Defence Science Journal, Vol. 3 , No. 1, January 1953, pp.35-41

OPTICAL DESIGN AND COMPUTATION*

BY MR. F. H. BOPP, WORKS MANAGER, ORDNANCE FACTORY, DEHRA DUN.

I have the honour and pleasure of addressing you with a few words on the subject of optical design and computation. Although there are many comparatively recent developments on the improvement of optical image formation, these are mostly only of considerable interest to the specialist and are liable to bore the majority of an audience with diverse interests. I shall therefore, confine myself to some general aspects on the subject, taking into consideration the fact that even in other countries the knowledge of optical design and computation is limited to a comparatively small group of specialists ; to India the subject is practically entirely new.

To begin with I should like to make a few remarks about the requirements and possibilities of an opitcal system. For the sake of conciseness I shall confine myself to those systems forming a real image of a real object as it is the case in the most usual instrument for visual and photographic purposes. As regards the desired performance of the system, that is, image quality, one often hears the simple statement that for best imagery the image should be as similar to the object as possible. Without certain reservations this is neither possible nor even desirable. If the object if a solid and is located nearer than practically infinity or, in other words, the magnification is not zero and is different for different depths of the object, the image, according to the above stipulation would again be a solid. But this solid image would by no means be similar to the object since the lateral size of the image would be directly proportional to the magnification, the depth of the image, however, would be proportional to the square of the magnification. Apart from this it is usually not even desirable that a solid image be formed since the receivers of the image, generally the eye or the photographic plate are not capable of simultaneously registering correctly image surfaces that lie substantially apart. We therefore must demand that the object be imaged on a surface and not in space. But strictly speaking also that is not possible for a solid object, except at zero magnification, since no optical system can have only one image conjugate for any number of object conjugates. All the foregoing can be easily deduced from the most elementary laws of first order or Gaussian Optics.

In actual practice there will always be a certain range of the solid object which for all practical purposes will be imaged on a surface. This range depends upon the relative aperture and focal length of the system, the magnification, and the image quality, and is called the depth of focus. The very nature of light causes an infinitely small object point to be imaged as a patch

* Paper communicated to the Defence Science Conference held in April, 1952

of finite size called the diffraction pattern, its size depending upon the angle of the light conforming the image.

This fact in conjunction with the limited resolving power of the image receiver gives every optical system a sort of inherent depth of focus within which the object surface may move without noticeable difference of image quality. An increase of this depth of focus can only be obtained at the expense of resolving power or image quality. As the angular depth of focus of any system of high resolving power is only a very small fraction of the convergence of the system it is practical for design purposes to consider also the object as a surface.

Having thus made the reservations that both object and image are surfaces we can define perfect imagery by saying that perfect imagery obtains when every detail of the image is as similar to every detail of the object as the nature of light will permit. As to shape and position of the object and image surfaces we, not necessarily, but mostly assume the object surface and require the image surface to be flat and perpendicular to the opitcal axis. This is also in conformity with Gaussain image formation, according to which a system is assumed to possess no aberrations and is therefore capable of perfect imagery. Aberrations in this connection mean deviations from perfect imagery mathematically inherent in the system, not aberrations caused by faulty construction.

It is quite beyond the scope of this paper to go into the question of the large number of possible aberrations, their intricate inter-relation and their effects upon the image. Let it suffice that the study of higher order optics leads us to the following conclusions. Every optical system possesses certain aberrations. These aberrations vary with the choice of the conjugates or magnification. In principle a general system can only form a perfect image of a single object point on the axis at only one given magnification. A strictly symmetrical system is capable of doing this at two reciprocal magnifications. An ideal optical system giving perfect imagery over a finite field is a fiction, it does not exist. The mentioned fact that the aberrations vary with the magnification impose another restriction from the design point of view, namely the predetermination of the conjugates. A telescope or photographic objective for outdoor work for instance, will be corrected for infinity or zero magnification. A reproduction or micro objective, on the other hand will be corrected for the magnification at which it normally is used. High aperture systems such as high power micro objectives are indeed very sensitive to any change of the predetermined conjugates. A change of only a few milimeters in the tubelength of a microscope can noticeably affect the performance of such systems. Regarding the possible aberrations in an optical system I would like to point out that these are, contrary to usual opinion, by no means restricted to the five monochromatic aberrations namely spherical aberration, coma, astigmatism, field curvature, distortion, and the two longitudinal and transverse chromatic aberrations. These are merely the so-called third order or Seidel aberrations which are particularly noticeable and usually predominant in systems of moderate aperture and field. Theoretically the number of aberrations is unlimited. In systems of large aperture or/and field, even in actual practice the number of effective aberrations and their chromatic variations is very large. In a new type of high power micro objective which we are developing at O.F. Dehra Dun aberrations as high as the eleventh order are still effective. In the eleventh order no less than 27 monochromatic aberrations are possible, to say nothing of the chromatic aberrations and variations.

The foregoing, particularly the statement that an ideal optical system does not exist sounds very discouraging and, indeed, it constitutes one of the greatest difficulties for the designer. On the other hand the fact that the nature of light itself will not permit the formation of an ideal image in the geometrical sense and the fact that the resolving power of the image receiver is always limited brings it about that certain magnitudes of aberrations may be present without noticeably affecting the image quality. This certainly is fortunate, for after all, what harm can aberrations do if they are not noticeable? Again, the limit up to which aberrations are unnoticeable is extremely small. The human eye is extraordinarily sensitive to aberrations. The conventional Rayleigh limit of one quarter wave-length holds good only for systems of moderate aperture. In high aperture systems aberrations even below a twentieth of a wave-length may become noticeable in the intensity distribution of the diffraction pattern and therefore can impair the resolving power. No hard and fast rules exist for laying down the aberration tolerances; the limits will vary from case to case. In all cases it must be borne in mind that even the best optical system is in principle only a more or less close approximation to the desired ideal. Every optical system however well corrected, will have a number of residual aberrations which are all more or less inter-related and dependent upon each other. The amounts of these aberrations vary from system to system, their relative importance varies from case to case. Certain important aberrations may be reduced by leaving or even purposely introducing other less important ones. The final system will always constitute a compromise between the possible and the desirable solution. It is the production and selection of the best compromise out of the almost infinite number of possible ones which forms one of the greatest difficulties for the designer.

Although in the course of time a very large number of different ingenious mathematical methods and procedures have been developed for the purpose the design of a precision corrected optical system still remains in principle a mathematical "cut and try" method. Using even most modern techniques and machines no method exists which, by solving a number of however complicated equations, will in general yield the parameters of a system giving a required precision performance. With very rare exceptions analytical methods will at best only yield more or less close approximations to the final design. The approximations are tried out mathematically and if the results are not satisfactory, certain variations of the parameters are introduced; the variated system is again tried out and so on until the system performs according to requirements or at least the performance is the best possible under the circumstances. From the foregoing it is quite obvious that also here very considerable experience on the part of the designer is required to interpret the trial results correctly, to select the right type and amount of parameter variations from the usually vast number of possible ones and to finally decide that the best possible solution has been achieved and that the performance, after taking the unavoidable manufacturing tolerances into account, will be adequate for the purpose in question.

A few words may also be included on the difference between the work of an optical designer and computer. After consideration of the performance requirements, the designer will lay down the basic tentative design which he believes will meet the requirements. The computer then computes the aberrations, usually in third order approximation to save time, from which the designer decides the variations necessary for improvement. After the system shows satisfactory figures in the third order approximations either higher order aberrations are taken into account or, more frequently, accurate ray traces are made by the computer. From these or from computed parameter-aberration differentials the designer again suggests the necessary improvement variations, and so on until the results are satisfactory. From the aforementioned, it may be seen that the computer, as his desingation implies, computes the aberrations of a given system. However complicated and tedious these computations may be particularly so in high aperture systems where geometrical optics are no longer sufficiently valid and wave-front deformations, phase differences, and intensity distributions must be considered, the computer is only required to produce the performance figures of the system according to certain well defined procedures and with the correctness of these figures his responsibility ends. Obviously a mathematically well-trained and talented person can, after a comparatively short period of specialised training, become a first class computer. The work of the designer is of quite a different nature. Only a very limited amount of his work is laid down by routine procedure. The choice of the basic design, materials, and the type of computation required, the interpretation of the results, the suggestions for improvement, the final decision for acceptance of the system for the purpose in question, the responsibility for possibility of economic production, etc., requires a very considerable amount of aptitude, experience, and last not least intuition.

Of interest may also be a few remarks about optical designs and computation methods. Without doubt, the very first attempts at optical design were made by actual experiments with lenses. It need not be pointed out why this method was highly unsatisfactory, slow, and costly. As soon as the general laws of refraction at curved boundary surfaces were sufficiently understood, trignometrical methods for tracing individual rays through a system were developed in about 1795 to replace the physical experiment. Optical design being, as already mentioned, in principle a mathematical " cut and try " method, it is significant that for a long time development mainly aimed at hastening the " tries " i.e. to devise methods to trace the rays through the system in the shortest possible time. Simultaneously there was a sort of obsession for accuracy. Until as recent as 25 or 30 years ago, it was quite general practice to dive from first order optics of even a tentative system straight into six or eight figure ray tracing. It must here be pointed out that despite the considerable work involved in an accurate ray trace, beyond the accurate penetration point of the ray in the image plane, the trace of a single ray gives us no information whatsoever. Only a comparatively large number of such ray-traces could give us sufficient information on the performance and the insight into the general conditions and potentialities of the system would still be very meagre indeed. Although in principle it demands very considerable and tedious labour to compute by this method all the aberrations and particularly the contributions of the individual surfaces, a knowledge which is of vital importance for systematic improvement. Far more valuable for the initial stages of a design would be an easy and quick method of computing performance and surface contributions even if the results were only rough approximations. For, after all, what is the value of accurate figures of a tentative design which in all cases will be modified and which may even be scrapped entirely. It is like attempting to measure a slip gauge interferometrically when it is only rough ground. It is indeed surprising that such analytical methods, particularly the third order or so-called Seidel approximations were not used until comparatively recently although they were known to exist, practically in the form now used, since about 100 years ! But initially they were almost ignored as being too inaccurate. Even Seidal himself, for the same reason, was sceptical as to their practical value. But in actual fact, with very little labour and with only pencil, paper, and a slide rule as facilities one may gain a rough idea of performance and surface contributions and above all a deep insight into the possibilities of a system.

Naturally in the more final stages of design accurate ray tracing or an equivalent thereof is imperative. Also in this respect development has taken place only in comparatively recent years. Before introduction of the desk calculator the usual method was purely trigonometrical using log tables. When using desk calculators log tables are of course unnecessary and the natural values of the trigonometrical functions are used, preferably in decimal degree fractions or radians, as arguments. The bulk of the tables can be reduced to a small fraction if one resorts to sine-radian difference tables, a method which we use exclusively where only non-automatic desk calculators are available. With fully automatic desk calculators we prefer the purely algebraic vector method which does away with trigonometrical tables entirely and requires only a one page square root table to give the approximate values for the extraction of the accurate square roots by a simple iterative process. In a certain modification even this table can be avoided by solving a simple quadratic equation by iteration. These purely algebraic methods are, of course, particularly adaptable to automatic punch card controlled calculators. Again, certain trigonometrical path difference methods require only four figure computation to yield a result which is equivalent in accuracy to a seven or eight figure ray-trace. I may recall my remark concerning early endeavours to carry out a ray trace in the least possible time. But such a ray trace will necessarily yield no safeguard against errors and above all will yield no additional information beyond that of the ray path. The present trend is to use such methods that will as a by product, supply the parameter differentials so valuable for further improvement and to introduce checks at each stage to disclose errors at once, even if the trace thereby takes a longer time. In the end this additional time is usually well invested.

I should like to add a few words on some rather unusual possibilities which we are trying to develop into a practical and useful method of tackling the problem. After a design is completed I always somehow feel discouraged when I compare the few valuable result figures showing the performance of the system with the sometimes huge number of figures on piles of computation sheets necessary to obtain this little result. When designing a basically new high performance system the number of computation sheets can run into a thousand and more, every sheet densely covered with figures. An enormous amount of work has been put into this ocean of figures but they are so highly specialised according to the particular ray-path that, with rare exceptions, they are practically useless for future reference. Even if, by mere coincidence, an identical ray path should occur sometime in the future it would be quite impossible to find the data of its predecessor. Obviously it would be very valuable if at least a part of the figures would be in the form of constants of the system elements which are independent of the particular ray path and could therefore, once calculated, be used over and over again. In principle this actually can be done. In first order optics the Gaussian expressions are admirably suitable to such a scheme particularly if dealt with by matrix algebra. Each system element, i.e. each surface and interspace is represented by a 2×2 matrix containing the necessary constants and the matrices are preferably entered singly on index cards which can be suitably filed for future use. These matrices are simply multiplied according to the usual rules of matrix multiplication in the order of their occurrence in the system and the resulting 2×2 product matrix yields the four Gauss constants of the entire system from which all necessary first order quantities can be derived. As the matrices of flat surfaces and zero interspaces reduce to unit matrices, they can simply be ignored. If the actual paraxial ray-path through the system is required it is only necessary to introduce the corresponding conjugate matrix at the beginning of the procedure. The intermediate product matrices will show the required quantities. Now matrix multiplication is in general of course not commutative but it is always associative. Therefore, when using this method, one will preferably group certain element matrices together to form whole components such as thin or thick lenses, cemented doublets or the like. The product matrix of each group can then also be entered on a card to be filed away for future use in case the same component should occur in another system. For first order optics this method is really ideal since not only the element matrices but also many of the intermediate or group product matrices can be used over and over again in any system. Also the two paraxial colour aberrations can easily be included into such a scheme. It is only necessary to introduce the colour differentials whereby each element is represented by a 4×4 matrix. Otherwise the procedure and the advantages are exactly the same. The resulting product matrix will now yield not only the first order quantities but also the longitudinal and transverse colour aberrations in the paraxial region. The five monochromatic third order aberrations can be included necessitating a 10×10 matrix for each element. The inclusion of the fifth order aberrations would presumably require a 20×20 matrix. In principle the method can be extended upto any order but it becomes more and more impracticable as the order of the aberrations rises. With automatic desk calculators the multiplication of the 2×2 matrices is quite fast and of the 4×4 matrices sufficiently so.

But the multiplication of 10×10 or higher order matrices, although in principle simple enough, becomes very tedious and tiresome and I fear that if I should request our computers again to compute a system with 10×10 matrices they would probably apply for permanent transfer or at least for leave on compassionate grounds. However, I am convinced that certain simplifications will be possible, probably at the expense of some of the complete generality mentioned and that the method has very great possibilities when using punch card controlled calculators.

Ever so much more could be said on the subject but I am afraid to imerge into boring details and to try your patience more than I have already done. I should therefore only like to conclude with a suggestion.

Since almost 2½ years I am concerned with optical data and optical design at Ordnance Factory, Dehra Dun, and in all that time I have never seen the constructional data of any optical system to be manufactured accompanied by performance data, no matter where the constructional data came from, whether from the T.D.E., from abroad or from anywhere else. Whenever I asked for performance data I was told that it was not available. Only in two cases did I receive a few figures which were so meagre and incomplete that only an expert could prove them to show unsatisfactory performance of the system in question. Surely the indentor and the user would like and are entitled to know what they are getting for their money. From certain quarters I have heard the reply in this connection : " The indentors and users are laymen who would not understand the figures anyway ". I do not agree with this remark, provided the performance figures are represented in a suitable manner. It is true that performance figures alone will usually mean nothing to a layman but every intelligent person can read a simple graph and if this graph, besides the important aberrations, also contains the respective conventional tolerances even a layman can see at a glance whether these tolerances are exceeded or not and can, above all, compare the performance of comparable systems. It is also true that in some cases an opinion or a comparison is not as easy as all that, particularly when certain aberrations have been reduced at the expense of others. But explanations can always be given by the supplier of the performance data or by any expert on the subject. I therefore suggest that the optical performance data for every new system to be manufactured be supplied to indentor and user for approval before manufacture. I also suggest that for this data certain standards of representation, depending upon the type of system under consideration, are strictly adhered to in order to make data of the same type directly comparable. If the inspection authorities are going to introduce adequate optical performance inspection, which must be done sooner or later to raise the standard of our optical instruments, inherent performance data is absolutely essential. Otherwise how is the inspection to know what the instrument can and should perform and how can they distinguish between the aberrations that are inherent in the design and those that are due to faulty construction ?