Global view of optical design space

Doron Sturlesi
Donald C. O'Shea, FELLOW SPIE
Georgia Institute of Technology
School of Physics
Center of Optical Science
and Engineering
Atlanta, Georgia 30332-0430

Abstract. The optical design space of some simple lenses is investigated systematically. Typical space topographies are visualized with 3-D graphics, where the complete set of available solutions is clearly identified. The space characteristics are then studied and compared through the use of several merit functions with differing degrees of complexity. A two-phase search algorithm, based on global optimization techniques, is proposed here. In the first phase, using a coarse sampling approach, the program finds the favorable regions that correspond to potentially promising configurations. In the second phase, conventional optimization routines are used to find the best solutions in each region. Then an optimum solution is determined according to the application at hand. The proposed algorithm is analyzed and compared to more conventional design approaches. A further refinement of the algorithm excludes from the systematic search some unfavorable configuration regions through the use of a simple expert system. Search times are further reduced through parallel-processing methods. The algorithm provides overall information about a given design space and offers a selection of "best" solutions to choose from. As an example, it is applied to a triplet objective.

Subject terms: optical design; lens design; parameter space; global optimization; sampling techniques; expert systems; parallel processing; Cooke triplet.

Optical Engineering 30(2), 207-218 (February 1991).

CONTENTS

1. INTRODUCTION

The design space of optical systems is typically a complicated multidimensional parameter space. By constructing a merit function space that expresses the departure of individual configurations from ideal required performance, it is possible to determine which of all the possible configurations in this space will yield the best solution to the design problem. One of the known characteristics[1] of the merit function space is the large number of stable configurations that are contained therein. These configurations correspond to a large number of local minima (of a merit function defined over the space), many of them showing comparable performance.

In a conventional design process, an optical designer chooses an initial configuration—a point in the design space—and with the help of an automatic optimization algorithm moves it around in the design space to find a solution that gives the best, or at least a good workable performance. The search for the "ultimate best," or "global optimum," is perhaps the most difficult part in the lens design art. Even the best optimization routines cannot, in general, find it. Most of the popular optimization routines like damped least squares (DLS), and others related to it, are based on systematic descent principles that accept only steps which decrease the merit function (improve the performance). Thus the designer's solution almost always gets trapped in a local minimum.[2]

The limited success of the conventional design methods in reaching the global minimum is mainly due to the fact that the search is confined to narrow corridors along the (downhill) optimization paths. The solution is therefore crucially dependent on the initial configuration. As a result, conventional design methods work satisfactorily and converge quickly to a good solution only if a good starting point has been chosen. Most often, the routine ends up in an unfavorable configuration that does not satisfy the design requirements. It is not unusual for this to happen several times during the first design sessions.

The conventional design methods are, in principle, local search methods and do not provide any global information on the design space. The only exception is simulated annealing (SA). This method, which was recently revived from an older statistical cooling algorithm,[3] is the first global optimization algorithm to be applied to optical design. This is a stochastic search algorithm that follows the principles of the annealing process, where the goal is to release residual stresses and to bring a substrate to a

Paper 2857 received Jan. 2, 1990; revised manuscript received July 26, 1990; accepted for publication Aug. 7, 1990. This paper is a revision of paper 1168-13, presented at the SPIE conference Current Developments in Optical Engineering and Commercial Optics, Aug. 7, 10–11, 1989, San Diego, Calif. The paper presented there appears (unrefereed) in SPIE Proc. 1168.

certain finite probability, and thus can theoretically get away from local traps. But even this algorithm, according to results published recently,[4–8] seldom converges to the global minimum, since it is too sensitive to the tuning of its internal constants and becomes increasingly inefficient as the number of degrees of freedom rises. There are efforts to overcome the poor internal constant tuning by dynamic compensation. Thus a number of approaches are being used to attack the problem of the global optimum in lens design. This paper represents a report on some initial results using another technique.

2. GLOBAL OPTIMIZATION

Global optimization (GO) is a separate area in mathematical optimization. Although the fundamental problem is not new, most of the work in this area had been done in the past two decades. Because of the availability of powerful computers, a considerable number of algorithms have been introduced as part of the general effort to solve the GO problem.[9,10] These minimum-seeking algorithms were designed to deal with multiminima spaces and are, in principle, different from the local optimization algorithms.

Since the work in this area is still going on, it is obvious that there is no general deterministic algorithm that can locate global optimum (or minimum) for every general multidimensional function.[9–11] The usefulness of the algorithms depends crucially on the type of (objective) function that forms the merit function space. The most difficult situation, which is unfortunately typical to many engineering problems, is the ''black box'' situation,[12,13] where the objective function (merit function), which is multidimensional, cannot be expressed in closed form and its evaluation requires massive numerical computation. As such, function evaluation, which is the only way to get information on the design space, is a time-consuming operation. Optical design is an example of just such a case.

The problem can be stated most generally as follows; Let R^n be the n-dimensional design space. A point in that space $x \in R^n$ is characterized by the vector, $x = (x_1, x_2, x_3, \ldots, x_n)$ where $\{x_1, x_2, x_3, \ldots, x_n\}$ are the set of coordinates or design parameter values that specifies the configuration x. The merit function is defined by $F(x)$, where $x \in R^n$ and $F: R^n \rightarrow R^1$ over a compact set $S \subset R^n$. The GO problem is to find $y^* = \min_{x \in S}\{F(x)\}$, for *every* $x \in S$.[13]

This problem is one of a class of NP-hard problems. This is a class for which no algorithm is known to give an exact solution within polynomial time (i.e., the computation time increases at least exponentially with the complexity of the problem).[3]

The simplest global search method is a systematic sampling of the function on a multidimensional grid. This deterministic method was originally called the factorial[12] method. In this method each design parameter (factor) is divided to a number of levels, and for each combination of levels of the set of the design parameters a sampling is performed. Sampling, in this context, means simply evaluation of the merit function. There are some modifications to this method, such as the fractional factorial method,[14] in which a systematic deletion of some parameter combinations is performed before the function evaluation begins. The main problem with this systematic sampling method is that the number of function evaluations increases exponentially with space dimension which sets some practical dimension limits to this method.

Probabilistic approaches have been introduced to GO in order methods there are two phases:[13] a global phase, during which the function is evaluated at randomly sampled points, and a local phase, during which the sample points are manipulated by means of some local searches to yield a candidate global minimum. The major drawback of the stochastic method is that the possibility of an absolute guarantee of success is sacrificed in favor of limiting the effort. The global phase can, however, yield an asymptotic guarantee in the stochastic sense.

Most successful methods for GO involve local searches from some or all of the sample points. This presupposes the availability of some local search (LS) procedure. LS is assumed to be strictly descendent,[13] such that if LS is started from any point in the $x \in S$ and converges to a local minimum x^*, there exists a path from x to x^* along which the function values are nonincreasing. A common feature in GO methods is domain partitioning.[15] This is basically a grid sampling of the design space that creates a collection of cells that can be analyzed later by LS procedures.

The efficiency of GO algorithms can, in principle, be measured by the probability of a specific algorithm to find the global minimum in a certain number of steps. However, in practice, the convergence of a certain algorithm is highly dependent on a number of external factors such as the properties of the objective function, the dimension of the space, the programming approach, and hardware characteristics. So it is difficult to construct general objective tests for the purpose of comparing methods, even if we were to attempt to apply them to a standard set of test functions.[10]

Although the GO methods are commonly classified to deterministic/probabilistic classes, a more profitable approach is constructed by combining elements of these two classes. In general, a pure deterministic approach has advantages up to certain design space dimensions, where the time required to perform a systematic search is reasonable. Above that dimension, introduction of some probabilistic elements is necessary to overcome the exponential growth in the number of function evaluations,[9,10] as noted above.

Another interesting feature of the optical design space is that the space dimension is itself a design parameter. Thus, the ideal search algorithm should allow for the addition and subtraction of surfaces and lenses dynamically during the optimization process, according to some previously stated criteria. To our best knowledge, the only attempt to approach this ''dynamical space'' property was recently reported in Refs. 16 and 17. In this work, a change of dimension from R^n to R^m $(m>n)$ occurs automatically if the algorithm cannot reach a certain figure of merit within R^n. This automatic dimension change is done by a special procedure called a sequential cluster algorithm.

3. THE OPTICAL DESIGN SPACE

The above analysis of GO gives rise to a fundamental postulate regarding the successful application of global search to the optical design procedure (and, in fact, to any other applied problem). To apply GO techniques efficiently, there needs to be an investigation of the general properties of the design space and a formulation of its basic characteristics.[18] With this information in hand GO techniques can be applied more efficiently. Consequently, we may expect considerable improvement in the search for the ultimate (optical) design solution. The basic topics for the investigation are (1) the type of merit function to be used, (2) the existence of discontinuities, and (3) the topographical

The design space, or parameter space, is a multidimensional space over which a single merit function is defined in terms of the degrees of freedom that are being made available to the specific design task. The merit function is a combination (usually sum of squares) of departures from the criteria that are considered important characteristics for a specific design task and that are to be minimized during the design process. A point in this space represents, therefore, a specific configuration of the optical system; and the merit function, a measure of its deviation from the required optomechanical performance. To this design space is added one more dimension, the merit function. It is this space consisting of n independent variables representing the design parameters and one dependent variable, the merit function, that is explored in the optimization process. A configuration is mathematically stable, for small changes in the degrees of freedom values, at the local minimum of this merit function space. However, the local minimum may not necessarily be acceptable as a good enough solution to the design task considered. A practical definition of a minimum in terms of optical design is a (stable) configuration that cannot further be improved by conventional optimization (i.e., DLS, etc.).

Typically the merit function is not an analytical function, and it is not expressible in closed form in terms of its independent variables. Its components may have been obtained from the various orders of the geometrical aberration theory or from numerical ray-tracing calculation. The ultimate goal of the design is to find the optimum among all the minima, which is the global optimum.

This design space is the playground of the optical designer. The correct construction and understanding of this space is essential for a successful design task. What follows is a report on investigations of the properties of this design space to provide an understanding of how one might attack the problem of finding the global optimum.

3.1. Merit function spaces

The merit function depends differently on the different design parameters and on the numerical scaling of the essentially nonmetrical space. For example, lens curvatures, as a class, are usually more effective than the surface separations in manipulating this function. An artificial metric is therefore essential in ''balancing'' the design space to unify the dynamic range of these numerical values. Even so, some dominant parameters always exist. As a result, the design parameters can be sorted according to their effect on the merit function. This property will be used in the sampling algorithm to be described here.

We can only visualize a three-dimensional space, sometimes resorting to stereo-viewing techniques to assist us. Multidimensional spaces are beyond our abilities to depict. However, a good deal of information can be obtained by looking at three-dimensional slices of these spaces. We can use a two-dimensional design space to visualize the distribution of our configurations, reserving the third dimension to represent the merit function value at each point. With today's computer graphics this space can be drawn in perspective on a sheet of paper or a computer monitor. In higher dimension spaces, we can represent the essential physics of the system by selecting the most significant pair of parameters to be plotted and freezing all others at some intermediate value. In this manner, we can develop an intuition for cases having three, four, and more degrees of freedom.

to play in the success or failure of a design task. We have chosen to highlight our approach by using three types of such merit functions, which we believe to be representative of three major types: an aberration-based merit function, a selective rays merit function, and a full-beam-analysis merit function.

The easiest of these to evaluate is one that seeks to minimize the third-order Seidel and color aberrations:

$$MF = \left[\sum_{i=1}^{5} (S_i)^2 + \sum_{j=1}^{2} (C_j)^2 \right]^{1/2} , \qquad (1)$$

where S_i = spherical aberration, coma, astigmatism, Petzval curvature, and distortion ($i = 1$ to 5), respectively, and C_1 = longitudinal color and C_2 = transverse color. The Seidel aberration coefficients,[19] represent the third-order terms in the development of the aberration function of a rotationally symmetric optical system. They can be weighted, of course, when the sum of squares is formed, to reflect the designer's view of a particular task, enhancing or excluding some of them.

A merit function, derived from the exact trace of a pair of meridional rays, can be used to provide a more realistic evaluation of the system:

$$MF = \left[\sum_{i=1}^{6} (A_i)^2 \right]^{1/2} , \qquad (2)$$

where A_i = spherical aberration, offense against the sine condition (OSC), Conrady color, sagittal and tangential curvatures (Coddington), and distortion ($i = 1$ to 6), respectively.

All that is required is a marginal ray passing through the rim of the entrance pupil (or a fraction of it) and a chief ray starting from the extreme point of the object (or, again, from a fraction of it). In this way a finite spherical aberration, the offense against the sine condition (OSC), the color according to Conrady (*d-D* method), the astigmatism (tangential and sagittal, using a Coddington trace), and a finite distortion measure can be included, weighted to reflect their relative importance.[20]

A third merit function uses the blur spot size for three points: an axial point, a 70% field point, and a point at full field:

$$MF = \left[\sum_{i=1}^{3} (SP_i)^2 \right]^{1/2} , \qquad (3)$$

where SP_i = polychromatic blur spot radius at three fields of view, corresponding to $i = 1$ to 3. The mean blur spot is measured as the root mean square of ray hits around their center of gravity obtained from a fully polychromatic (3 colors) exact ray trace, at three selected field points, at its best position along the optical axis. The number of rays traced per field point per color depends on the lens type and varies between 15 and 100.

3.2. Merit function space examples

With these three merit functions as tools, we chose two simple lenses to explore techniques for searching for the global optimum.

3.2.1. Cemented doublet (CD)

A crown-first, cemented, two-glass, *f*/4 achromat, working at a 10° full field and infinite object distance, was investigated and

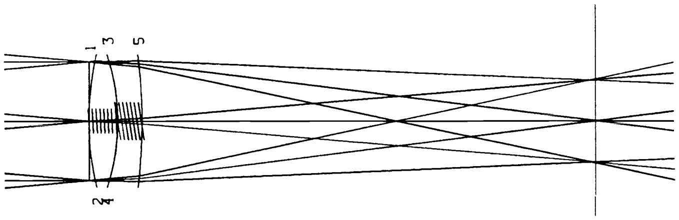

Fig. 1. Cemented doublet, the first example studied here.

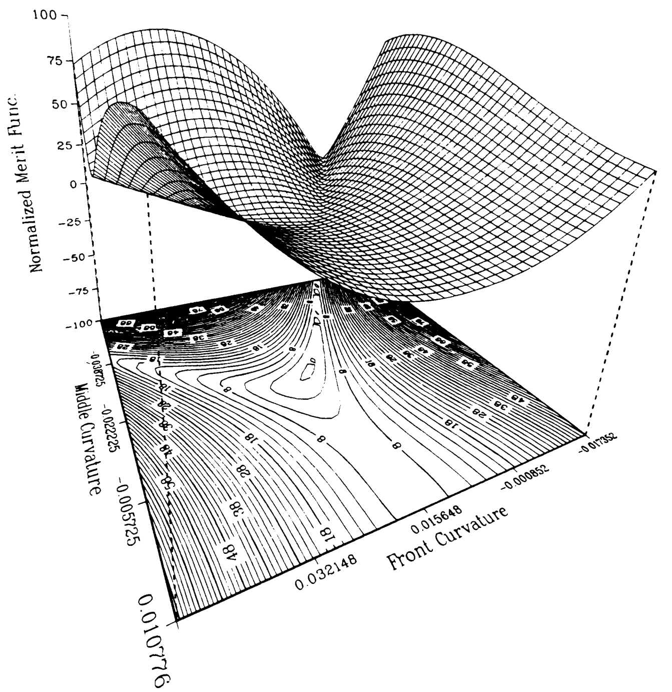

Fig. 2. Cemented doublet design space, with Seidel-aberrations merit function.

dimensional design space, if the four curvatures only are considered as free degrees of freedom. The last surface curvature (C_5) is solved to maintain a constant effective focal length (EFL); the design parameters are C_2,C_3 (C_1 is the stop and $C_4 = C_3$). The two component thicknesses were kept constant and the image distance was adjusted to the ''best'' average blur spot position. Figures 2 through 4 show the merit function spaces of the three merit functions described above.

It can clearly be seen, for all three different merit functions, that within the permitted region of the variation of the variables (C_2,C_3), the landscape is fairly smooth and there is a well-defined minimum that corresponds to the single optimum solution.

3.2.2. Air-spaced doublet (ASD)

This crown-first, *f*/5, infinite-conjugate, air-spaced achromatic doublet with a 2° full field was extensively explored. The space was found to contain at least nine discrete stable configurations, shown in Fig. 5. Two of these are traditionally identified as Fraunhofer (A) and Gauss (D) configurations,[4,21–23] bearing the names of their creators. In the ASD case, if we let only the curvatures vary and solve for the last one (C_5) to keep the EFL at a constant value, we need three dimensions to specify all possible configurations and four to represent it: C_2, C_3, C_4, and

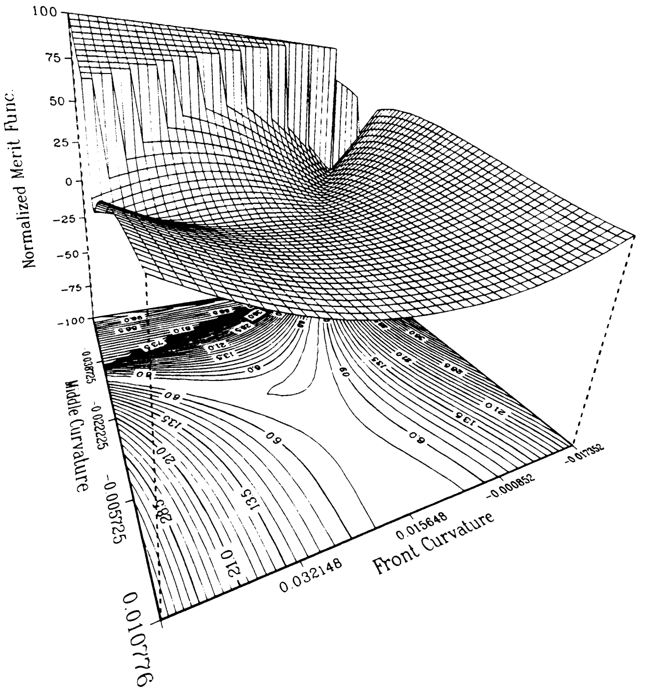

Fig. 3. Cemented doublet design space, with finite-ray-aberrations merit function.

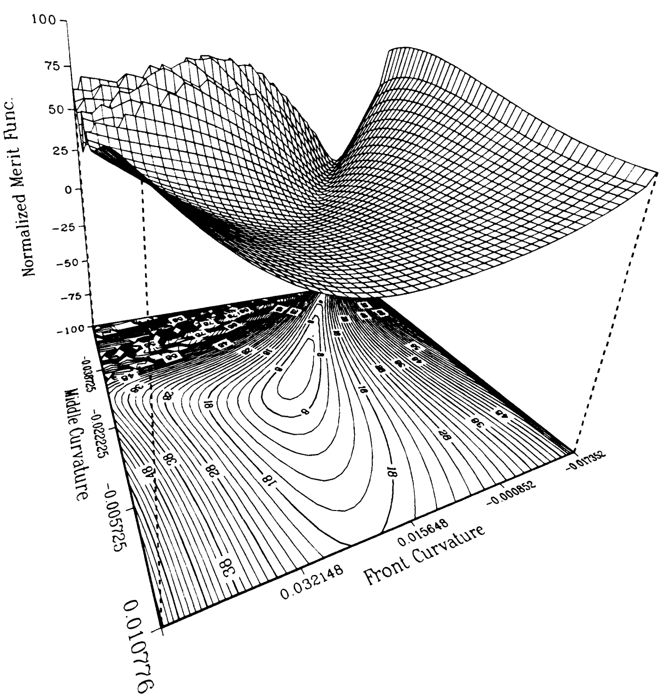

Fig. 4. Cemented doublet design space, with three object points, polychromatic blur-spot merit function.

first curvature (C_2) is the least significant in terms of its effect on the merit functions used, so that an instructive general view can be achieved by projecting the design space on the two remaining dimensions, C_3 and C_4, and keeping C_2 at an intermediate constant value. Part of the merit function spaces so obtained, one for each type of the three merit functions, are shown in Figs. 6 through 8. By inspection, and by following

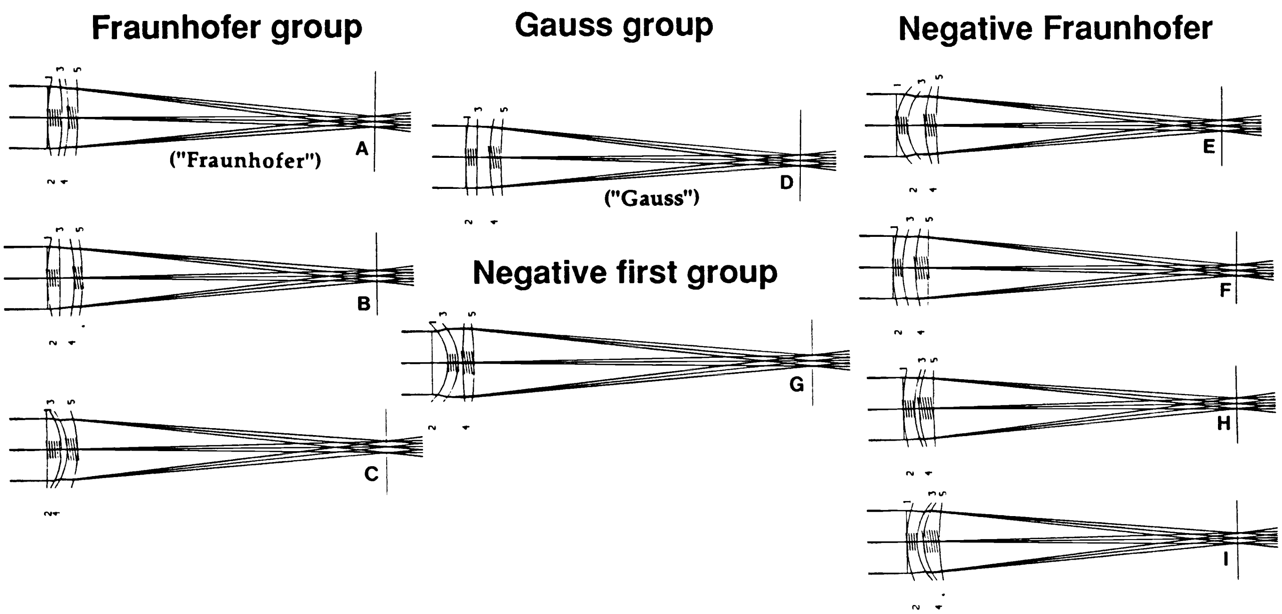

Fig. 5. Nine different configurations for the air-spaced doublet design space.

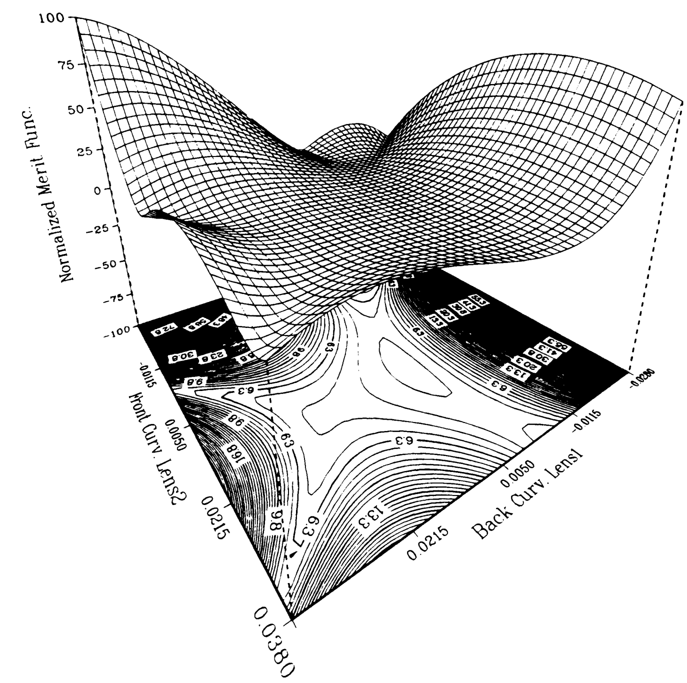

Fig. 6. Air-spaced doublet design space, with Seidel-aberrations merit function.

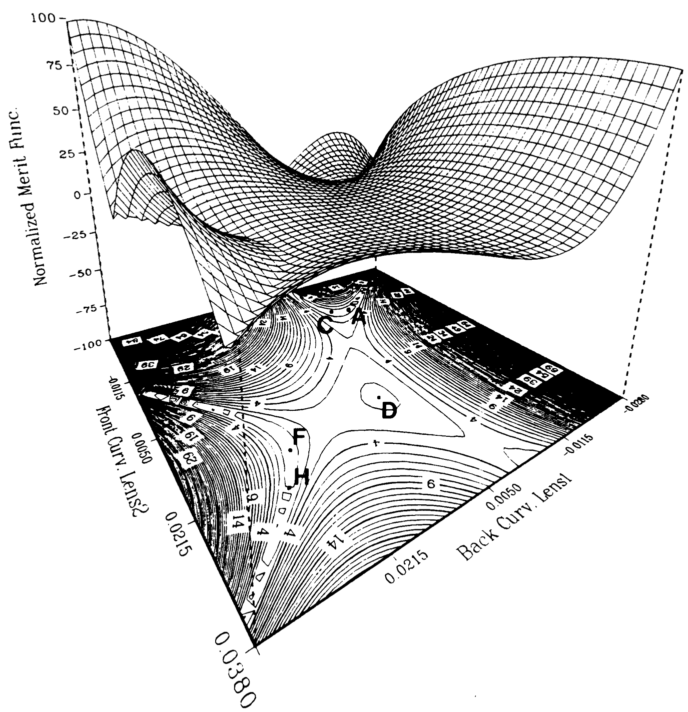

Fig. 7. Air-spaced doublet design space, with finite-ray-aberrations merit function. The position of five of the configurations in Fig. 5 are shown here distributed in three solution regions: Fraunhofer (A and C), Gauss (D) and negative Fraunhofer (F and H).

an overall view of the ASD design domain, a view that is not obtainable through local analysis, which is used by the current design methods. Five of the configurations mentioned above, namely A,C,D,F, and H, are seen in Fig. 7, distributed in the shallow area. The whole solution set can be classified into four major groups, each group occupying a different shallow valley, as follows:

Fraunhofer group: Configurations A,B,C;
Gauss group: Configuration D;
Negative Fraunhofer group: Configurations E,F,H,I;

The nine are shown and compared for performance and sensitivity in Table 1. The column labeled Sensitivity provides an objective evaluation of the lenses on another criterion. All of the configurations are evaluated to determine the RMS of the partial derivatives of the merit function with respect to each of the design parameters at the minimum. This provides some assessment of the lens to manufacturing tolerances. (The inclusion of tilt and decenter in the sensitivity analysis did not change the

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase®
Smarter legal research.