or even aspheric surfaces for non-stressing specifications that do not justify such extreme measures. These designs have probably arisen from the tendency of the designer to increase the number of variables as soon as the optimization appears to have stalled, when in fact the correct approach would have been to take a second look at the merit function and the optimization technique in order to determine the cause of the stagnation.

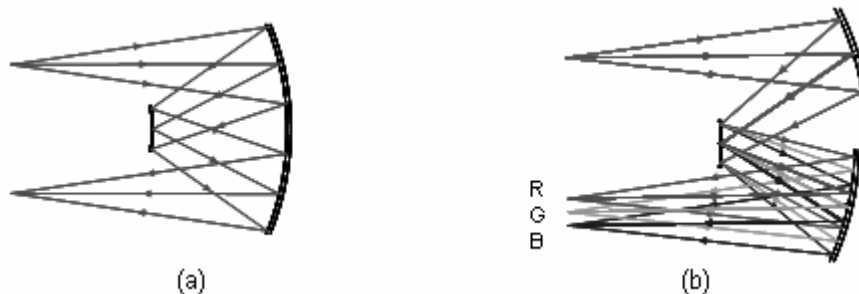


Figure 1 (a): The original Offner relay. (b): an Offner spectrometer with different primary and tertiary.

The above discussion demonstrates some of the benefits of simplicity and cost-consciousness. Let us now consider how complete system considerations must be inserted in the design process. The Offner design is capable of covering a wide spectral range with appropriate order-sorting filters. This has led to a demand for grating designs that are efficient over a broad band, for example, 400-2500 nm. Such broadband efficiency can be achieved by modifying the grating groove shape to depart from the ideal blazed sawtooth profile^{10,11} or by splitting the grating into areas with different blaze angles¹¹ in conjunction perhaps with utilizing more than one diffracted order. The second technique is the more mature of the two, and has been employed in the CRISM imaging spectrometer which is due to launch towards Mars orbit in 2005.¹² A new instrument, the Moon Mineralogy Mapper (M^3) is currently under development for a two-year Moon-orbiting mission. This instrument intends to cover a very broad band (400-3000 nm with 10 nm sampling) using a single grating in first order. As part of the optical design, the grating blaze angles and relative areas have to be optimized and then specified for fabrication.

With lithographic grating fabrication techniques, the grating efficiency function can be tailored to requirements at least to some extent, and thus affords a degree of freedom to the designer that can be used to balance the system response (signal-to-noise ratio or SNR) through wavelength. An ideally balanced

system has a flat SNR over the entire spectral band. This is very difficult to accomplish as the solar blackbody spectrum is heavily weighted towards the short wavelengths. In any case, the grating response must be high in the long wavelength end where source photons are scarce and where thermal background noise begins to intrude.

By separating the grating into three zones with different blaze angles, a composite response such as the one shown in Fig. 2 can be obtained. This response is the weighted sum of the individual blaze responses also shown in the figure. The question that then arises is how to specify the shape of the zones. This is an important question that is usually either forgotten or not addressed with sufficient care in these designs. It is generally advantageous to retain the symmetry, so we may assume that the zones will be concentric rings. But which ring should go on the outside, inside, or middle?

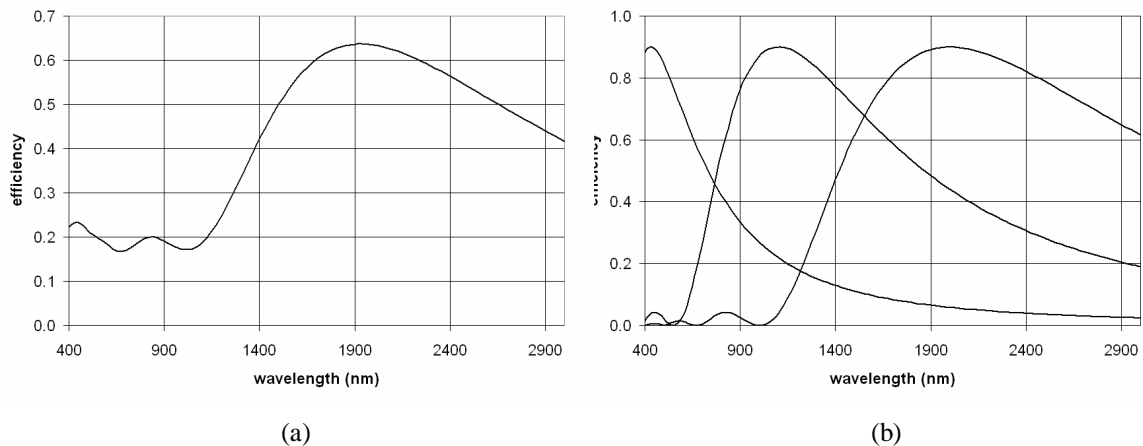


Figure 2. (a): Composite triple-blaze grating broadband grating efficiency response, made to emphasize long wavelengths. (b): Individual blaze responses. These are multiplied by the corresponding area they occupy on the grating surface, and summed to result in the composite response (a). Notice the secondary lobes to the left of the main peaks occurring at around 850 nm and 450 nm.

The importance of the question can be understood by inspection of Fig. 2, which shows that the efficiency of certain zones will drop to zero at some wavelengths. This produces a heavily apodized beam, which can have a significant effect on the PSF. Further, there can be a phase shift between the two zones. Even if the grating is perfectly balanced and fabricated with no phase jumps across the boundaries, the secondary minor lobe to the left of the main one in every efficiency curve is inherently phase-shifted by π . A model of the spectrometer PSF would be severely lacking if it were to ignore these effects, and yet they are most often ignored and the grating is treated as if it had uniform efficiency and constant phase across its face.

The effect of the apodization as well as the phase shift can be seen in Fig. 3, which shows the PSF and the ensquared energy for a wavelength of 850 nm, at which point the long wavelength blaze displays a phase shift of π (Fig. 2(b)). It can be seen that the effect of the phase shift is to completely suppress the central maximum and send a lot of energy into the rings, which are already strong due to the lack of any significant intensity in the middle of the aperture. The corresponding ensquared energy is also shown in Fig. 3(b), and represents the worst case for all wavelengths and fields. This PSF is obtained for an arrangement that puts the long wavelength blaze in the middle ring. Alternative arrangements have similarly low ensquared energy although at different wavelength bands.

The ensquared energy of Fig. 3 is not so low as to make the spectrometer useless; indeed, many spectrometers operate at even lower values than this. The detrimental effect only becomes apparent if one considers the effect of the PSF on the system characteristics, and specifically the spatial uniformity of the response. At other wavelengths, where the grating apodization is less pronounced and there are no phase shifts, the ensquared energy approaches the (unapodized) diffraction limit of 90%; and it is the difference

between that high value and the lower (~65%) value of Fig. 3 that causes a problem. If the various wavelengths have very different PSFs (or spatial response functions) then the signals arriving at the detector do not arise from the same spot on the ground, thus violating a fundamental assumption of spectroscopy. The ultimate effect is to reduce the spatial resolution of the spectrometer system by impairing its ability to discriminate between spectra in images with high amounts of spatial detail. Stated differently, the spectrometer produces non-physical spectra at all those pixel boundaries separating two different materials. For a high-performance instrument, where such variation of the spatial response function is required to stay below 10% of a pixel width, the PSF of Fig. 3 is not acceptable if at other wavelengths the response is much better. Indeed, for the M³ instrument a different solution had to be sought. In any case, the example illustrates how the designer must understand the system requirements in great detail so as to provide a proper solution, as it would be unrealistic to expect the customer to specify a grating in such detail.

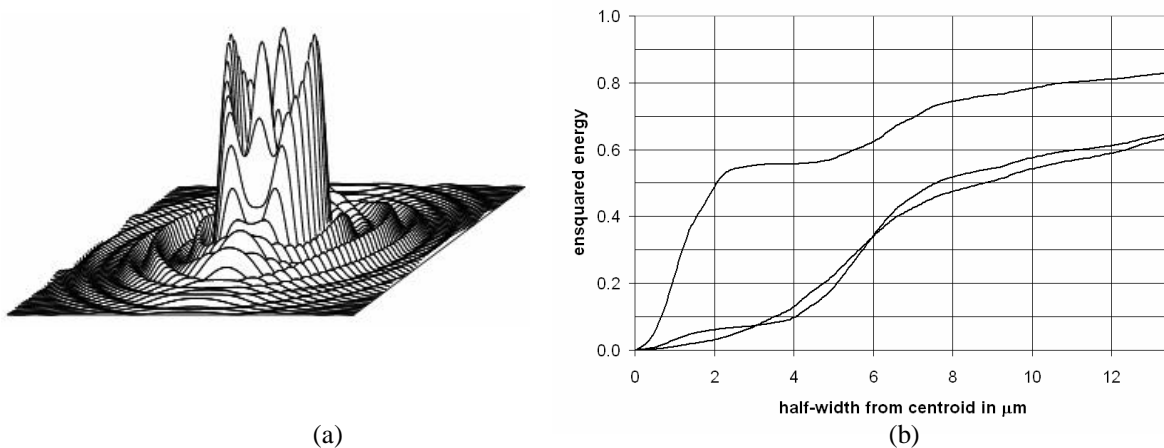


Figure 3. Left (a): Point Spread Function corresponding to the wavelength of 850 nm, using the grating efficiency curves of Fig. 2. The size of the window equals a 2x2 detector pixel area. Right (b): Ensquared energy inside a 27 μm pixel for the PSF shown in (a). The top curve shows the ensquared energy for a similarly apodized aperture but without the pi phase shift. The two lower curves include the pi phase shift and represent different field points with different aberrations showing that the residual aberration is entirely swamped by the apodization and phase shift effects.

3. THE ELEPHANT IN THE ROOM: OPTICAL DESIGN SOFTWARE AND ITS USES

There are two aspects of interest with regard to optical design software: its use by a professional, and its use as a teaching tool.

With respect to the first aspect, a famous optical designer and engineer is reputed to have said that designers should keep their hands away from the keyboard until a satisfactory first and third order solution has been found (I am omitting the name, as I cannot find the original quote). I have always thought of this as good advice and repeated it often in word and deed. However, the evolution of software and the proliferation of lens design databases forces us to see the previous advice in a new light. It must now be seen as a recommendation for a solid understanding of the theory, rather than as describing the steps of the design process. The designer must seek the most efficient way of performing a given task, and for me, the most efficient way is usually to solve the theoretical problem up to the point where I understand that there is a solution, and shortly after that, let the computer find that solution. This does not mean blindly submitting a poor design to optimization, but rather performing explorations of the design space by varying parameters and asking “what if” questions bounded by a general understanding of the problem. Compared to solving equations by hand, such questions can be answered with lightning speed by the computer, and in any case the problem is not always amenable to closed-form solution. I make this statement with considerable regret, since the reduction of a physical problem to a set of equations and their subsequent solution is one of the most satisfying activities I can think of. Nonetheless, I cannot see how I could have responded to the workplace demands for instant designs without letting the computer do most of the work

while I tended to those activities that, regrettably, the computer could not handle such as management meetings. Clearly, one must be armed with the knowledge of certain existence theorems (e.g., correction at multiple conjugates, stop-shift theorems) and a thorough knowledge of first-order layout so the computer is not asked to perform impossible tasks. But there is no question that a significant change of attitude took place by necessity, from seeing the computer as a necessary evil to seeing it as an irreplaceable productivity tool.

3.1 Example: A broadband achromatic lens

For this example, it was required to produce a miniature lens of 12.5 mm focal length that would be diffraction-limited and achromatic in the range 280-2500 nm. The object is at infinity and the required f-no is 5. The lens has a limited field of view, $\pm 2.2^\circ$. These specifications together with the geometry of the problem ruled out a reflective single mirror or multi-mirror system.

How might this problem be attacked? Given sufficient time, the designer might be tempted to extend the traditional theory of achromatization over a broad wavelength range. But under time pressure, an alternative way is forced upon the designer. This alternative hinges on 1) the realization that only few well-behaved materials transmit well enough over the required range, and 2) the ability of the global optimization algorithm to work with discrete glasses* as opposed to treating glass as a continuous variable. This eliminates the need to reoptimize the design after substituting real glasses and also prevents the optimization from wandering off to unobtainable index and dispersion values in the case where the glass catalog is very sparse. The solution then was to simply construct a glass catalog with the few available materials (fused silica, CaF_2 , MgF_2 , BaF_2 , LiF , SrF_2 , sapphire, KBr , KCl) and then let the computer find the most suitable combination. In fact, one can further reduce this catalog by concentrating on the most common materials and avoiding hygroscopic ones such as the potassium compounds. Since diffraction-limited performance is sought over some small but non-negligible field, a triplet construction is a reasonable guess; air-spaced form is also desirable.

It may come as no surprise to the initiated that the optimization settled on a MgF_2 -silica- MgF_2 solution. This is no different from the typical positive crown, negative flint, positive crown triplet, except that the extremely low index and dispersion of MgF_2 allows fused silica to play the role of the flint. MgF_2 has the lowest dispersion of all available materials, while fused silica has the highest. However, although the solution may seem obvious with hindsight, it remains a fact that the computer arrived at it faster than the designer, and in the meantime allowed the designer to work on a different task.** In any case, even if the designer could have guessed correctly the best choice of glass, s/he would have still needed to submit the design to optimization, not only to come up with the best lens shapes, but also because the simple intuition based on a limited achromatization range would require confirmation.

A wrinkle in the design process is that the chosen material, MgF_2 , is birefringent. Hence there are two alternatives: evaluate the resulting lens accounting for the birefringence, or choose an alternative material. In practice, the only reasonable alternative material to MgF_2 is CaF_2 which has the next lowest dispersion. Upon replacing for CaF_2 and re-optimizing, the desired performance is not achieved. On the other hand, since the field and aperture angles are relatively low, the rays will remain close to the optic axis of MgF_2 . After adding the birefringence, and without further changes, the extraordinary rays do not provide a diffraction-limited spot, but upon re-optimization of the lens shapes a good compromise is found that works for both the O and the E-rays simultaneously.

For the sake of completion, the prescription of the lens is given in Fig. 4, and its performance in Table 1. It may be seen that the Strehl ratio is everywhere above 0.8, so the lens satisfies the conventional definition of being diffraction-limited. These values are of course calculated at a single, common focal plane otherwise they would both be higher yet. The total secondary spectrum is less than 0.1mm. The longest wavelength

* As far as the author has been able to ascertain, only ZEMAX offers this capability from among the major commercial packages.

** Actually, ref. 13 contains a CaF_2 -silica- CaF_2 triplet operating over a narrower band (250-730 nm) that could also have provided a clue, although CaF_2 does not work over the much broader band of this example.

(2500 nm) is responsible for most of the secondary spectrum, but the solution takes advantage of the long depth of focus afforded at the longest wavelength (the corresponding secondary spectrum of the CaF₂ solution was more than 0.4 mm). In terms of transverse error, the birefringence produces an image shift of only ~0.5 μm at the edge of the field, when the smallest Airy disk radius for the 280 nm wavelength is 1.7 μm. Because the birefringence is relatively small, the lens can be used at a wider field if the diffraction-limited performance requirement is somewhat relaxed.

Surf	Type	Radius	Thickness	Glass
OBJ		Infinity	Infinity	
		Infinity	3	
STO	birefr.	6.719	1.507	MGF2
	birefr.	-4.097	0.288	
	std	-3.424	1.000	F_SILICA
	std	19.478	0.135	
	birefr.	49.112	0.990	MGF2
	birefr.	-4.649	10.938	
IMA		Infinity		

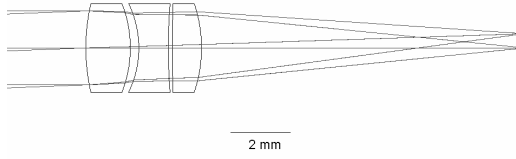


Figure 4. Prescription and schematic of F/5, ±2.2°, F = 12.5 mm broadband achromatic objective.

Table 1
Strehl ratio for various wavelengths and field positions (values for the E-rays in brackets)

	280 nm	400 nm	530 nm	800 nm	1200 nm	2500 nm
On-axis	0.99 (0.89)	0.97 (0.89)	0.92 (0.85)	0.94 (0.90)	0.98 (0.97)	0.92 (0.93)
max. field	0.81 (0.87)	0.95 (0.95)	0.98 (0.95)	0.98 (0.96)	0.99 (0.99)	0.89 (0.90)

Perhaps the ultimate case of substituting the computer for the designer is provided by the design that starts from a set of plane parallel glass plates. While I have not yet employed this as a routine method, I have submitted a set of plates to optimization using a merit function which was initially developed after a considerable amount of iterative optimization that finally succeeded in converging to a satisfactory solution. The system was a wide angle (30° half-field of view), relatively fast (F/3.1), fully achromatic and diffraction-limited, with significant size and other restrictions. Using the same merit function, and starting from a set of plane parallel plates, the computer converged, within about an hour, to the same solution that I had reached after some weeks of on-and-off work. The deck was stacked a little in favor of the computer, since I retained the same number of plates as elements, including cemented interfaces, and since the merit function contained important constructional considerations that were not obvious at the beginning of the design process. All the same, the experience was certainly humbling.

3.2 Educational uses of optical design software

Returning to the use of optical design software in teaching, some of the latest books in optical design are practically integrated with the software.^{14,15} Indeed, it has been claimed that nowadays it would be a disservice to the students to teach optical design without familiarizing them with at least one commercial software package.

In my teaching experience I found it very difficult to use advanced software as a teaching tool. The use of the software tended to raise more questions than could be reasonably answered within a one-quarter course. Its many features can overwhelm all but the best prepared of students, and force the instructor to give cursory explanations. There is a lot of theory underlying the use of optical design software, and not enough time to cover both theory and use of software in one class. At first, my solution to this dilemma was to place the emphasis on the theoretical preparation of the students, and then to rely on the fact that the

increasing user-friendliness of the software would make it possible for the increasingly computer-savvy students to pick it up on their own as the need arose. Indeed my own personal example was in a way a validation of that approach, when I was forced to become a lot more familiar with the software as an engineer than I ever was as an instructor – I found no case where I regretted knowing more of the theory and less of the software.

However, an important change has taken place in the last ten years or so. Optical design software has evolved from a mere lens design tool to a modeling and analysis tool. This affords opportunities for introducing the software at an earlier stage and familiarizing the student with the interface as well as at least some of the advanced capabilities of the software without ever mentioning optical design, which is a specialized activity and requires courses in aberrations before it can be taught. Optical design software can and should be used to model the effects of interference, beam propagation and diffraction, polarization, and thin films, and to some extent even coherence. The analytical capabilities of software are constantly expanding, so its use will ultimately be limited mostly by the instructor's imagination. Idealized optical systems using perfect lenses or no lenses at all can be set up to explore wave optics problems. For example, ZEMAX contains ready-made examples of such things as an amplitude zone plate, Talbot imaging, etalon transmittance, Twyman-Green and Mach-Zehnder interferometers, and several others. By discussing these problems not just as theoretical, closed-form solutions but in the context of a practical computation, the student will be introduced to the problems of numerical analysis and computational algorithms that are in practice as important as the theory in terms of producing accurate quantitative answers.

Perhaps the simplest and most instructive of all these examples is the propagation of a plane wave through a circular aperture, which can also be stated in the context of a pinhole camera.¹⁶ A simple laboratory experiment or even classroom demonstration can show the evolution of a plane wave as a function of the distance from the diffracting aperture. One need not know anything about lenses in order to set up this problem using commercial software. The problem can be stated in terms of trying to find the best pinhole size for a given camera box size. In fact the number of questions that can be investigated will likely be limited mostly by the available time. Thus geometrical and wave theory, experiment, and practical computation can all come together in a most fruitful way that is bound to leave a lasting impression upon the student. A side advantage is that even those students who do not choose to take the optical design elective will become familiar with the software and may at some later stage be able to use it to satisfy modest design needs. From the opposite viewpoint, the incorporation of so many advanced analysis features into the optical design software reduces greatly the need for writing one's own code, but at the same time creates an expectation for the designer to perform sophisticated analysis that would be handed to experts some years ago. Thus it is more critical than before to familiarize the student with the advanced features of the software and the necessary theory to understand its many functions.

4. SOME ISSUES SPECIFIC TO OPTICAL DESIGN

It is within this section that one's own experience is most evidently personal and limited. One's list of useless topics may be another's list of useful ones. One might be tempted to offer objective-sounding arguments in support of one's views, but the value of such arguments is probably limited. Here, in any case, are some comments on what I have found useful and what less so, as well as what I could have used more of at the outset.

4.1 Missing or underrepresented topics in traditional curricula

Of course, it is impossible for optics curricula to cover everything. The engineer will have to figure out many details on his or her own after graduation. For this reason, the topics in this subsection are only the few most important ones.

1) *Reflective systems, especially unobscured.*¹⁷ From reading the literature, one might get the notion that optical design is all done with lenses. Yet in my years of practicing, I came closer to believing that it's all done with mirrors. About 70% of the systems I have had to design were basically reflective. This may well be a biased sample, and of course at some deep level there is no fundamental difference between reflective

and refractive systems, but there are important differences all the same. Unobscured mirror systems deserve some more extensive than usual treatment as they are not intuitive unless one is an inventive genius. It is necessary to spend at least a little time with the basic mirror systems and instill the simple intuitions of how an unobscured system can be an off-axis part of an obscured one, and further how the theory of axisymmetric systems applies or does not in this case. Also, the ring-field system and how it leads to a wide FOV in one direction should be discussed.

2) *Zernike polynomials*. Few books pay sufficient attention to the Zernike coefficients, yet they are of more use in analysis than the Seidel terms. In this respect, ref. 16 provides a detailed description and analysis. It is necessary to become as familiar with the Zernike coefficients as one is with the Seidel ones.

3) *Stray light*. Usually, little or nothing is taught about this topic. Yet an optical design is not complete without an evaluation of ghosts and scatter. Probably every optical engineer has either personal or proximal experience of a stray light or ghost horror story. Perhaps not much more needs to be taught than the basic principles of stray light control and good practices with a few examples. But even that is often missing.

4) *Optical design for the thermal infrared*.² One need not expend time with the many details, but just make sure that the key insights and differences between IR and visible systems are understood, for example that the energy seen by in an infrared system is not necessarily only what passes through the lens aperture, as well as introducing the concept of the cold stop.

5) *Ray aiming*. The term is not usually encountered in optics texts, yet it is an important topic in the practice of using optical design software. Ray aiming is a way of taking into account pupil aberration in the launching of rays. Pupil imagery in general is a tricky topic that does not seem to receive much attention but asserts itself as soon as one tries to understand the definition of the chief ray. In this respect, there is a danger that the subtleties of the topic will not be appreciated if all one does is to rely on the various default options of commercial software. The problem can probably be fully appreciated only if one attempts to write one's own routines for ray tracing and image evaluation. But at the very least it should be explored through suitable examples.

4.2 Material used rarely or never

Certainly, one risks incurring wrath upon pronouncing any part of the theory to be useless. To be clear then, the topics below are not useless; they are just topics that this engineer has had no reason to use until now.

1) *High order aberration theory*. Beyond the basic idea that higher orders sometimes can balance lower orders, such balancing is always done automatically by the software. The only caution found necessary is to make sure that the pupil is sampled adequately in order to avoid the case where the aberration is corrected exactly at a few zones but is larger than expected in between.

2) *Coddington's equations, Conrady's D-d chromatic sum, offence against the sine condition, and other similar formulas*. These seem to have been derived at a time when it was important to provide computational shortcuts for evaluating the lens performance. Since the lens can now be analyzed with lightning speed and several alternative measures of performance computed almost instantaneously, these shortcuts seem to have lost much of their appeal. Each one may provide some particular insight into the behavior of rays, but in the competition for biological memory space they would tend to lose out to more modern concepts of general optics and advanced technology.

3) *Third order correction by hand computation and thin lens pre-design*. This really comes under the category "preached but rarely practiced". To be clear, the theorems relating to third order aberrations have been found very useful as they illuminate the behavior of a system. But setting out to derive a system by correcting third order aberrations of a thin lens design is too laborious a process compared with the alternative of submitting a preliminary solution to optimization, assuming of course that suitable software is available. A solid merit function can be composed in a small fraction of the time it would take to derive third order correction, and that merit function will eventually correct not only third but also higher orders, as well as control constructional parameters, manufacturability, and other important constraints.

4) *Ray fans*. Despite their popularity among many experienced designers, I have never found an occasion where their use would have changed the design approach. The most profitable way to evaluate a design is by devising a figure of merit that suits its intended use. Such merit figures are the MTF, encircled/ensquared energy, Strehl ratio, distortion, etc. Ray-based evaluation is almost never sufficient. As for the diagnostic power of ray fans, I believe it is at most no better than the diagnostic power of a wavefront map, but the latter is also useful in assembly, alignment, and testing.

5) *Eikonals*. What?

4.3 Tremendously useful topics

The topics below are those that have been found useful not just once or twice but practically all the time. While they may appear to be rather elementary, they are also endless in their application and not as fully appreciated as they should be, sometimes even by practicing engineers. Certainly, any time spent in an optics curriculum trying to instill a deeper understanding and more detailed knowledge of these few topics is time well spent.

1) *A clear understanding of pupils and stops and pupil matching*. This seems to be an almost interminable topic with a different twist in almost every new system.

2) *The process of first order system layout using the marginal and pupil rays*.¹⁸ This is useful even when the system is copied from previous designs in providing a clear understanding of what each component is doing.

3) *The Lagrange invariant and its relation to resolution and throughput*. One suspects that not a day passes without someone, somewhere, attempting to design a system that will violate the Lagrange invariant. This must be the optical equivalent of the perpetual motion machine – a quest that will go on forever.

4) *A detailed knowledge of wavefront and PSF shapes as a function of various aberrations*. All the time spent looking at star images or interferograms has been time well spent.

5) *An understanding of how certain aberrations may arise within any particular system* (e.g., due to symmetry or lack thereof). This, as well as the previous item have been useful mostly during system alignment in deciding what is the likely component causing a certain aberration and what motion of that component would fix it.

6) *Computational principles and issues regarding diffraction calculations*. These are items relating to the accuracy of numerical Fourier transformation, sampling, zero padding, re-sampling, etc.

4.4. Example: Amplitude distribution at the pupil and use of ray aiming

An example will illustrate a pitfall resulting from a potential misunderstanding of the use of ray aiming as well as inadequate understanding of the features of commercial software. If light is incident on a powered mirror at an off-axis angle, the ray density after the mirror is generally non-uniform, generating an inhomogeneous wave, although the effect is usually small enough to be unnoticed. If, however, an aperture stop is placed after the mirror and ray aiming is turned on, the stop will appear artificially as having uniform ray density. Consider a system of two off-axis parabolic mirrors with a single object point at the focus of the first one (Fig. 5). Place an aperture in the middle of the collimated space as shown. One might immediately expect that such a system would produce a perfect image, implying a perfect Airy pattern. But such is not the case. If a spherical wave is launched at point A, a set of uniformly spaced rays emanating from A will not remain uniform in the collimated space, and the effect will be further exacerbated by reflection from the second mirror. Thus light focusing at B should show some effect from this non-uniform amplitude. Ray aiming however forces the distribution in the collimated space to become uniform, thus artificially producing a much smaller effect.

To appreciate this fully, consider the ray density distribution at the intermediate plane between the two mirrors, shown in Fig. 6, and the corresponding Point Spread Functions shown in Fig. 7.

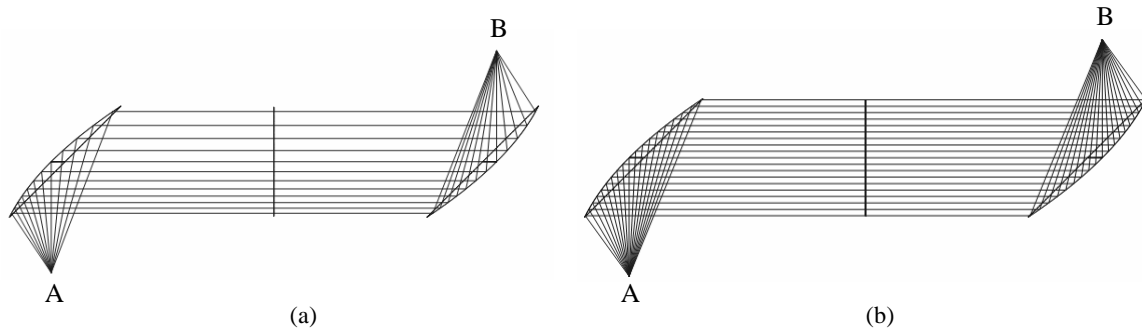


Figure 5 (a): Two parabolic mirrors with no ray aiming. (b): with ray aiming. In the second case, observe that the ray density in the collimated space is constant, implying a homogeneous plane wave. In the first case, the collimated space shows an inhomogeneous wave, but the spherical wave launched at point A is homogeneous, corresponding more closely to the physical situation.

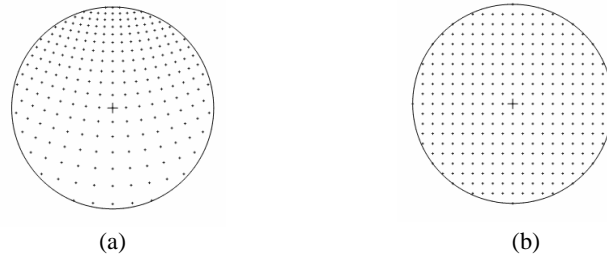


Figure 6 (a): ray distribution in the collimated space without ray aiming. (b): with ray aiming.

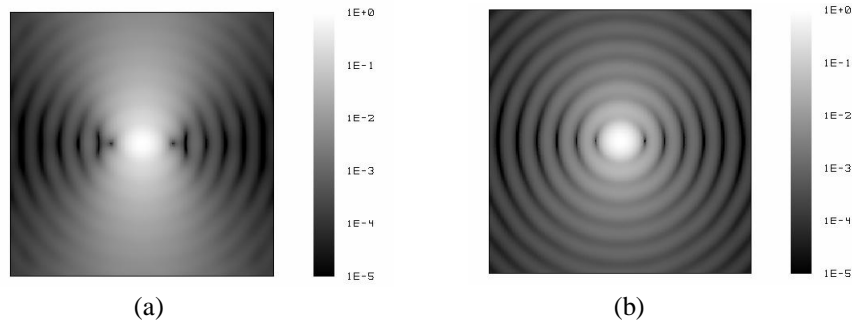


Figure 7. Point Spread Function (logarithmic scaling) for light focusing at point B, Fig. 5. (a): no ray aiming. (b): with ray aiming. Notice that the PSF on the right still shows some small asymmetry (the minima are darker along the horizontal than the vertical).

While the asymmetry shown in Fig. 7 is not particularly large, it can be significant in systems which require extreme wavefront and amplitude correction (e.g. a coronagraph). In any case, the understanding of the physical situation must be reflected in the model, or unexpected errors may arise.

This simple example may be taken even further. Suppose the second mirror is rotated to the orientation shown in Fig. 8. Evidently, the effect of the first mirror is now canceled (rather than exaggerated) by the second one, resulting in a perfect Airy pattern (a uniform spherical wave coming to focus). Again, this effect is best represented without ray aiming. Turning ray aiming on produces a situation similar to the

previous one and introduces a small but nevertheless artificial asymmetry to the PSF such as shown in Fig. 7(b).

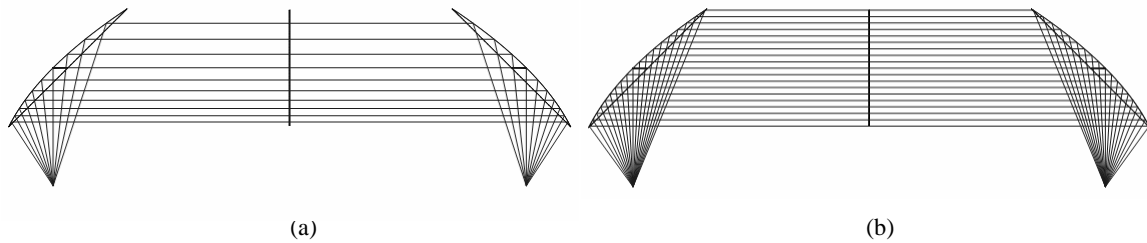


Figure 8 (a): with no ray aiming, the second mirror cancels the amplitude nonuniformity introduced by the first one. The result is a perfect Airy pattern. (b): With ray aiming, the situation is no different from that of Fig. 5(b), resulting in a slightly asymmetric PSF such as shown in Fig. 7(b).

A different way of seeing this example is to consider what would have happened if the system had been designed through optimization. If, as is common practice, the merit function included only OPD or ray aberration, then the two-parabola system would appear to be perfectly corrected with zero merit function, since there is no phase aberration. This result would hold irrespective of the orientation of the second mirror and also irrespective of whether the designer chose to employ ray aiming or not. Yet, upon illuminating the system of Fig. 5(a) with a uniform spherical wave, one would discover a less than perfect diffraction image. The system of Fig. 5(a) would optimize correctly only if the amplitude uniformity of the exit pupil was specifically included in the merit function. And, if the orientation of the second mirror was allowed as a variable, then the optimizer would hopefully settle on the solution shown in Fig. 8, which indeed will provide a perfect image.

5. EPILOGUE

As optical and non-optical technologies advance, the question that arises is how much to teach the specifics of the present technologies, which are almost guaranteed to be out of date within the working life of a young engineer, as opposed to teaching basic areas of physics that not only will retain their relevance but will also provide the basis for new technological advancements. The optical engineer cannot afford to be a narrow specialist, but must have sufficient depth of knowledge to comprehend and incorporate, if not invent, new technologies as they arise.

The development and wide dissemination of optical design software has democratized the practice of optical design to the point where only a few esoteric areas are left as the exclusive domain of the specialist. Many engineers can now satisfy their relatively modest design needs without resorting to an expert designer. At the same time, the demand for sophisticated system analysis is increasing. This means that the optical engineer will need a broad education, achieved not through a superficial survey of technology but by teaching physical and mathematical principles at a sufficient depth so they can serve as a springboard for understanding the operation of new devices. Moreover, it is absolutely essential to teach at least the necessary minimum of computational techniques, as it is not possible to employ successfully the various codes to perform advanced analysis such as detailed diffraction propagation without understanding the features and limitations of the underlying computation.

ACKNOWLEDGMENT

Parts of the work described here have been performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. I am indebted to the many colleagues who taught me these lessons, whether intentionally or otherwise (names withheld to protect the innocent). I am also particularly thankful to Drs. Virendra N. Mahajan and Azmi Q. Kaskaly for

reviewing this manuscript and suggesting many improvements, most of which I gladly adopted. If indeed the final result remains highly imperfect, the fault lies exclusively with me.

REFERENCES

- 1) W. J. Smith: *Modern Optical Engineering*, 3rd ed. McGraw-Hill, 2000
- 2) R. E. Fischer and B. Tadic-Galeb: *Optical System Design*, SPIE Press, 2000.
- 3) R. R. Shannon: *The Art and Science of Optical Design*, Cambridge University Press, 1997
- 4) R. H. Vollmerhausen and R. G. Driggers: *Analysis of Sampled Imaging Systems*, SPIE Press, 2000
- 5) P. Mouroulis: "Beyond spot diagrams: End-user oriented optical design", SPIE Proc. **4832**, 181-185 (2002)
- 6) D. Kwo, G. Lawrence, and M. Chrisp: "Design of a grating spectrometer from a 1:1 Offner mirror system", SPIE Proc. **818**, 275-279 (1987)
- 7) A. Offner: "Unit power imaging catoptric anastigmat", U.S. Patent No. 3,748,015 (1973)
- 8) P. Mouroulis: "Low-distortion imaging spectrometer designs utilizing convex gratings", SPIE Proc. **3482**, 594-601 (1998)
- 9) P. Mouroulis and M. McKerns: "Pushbroom imaging spectrometer with high spectroscopic data fidelity: experimental demonstration", Opt. Eng. **39**, 808-816 (2000)
- 10) P. Backlund, D. W. Wilson, and R. E. Muller, "Structured-groove phase gratings for control and optimization of spectral efficiency," in *Diffraction and Micro-Optics Topical Meeting*, Optical Society of America, Washington, DC, 2004), paper no. DMB6
- 11) P. Mouroulis, D. W. Wilson, P. D. Maker, R. E. Muller: "Convex grating types for concentric imaging spectrometers", Appl. Opt. **37**, 7200-7208 (1998)
- 12) S. L. Murchie and 42 co-authors: "CRISM (Compact Reconnaissance Imaging Spectrometer for Mars) on MRO (Mars Reconnaissance Orbiter)" SPIE Proc. **5660**, 66-77, (2004)
- 13) W. J. Smith: *Modern Lens Design*, 2nd ed., McGraw Hill, 2004
- 14) G. H. Smith: *Practical Computer-Aided Lens Design*, Willmann-Bell 1998
- 15) J. Geary: *Introduction to Lens Design with Practical Zemax Examples*, Willmann-Bell 2002
- 16) V. N. Mahajan: *Optical Imaging and Aberrations Part II: Wave Diffraction Optics*, SPIE Press, 2001
- 17) D. Korsch: *Reflective Optics*, Academic Press 1991
- 18) P. Mouroulis and J. Macdonald: *Geometrical Optics and Optical Design*, Oxford University Press, 1997