The CECOM Center for Night Vision and Electro-Optics

OPTOELECTRONIC WORKSHOPS

XXIII

OPTICAL SYSTEM ASSESSMENT FOR DESIGN AND SIMULATED ANNEALING

August 2, 1990

sponsored jointly by

ARO-URI Center for Opto-Electronic Systems Research
The Institute of Optics, University of Rochester

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Optoelectronic Workshop XXIII: Optical System Assessment for Design and Simulated Annealing

5. FUNDING NUMBERS

6. AUTHOR(S)
Gregory Forbes (UR) and Robert Spande (CCNVEO)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Institute of Optics
University of Rochester
Rochester, NY 14627

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This workshop on "Optical System Assessment for Design and Simulated Annealing" represents the twenty-third of a series of intensive academic / government interactions in the field of advanced electro-optics, as part of the Army sponsored University Research Initiative. By documenting the associated technology status and dialogue it is hoped that this baseline will serve all interested parties towards providing a solution to high priority Army requirements. Responsible for program and program execution are Dr. Nicholas George, University of Rochester (ARO-URI), and Dr. Rudolf Buser, CCNVEO.

14. SUBJECT TERMS
optical system; assessment; annealing; simulated annealing

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

The CECOM Center for Night Vision and Electro-Optics

OPTOELECTRONIC WORKSHOPS

XXIII

OPTICAL SYSTEM ASSESSMENT FOR DESIGN AND SIMULATED ANNEALING

August 2, 1990

sponsored jointly by

ARO-URI Center for Opto-Electronic Systems Research
The Institute of Optics, University of Rochester

OPTOELECTRONIC WORKSHOP

ON

OPTICAL SYSTEM ASSESSMENT FOR

DESIGN AND SIMULATED ANNEALING

Organizer: ARO-URI -University of Rochester and CECOM Center for Night Vision and Electro-Optics

1. INTRODUCTION

2. SUMMARY -- INCLUDING FOLLOW-UP

3. VIEWGRAPH PRESENTATIONS

   CECOM Center for Night Vision and Electro-Optics
   Organizer -- Robert Spande

   ARO-URI Center for Opto-Electronic Systems Research
   Organizer -- Gregory Forbes

   Overview
   Robert Spande, CCNVEO

   Towards Global Optimization with Adaptive Simulated Annealing
   Gregory Forbes and Andrew Jones, ARO-URI

   Lens Design with Adaptive Simulated Annealing
   Andrew Jones, ARO-URI

4. REFERENCES

5. LIST OF ATTENDEES

1. INTRODUCTION

This workshop on "Optical System Assessment for Design and Simulated Annealing" represents the twenty-third of a series of intensive academic / government interactions in the field of advanced electro-optics, as part of the Army sponsored University Research Initiative. By documenting the associated technology status and dialogue it is hoped that this baseline will serve all interested parties towards providing a solution to high priority Army requirements. Responsible for program and program execution are Dr. Nicholas George, University of Rochester (ARO-URI), and Dr. Rudolf Buser, CCNVEO.

2. SUMMARY AND FOLLOW-UP

The meeting began with a brief presentation by Bob Spande of the Night Vision Laboratory who described a variety of optical systems that his group was involved in designing. These included a number of many-element systems and systems which incorporate diffractive optics and gradient index elements. Since the design of such systems presents entirely new challenges, it was felt that global optimization may have the most to offer in such contexts on account of the paucity of "rules of thumb" to guide the designer in finding appropriate starting points for the regular, local optimization algorithms.

The second and third presentations were given by Greg Forbes and Andrew Jones. After illustrating the impracticality of an algorithm that is guaranteed to find global the minimum for problems such as the design of non-trivial optical systems, Forbes described the ideas behind the simulated annealing algorithm. He also presented the concepts introduced in order to automate the temperature control and step generation components of such an algorithm in an adaptive fashion. Jones described some of the principal difficulties that must be addressed in applying this algorithm in the context of lens design. The most important of these being the problem of dealing with the issues of constraint enforcement and non-linear transformations of the coordinates and merit function. This presentation was concluded with a report on applications of this adaptive simulated annealing algorithm to a number of optical design problems.

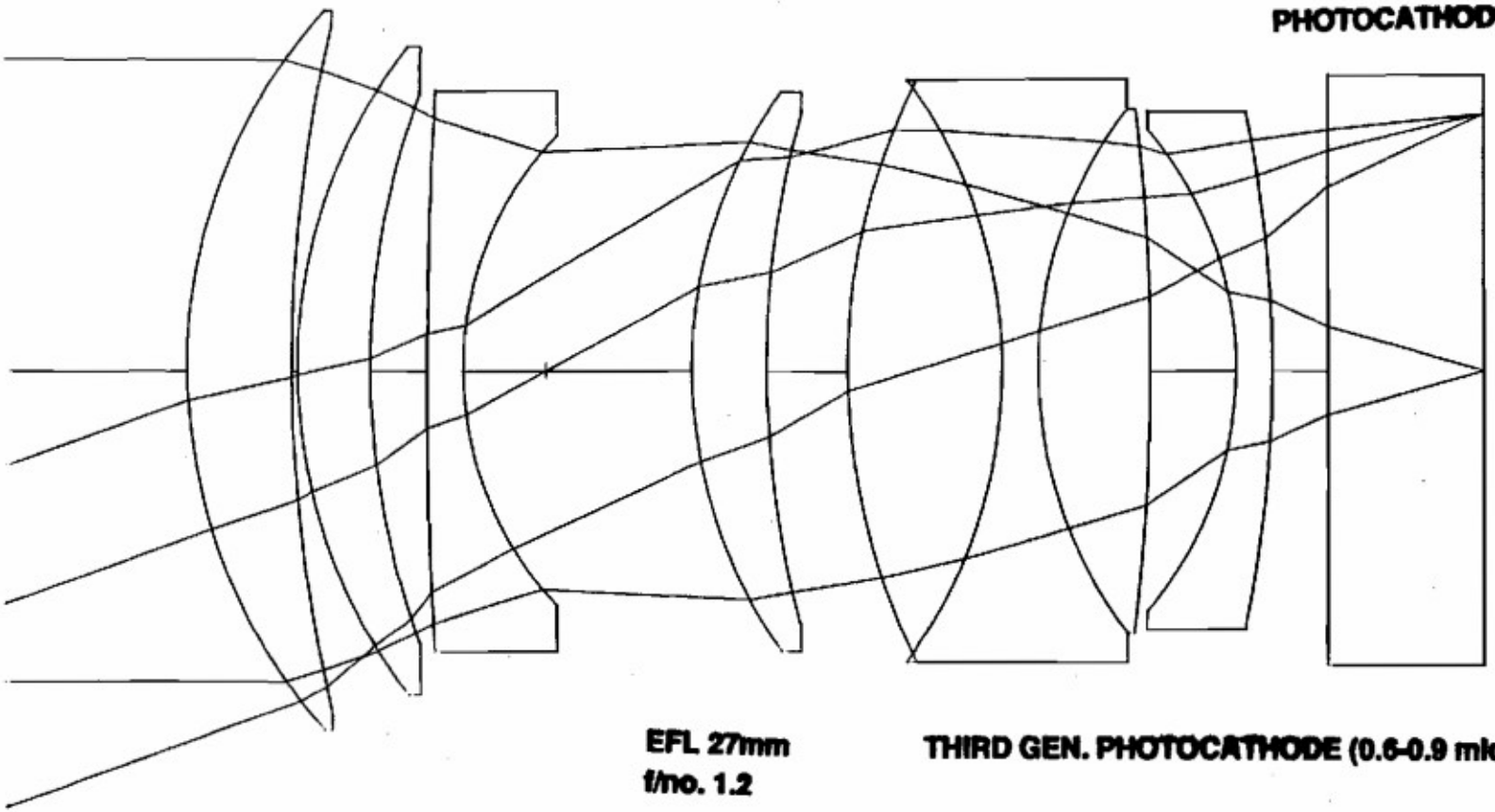
The workshop ended following a discussion of potential joint projects and collaboration. It was decided that it would be of benefit and interest to all concerned if the adaptive simulated algorithm and dedicated hardware at The Institute of Optics be applied to a variety of optical design problems supplied by NVL for comparison with their proposed systems.

WORKSHOP ON GLOBAL OPTIMIZATION
WITH
ADAPTIVE SIMULATE ANNEALING
2 AUGUST 1990
AGENDA

INTRODUCTION	**Robert Spande, C2NVEO**
OPTICAL APPLICATIONS FOR NIGHT VISION EQUIPMENT	**Robert Spande, C2NVEO**
GLOBAL OPTIMIZATION AT UR	**Greg Forbes** **Andrew Jones**
DISCUSSION	**ALL**
CLOSING REMARKS	**Robert Spande, C2NVEO**

CECOM CENTER FOR NIGHT VISION AND ELECTRO-OPTICS
OPTICAL APPLICATIONS FOR NIGHT VISION EQUIPMENT

ANVIS OBJECTIVE



EFL 27mm
f/no. 1.2

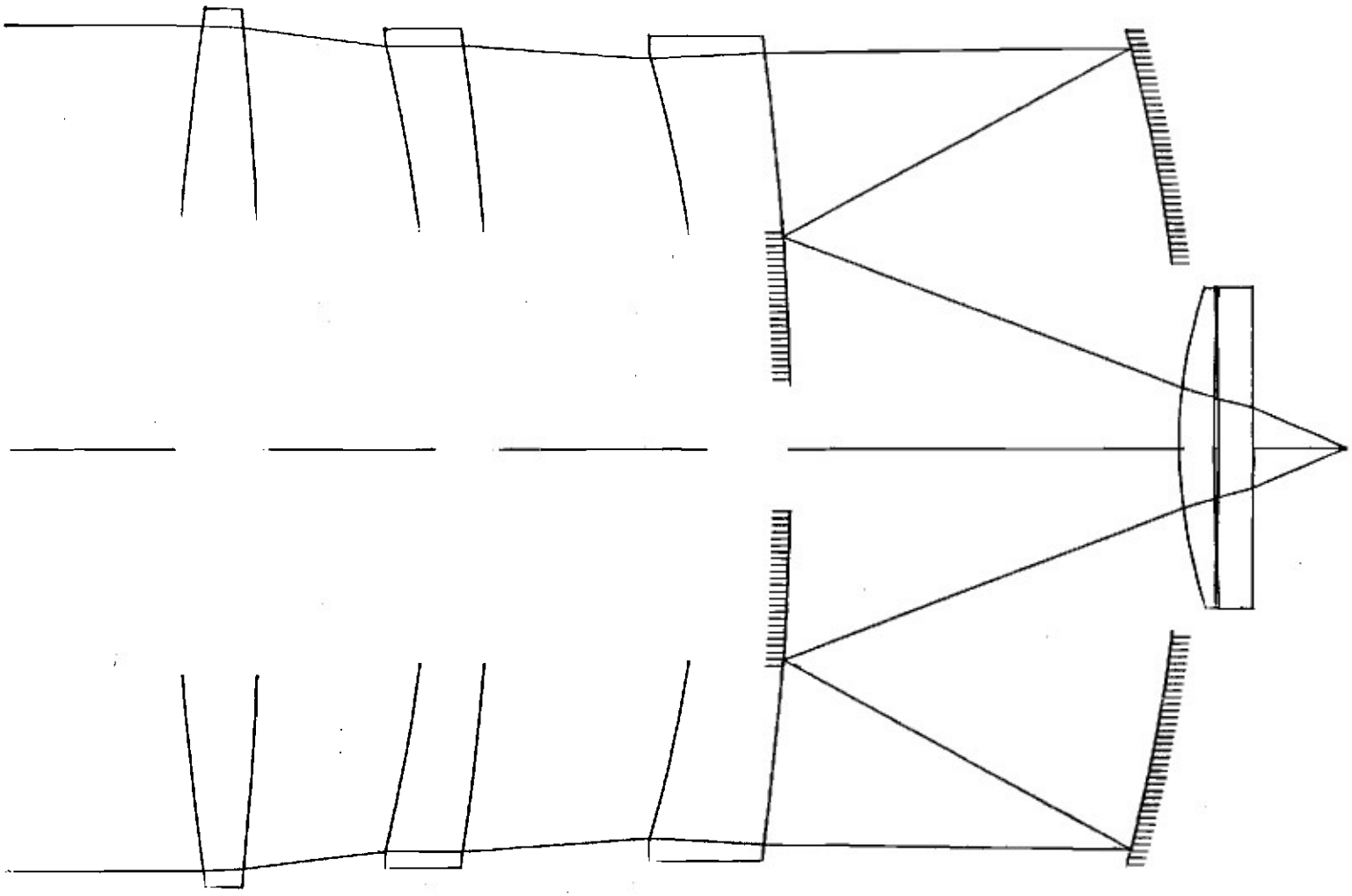
FOV 40 DEGREES

THIRD GEN. PHOTOCATHODE (0.6-0.9 microns)

6.25 MM

ANVIS OBJECTIVE SCALE 4.0 13-FEB-91

AN/TVS-5 OBJECTIVE

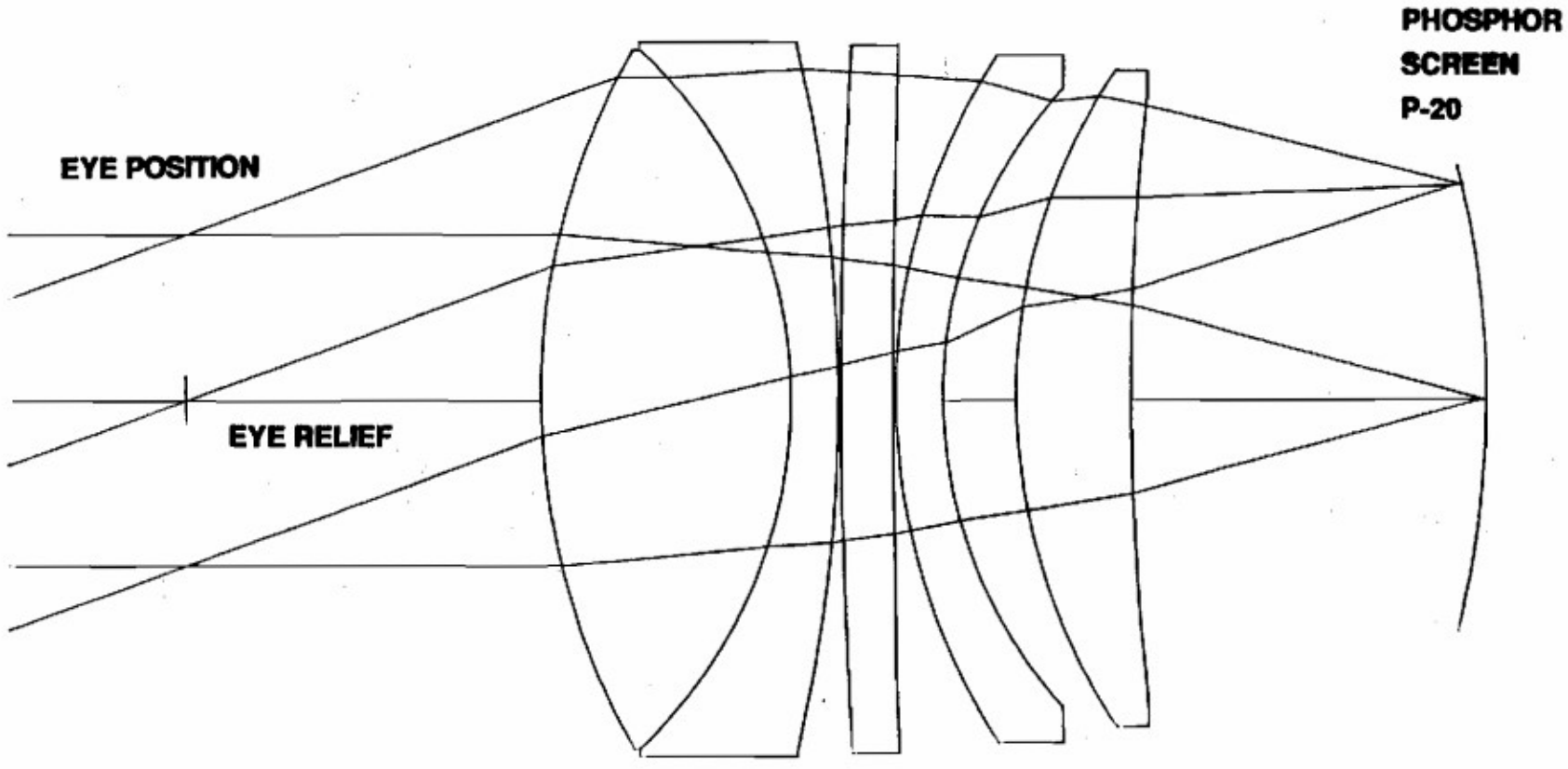


EFL 155mm
f/no. 1.2

FOV 9 Degrees

25.0 MM

AN/TVS-5 OBJECTIVE LENS ASSEMBLY FULL SCALE RAS 13-FEB-91

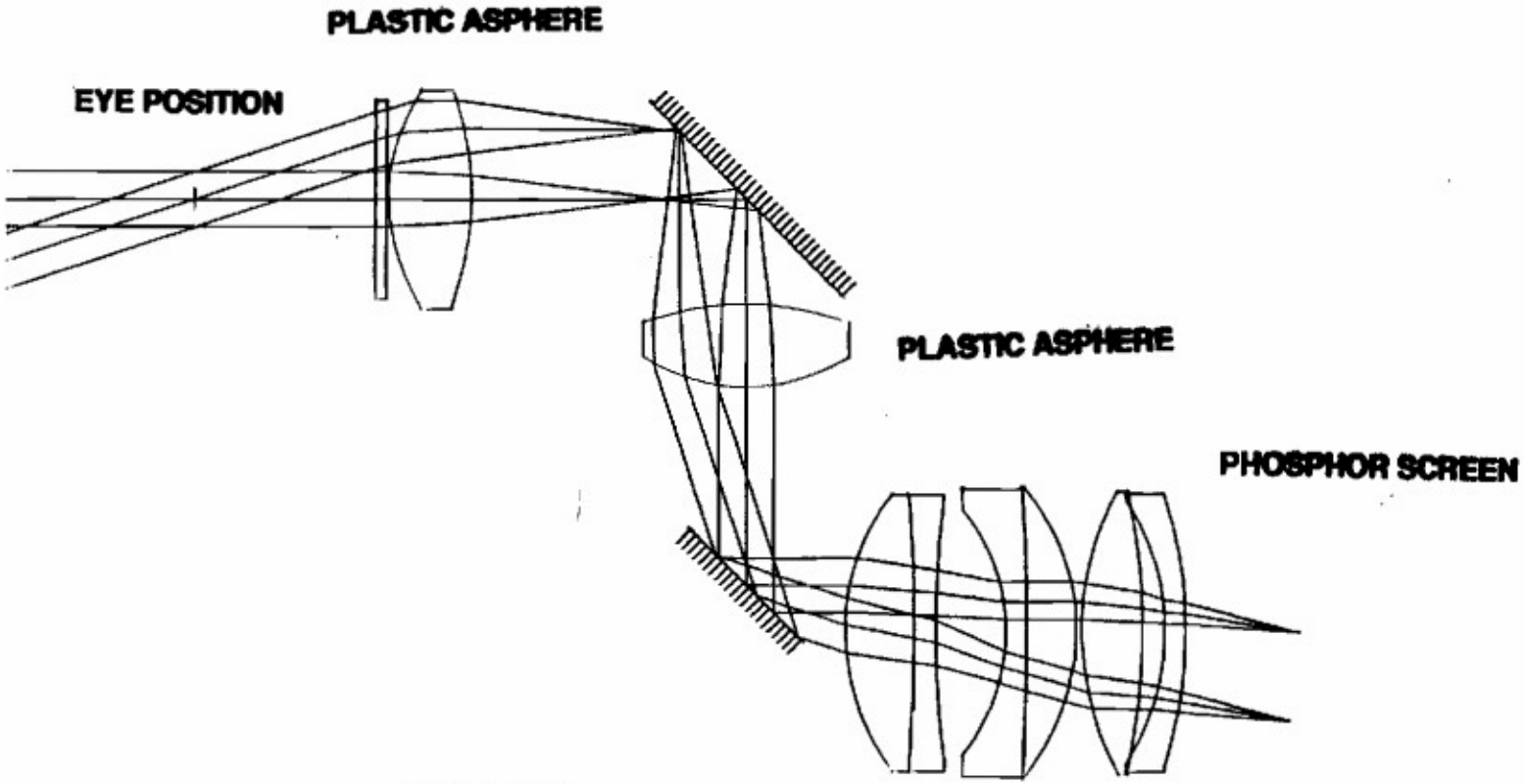


POWER 10X
EYE RELIEF 15mm
EYE PUPIL 10mm

7.35 MM

GOGGLE EYEPIECE

AN/PVS7B EYEPIECE

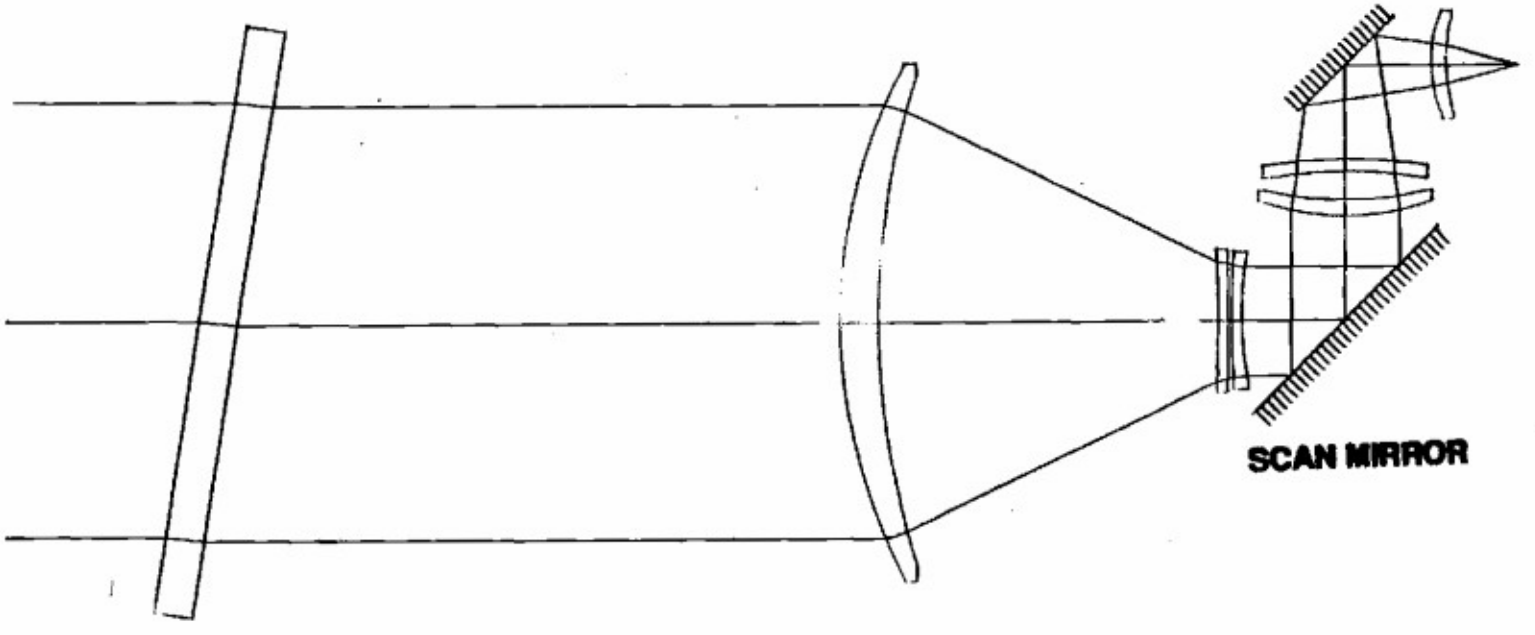


AN/PVS7B EYEPIECE SCALE 1.5 13-FEB-91

FIRST GENERATION FLIR

IMAGERY: SCANNER PANS SCENE ACROSS LINEAR ARRAY

COLD SHIELD ON LINEAR DETECTOR ARRAY



EFL 10.64 INCHES
f/no. 1.9
FOV 5 x 2.5 Degrees

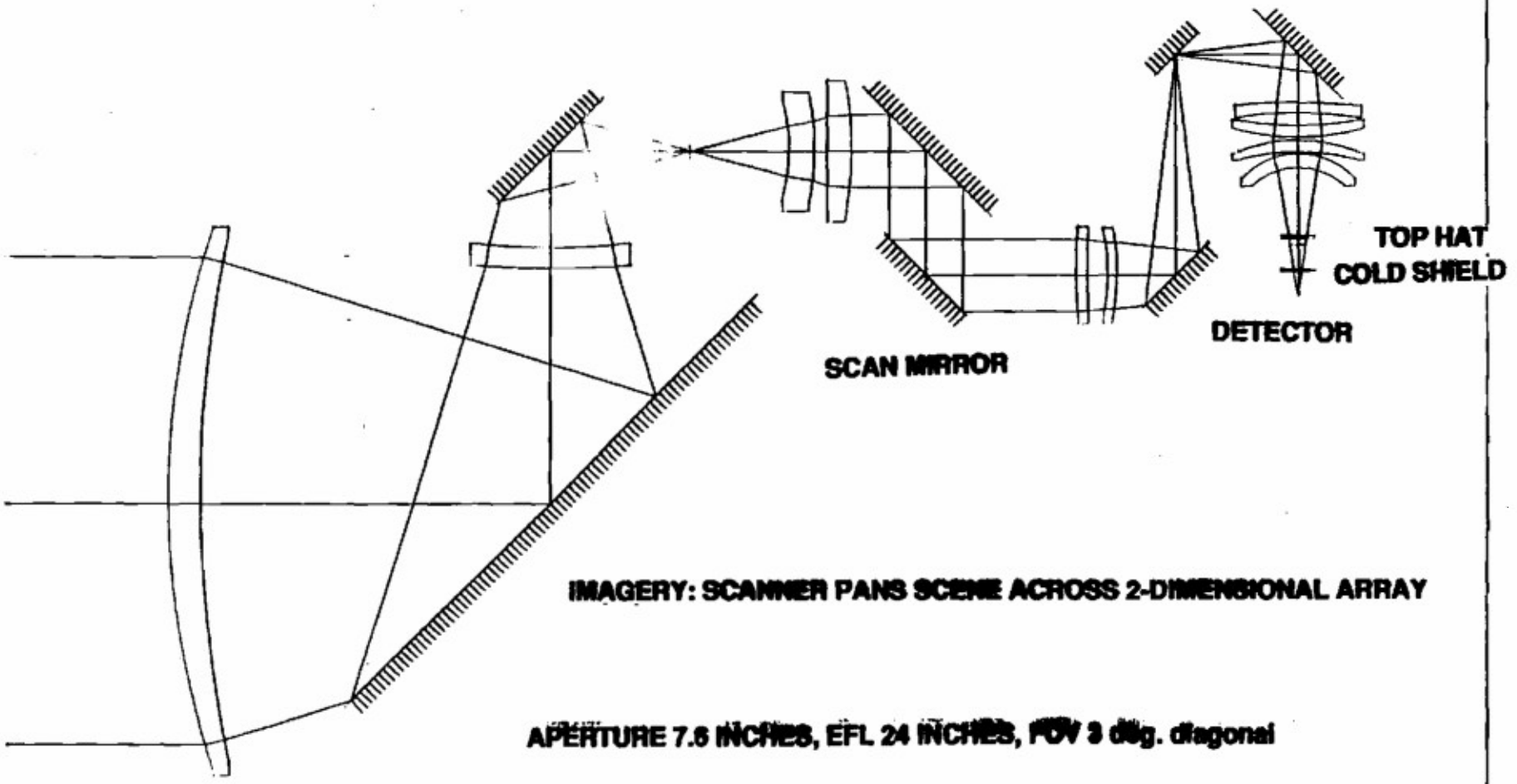
2.50 IN

FIRST GENERATION FLIR

POSITION 1
SCALE 0.400

MRG 13-FEB-91

SECOND GENERATION FLIR

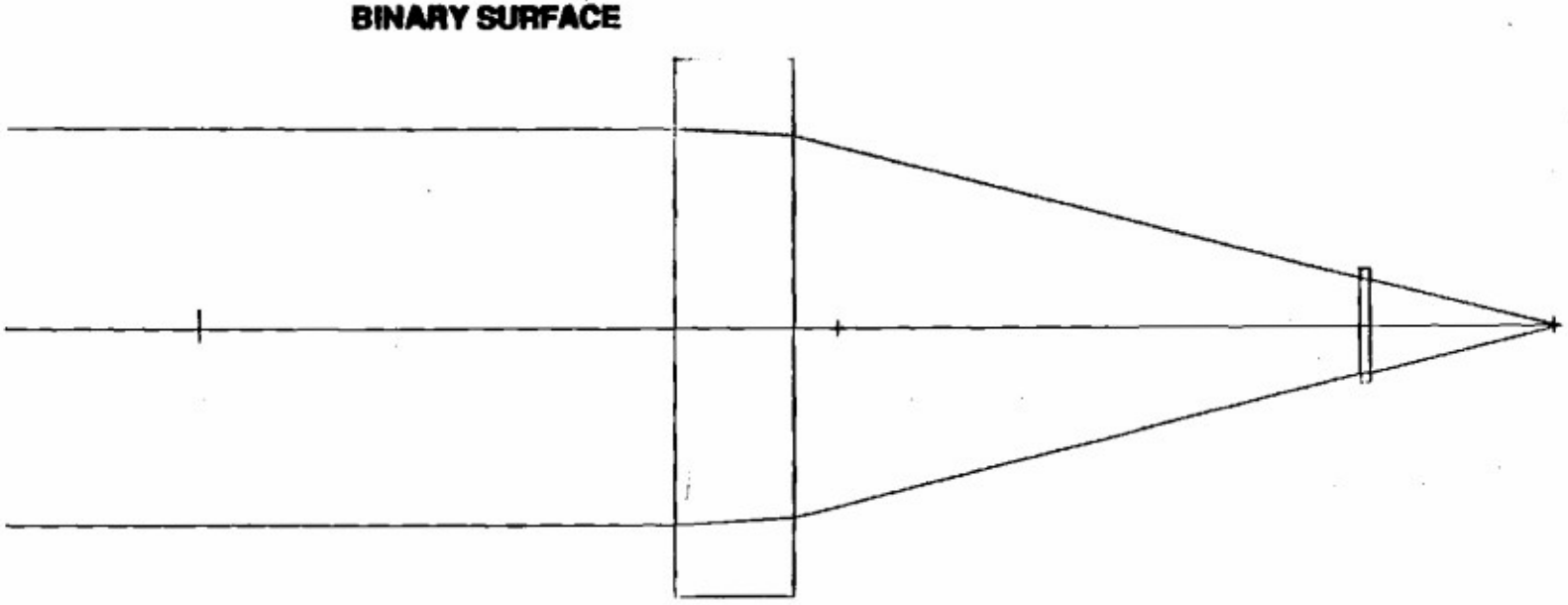


SECOND GENERATION FLIR

POSITION 1
SCALE 0.400

MRG 13-FEB-91

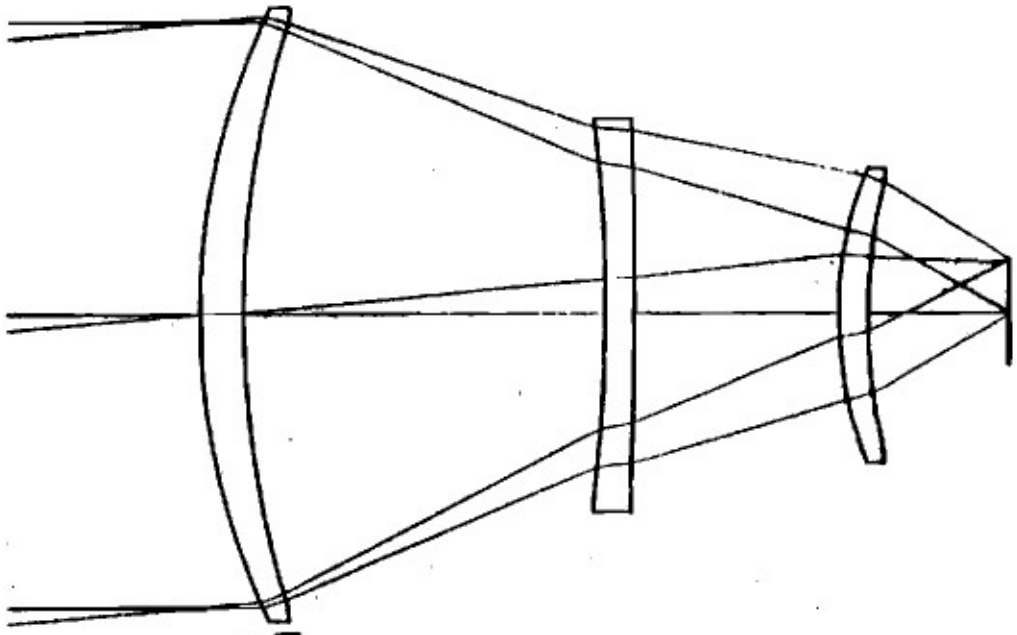
BINARY OPTICAL ELEMENT



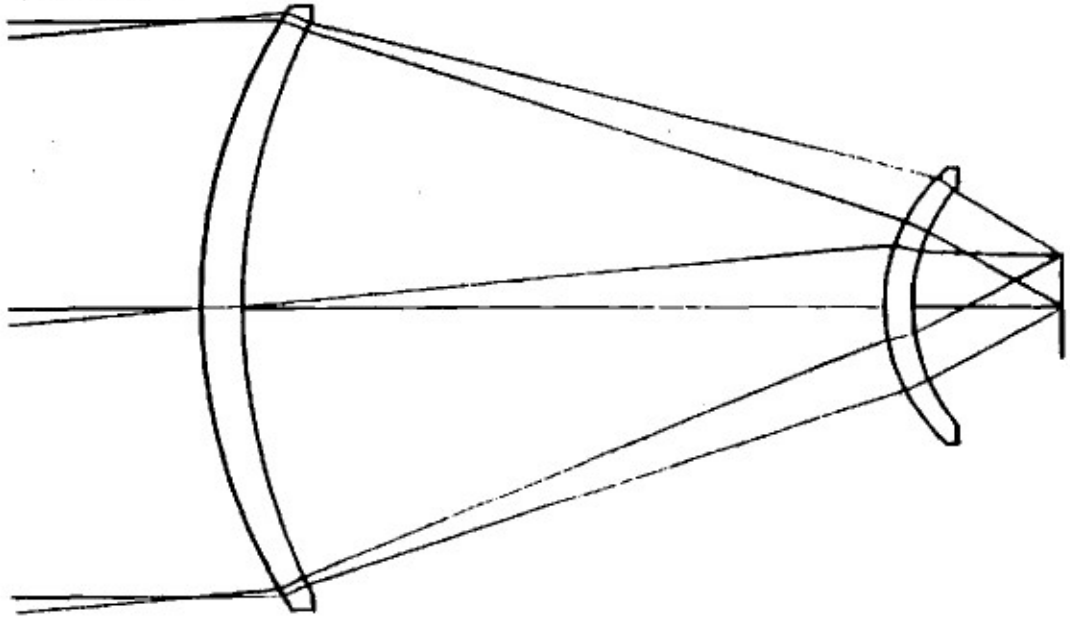
BINARY OPTICAL ELEMENT SCALE 2.5 13-FEB-91

ADVANCED OPTICS RESEARCH
OPTICS DESIGN
BINARY OPTICS

TRIPLET
3" APERTURE, f/1.0
10 DEGREE FOV
COST $1000

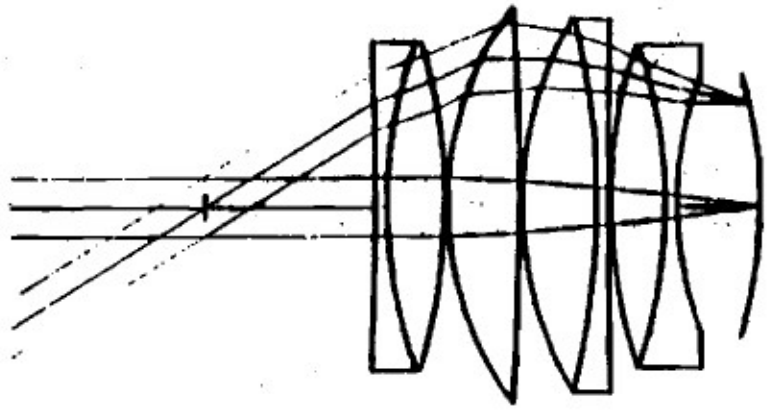


BINARY DOUBLET
3" APERTURE, f/1.0
10 DEGREE FOV
COST $700

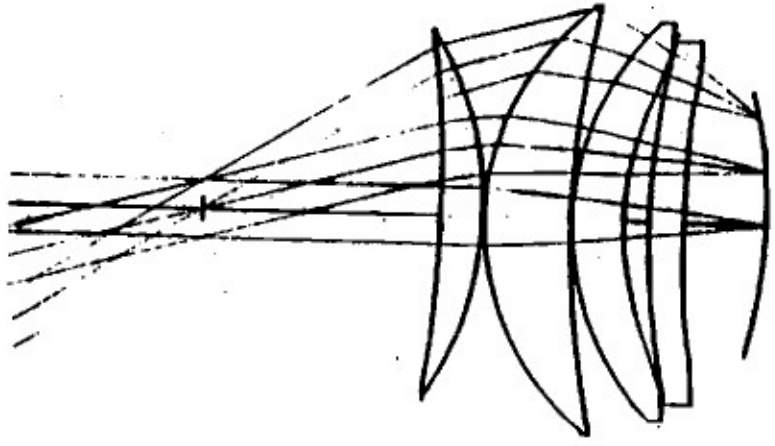


BINARY DISPLAY EYEPIECE

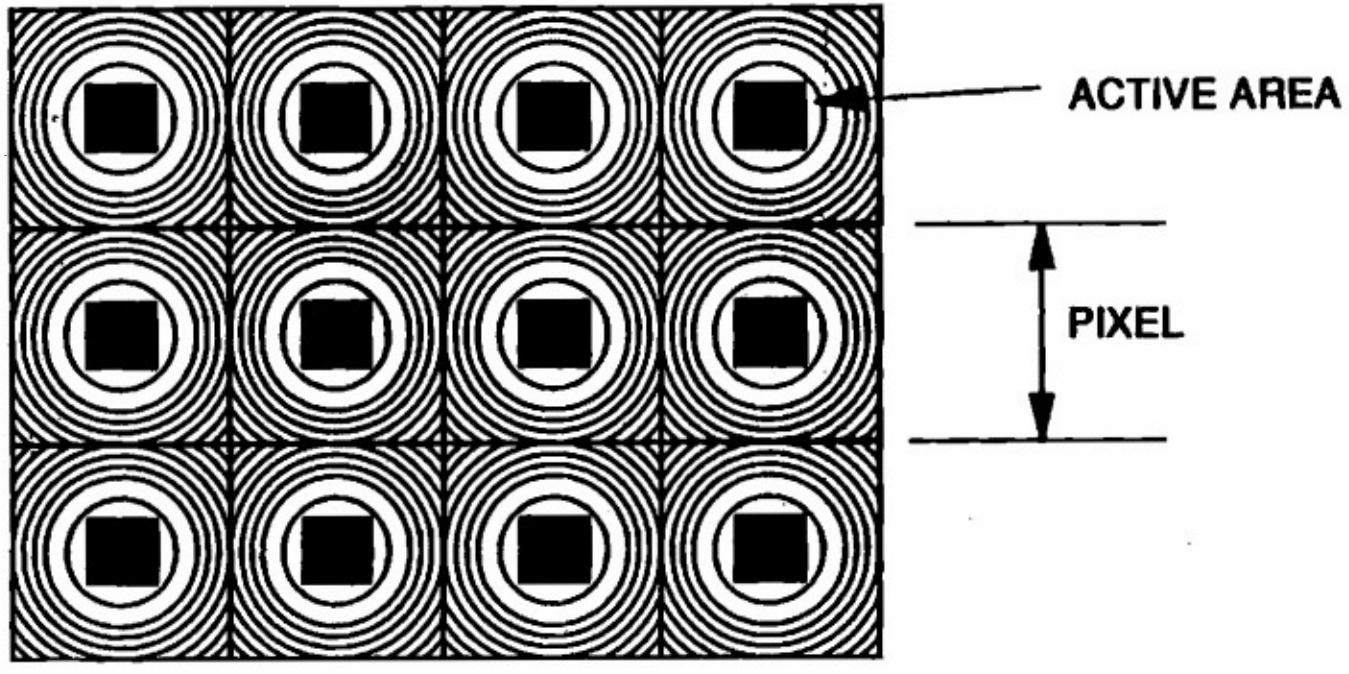
CONVENTIONAL ALL SPHERICAL
EYERELIEF 15mm
WEIGHT 59 grams

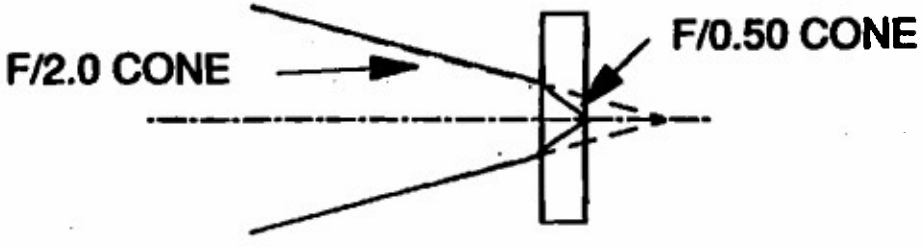


BINARY EYEPIECE DESIGN
EYERELIEF 21mm
WEIGHT 43 grams



MICROLENS ARRAY CONCEPT FOR FLIRs

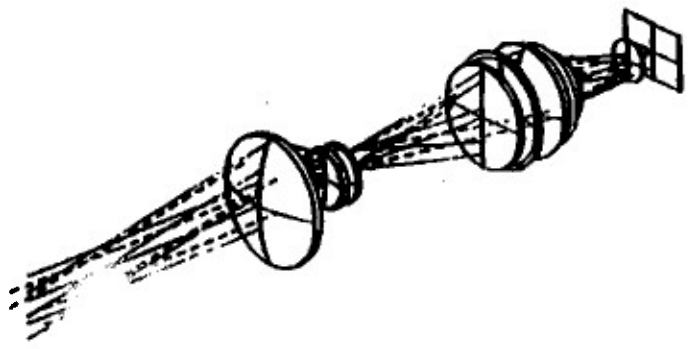




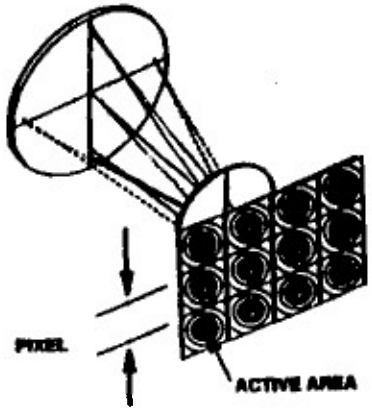
BINARY OPTICS ADD IMMERSION LENS TO EACH DETECTOR ARRAY ELEMENT

BINARY MICROLENS CONCEPT

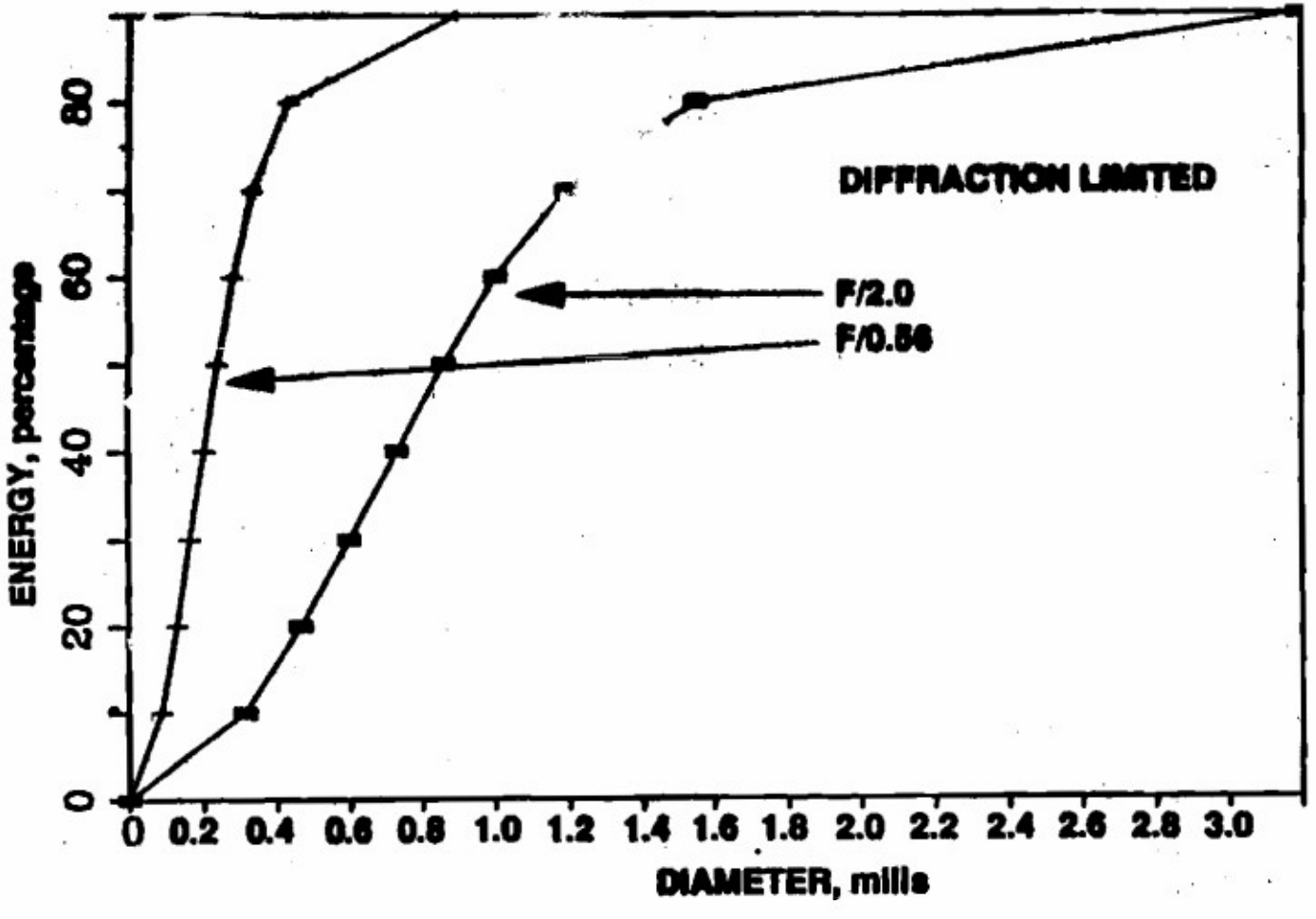
IR LENS SYSTEM



IMPROVES DETECTOR FILL FACTOR
ADDITIONAL REAL ESTATE FOR PROCESSING
INCREASED D-STAR FOR NON-BLIP DETECTOR
EACH DETECTOR HAS MICROLENS

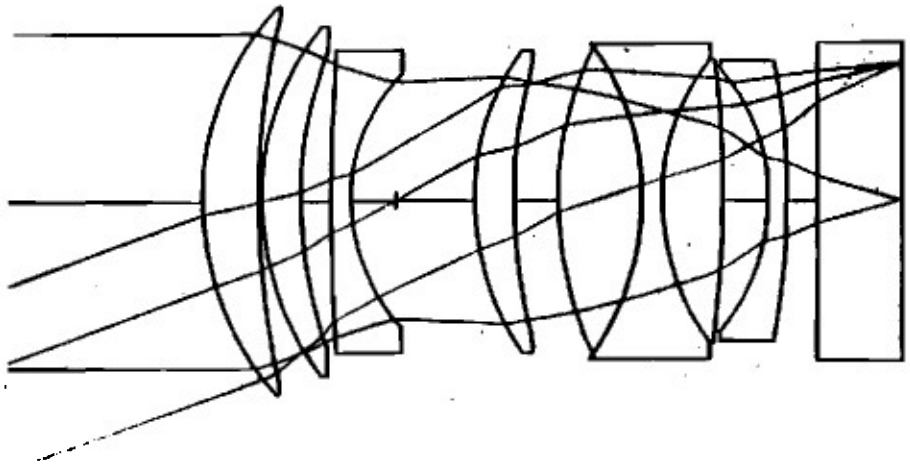


RADIAL ENERGY DISTRIBUTION

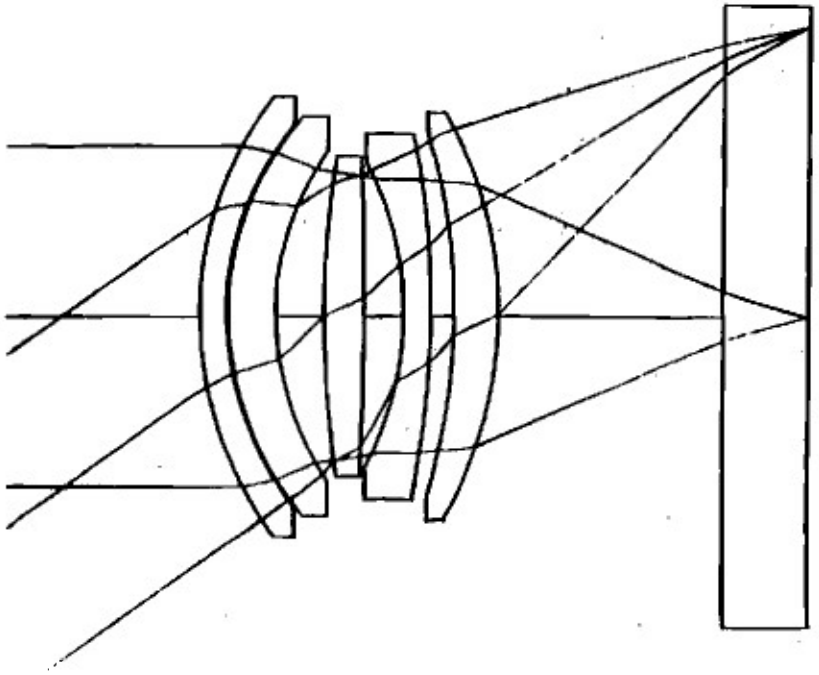


ADVANCED OPTICS RESEARCH
OPTICS DESIGN
NOTIONAL GRADIENT INDEX OPTICS

AN/AVS-6 OBJECTIVE
27mm, f/1.2
40 DEGREE FOV
WEIGHT 31 GRAMS



NOTIONAL GRADIENT INDEX OBJECTIVE
27mm, f/1.2
70 DEG. FOV
WEIGHT 27 grams



HOLOGRAPHIC OPTICS

- HOLOGRAPHIC

  THE USE OF PHOTOGRAPHIC MATERIAL TO STORE A REFERENCE PHASE WAVEFORM THAT CAN WHEN BE USED AS AN OPTICAL ELEMENT.

- MATERIALS

  DICHROMATED GELATIN

  PHOTOPOLYMER (POLAROID DPD 1208/DUPONT)

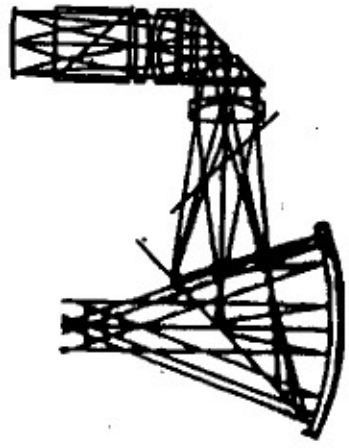
- PROBLEM AREA - DEVELOP A MATERIAL THAT CAN WITHSTAND THE MILITARY ENVIRONMENT AND PERFORM OPTICALLY.

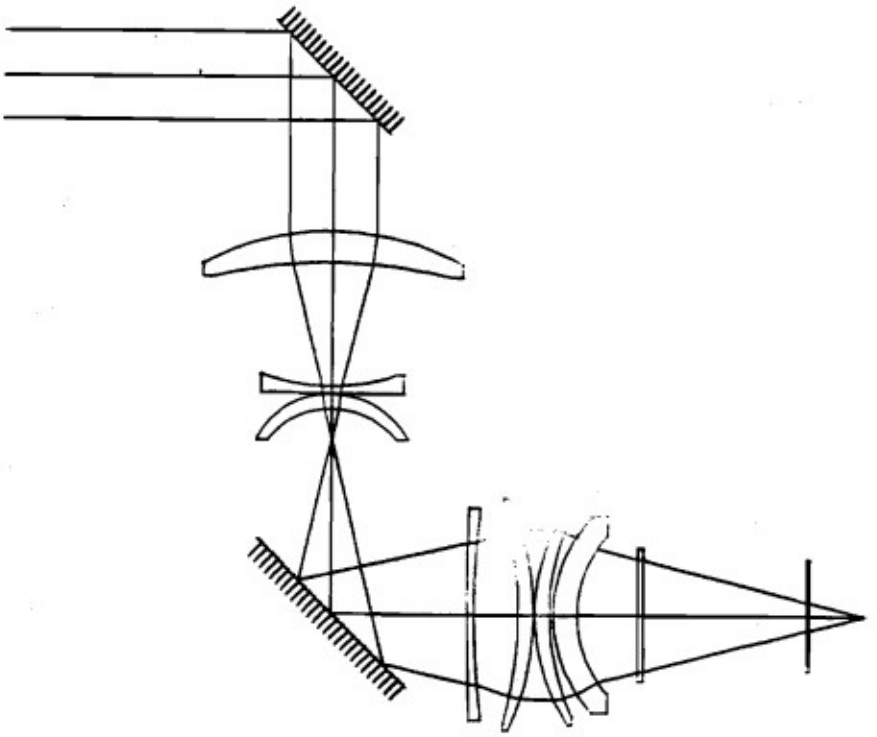
- PAYOFF FOR THE ARMY

  WIDER FIELD-OF-VIEW DISPLAYS

- POTENTIAL APPLICATIONS

  HMD OPTICS FOR AIMS ATTD





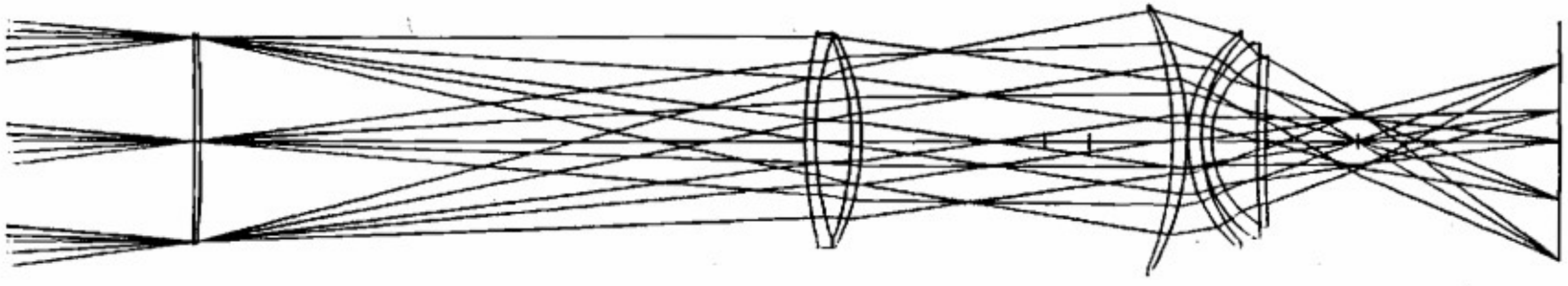
2nd GEN REIMAGING IM
AGER

SCALE 2.700 XZ
1.43 IN

POSITION 1
RAS 13-FEB-91

AUTOSPEC AFOCAL (8-12 micron)

EFFECTS CHANGE OF 7.5 DEGREE FIELD OF VIEW COLLIMATOR
TO 60 DEGREE FIELD OF VIEW COLLIMATOR



AUTOSPEC COLLIMATOR EXIT PUPIL

PILOTAGE FLIR ENTRANCE PUPIL

9.09 IN

AFOCAL. INCREASE FOV AUTOSPEC COLLIMATOR SCALE 0.110 MRG 13-FEB-91

CENTER FOR OPTO-ELECTRONIC SYSTEMS RESEARCH
TOWARDS GLOBAL OPTIMIZATION WITH ADAPTIVE SIMULATED ANNEALING

Towards Global Optimization with Adaptive Simulated Annealing

Greg Forbes and Andrew Jones
The Institute of Optics
University of Rochester

Statement of Problem

Configuration space :

$$\boldsymbol{r} = (x_1, x_2, ..., x_n).$$

Merit function :

$$f(\boldsymbol{r}).$$

Constraints (region of interest) :

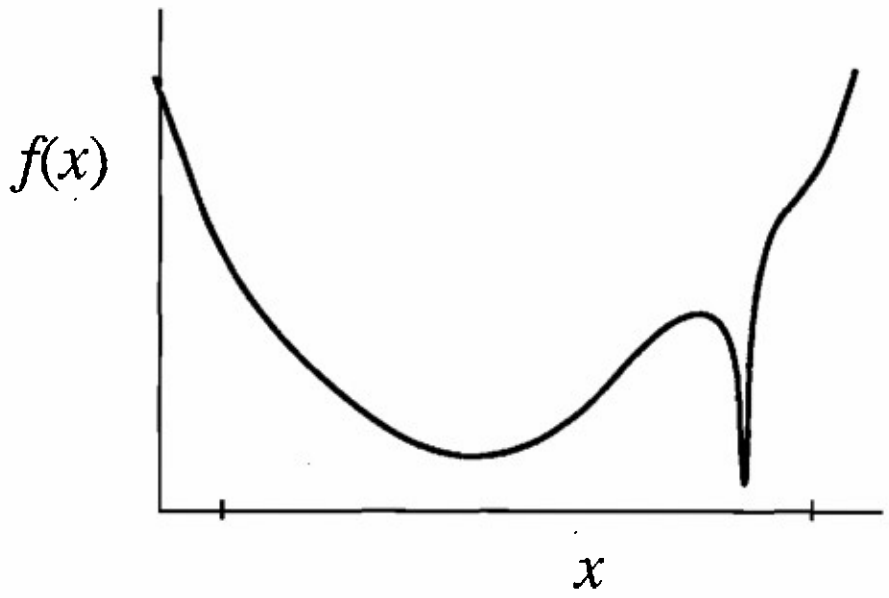
$$C_i(\boldsymbol{r}) < 0 \quad \text{for } i = 1, 2, ..., p,$$
$$E_i(\boldsymbol{r}) = 0 \quad \text{for } i = 1, 2, ..., q.$$

Task :

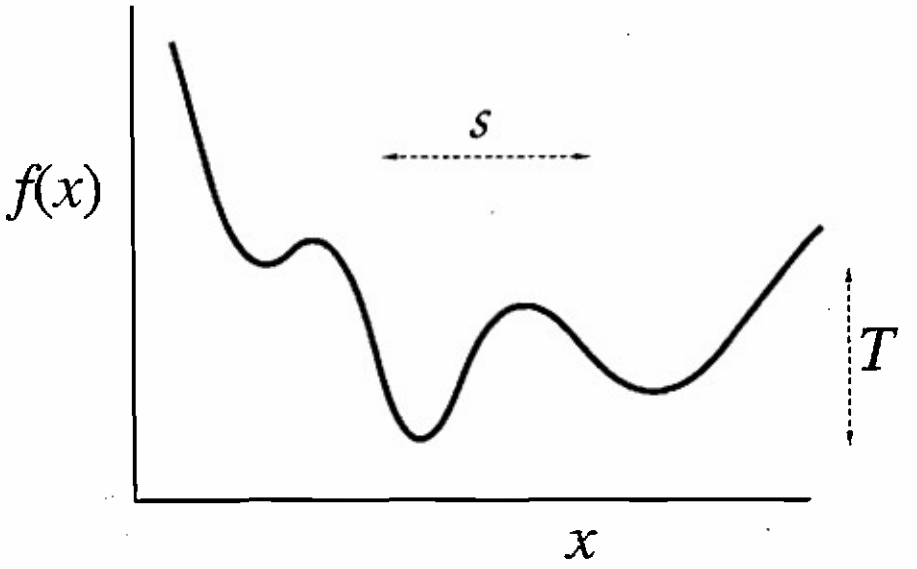
Find the point $\boldsymbol{r}_{\text{Min}}$ [inside the region of interest] where $f(\boldsymbol{r})$ takes its lowest value.

Lesson One

There is no algorithm that is guaranteed to find the global minimum in a finite time.



Simulated Annealing Algorithm



Repeat:

- Take random step of mean distance s
  - if downhill ☞ accept it as new position
  - if uphill ☞ accept with probability $e^{-\Delta f/T}$

Concepts for Annealing

Probability of accepting a step :

$A(\boldsymbol{r}, \boldsymbol{r}')$ = probability of accepting the step from $\boldsymbol{r}$ to $\boldsymbol{r}'$ given that $T = 1/\beta$.

$$A(\boldsymbol{r},\boldsymbol{r}') = \begin{cases} 1, & f(\boldsymbol{r}') < f(\boldsymbol{r}), \\ e^{-\beta[f(\boldsymbol{r}')-f(\boldsymbol{r})]}, & f(\boldsymbol{r}') > f(\boldsymbol{r}). \end{cases}$$

Step distribution :

$S(\boldsymbol{r}, \boldsymbol{r}')\,d\tau'$ = probability of generating a step to within volume element $d\tau'$ at point $\boldsymbol{r}'$ given that the start point is $\boldsymbol{r}$.

Occupation density :

$\rho(\boldsymbol{r})\,d\tau$ = fraction of cases where current point is within volume element $d\tau$ at position $\boldsymbol{r}$. [So, $\int \rho(\boldsymbol{r})\,d\tau = 1.$]
{ Same set-up observed many times. }

Modelling Annealing

Effect of one cycle of algorithm :

If occupation density is $\rho^{[n]}(\boldsymbol{r})$ after n cycles, then

$$\rho^{[n+1]}(\boldsymbol{r}) = \rho^{[n]}(\boldsymbol{r}) + arrivals - departures,$$

where

$$arrivals = \int \rho^{[n]}(\boldsymbol{r}')\, S(\boldsymbol{r}',\boldsymbol{r})\, A(\boldsymbol{r}',\boldsymbol{r})\, d\tau',$$

$$departures = \int \rho^{[n]}(\boldsymbol{r})\, S(\boldsymbol{r},\boldsymbol{r}')\, A(\boldsymbol{r},\boldsymbol{r}')\, d\tau'.$$

Equilibrium density :

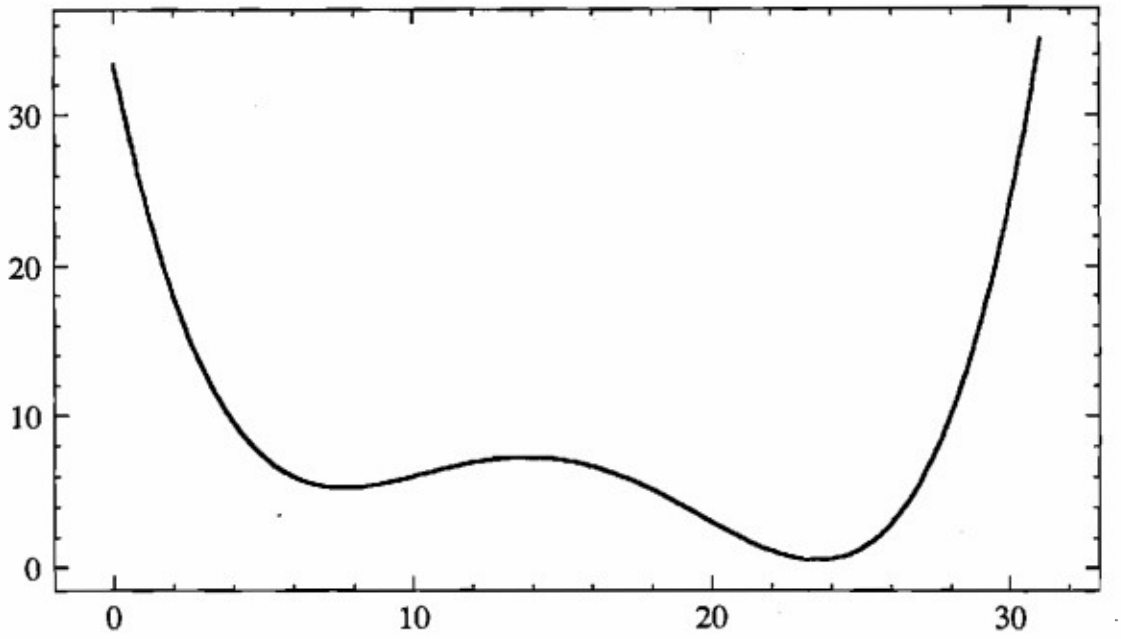
As $n \to \infty$, $\rho^{[n]}(\boldsymbol{r})$ tends towards a fixed distribution:

$$\rho_{eq}(\boldsymbol{r};\beta) = \frac{1}{N(\beta)} e^{-\beta f(r)},$$

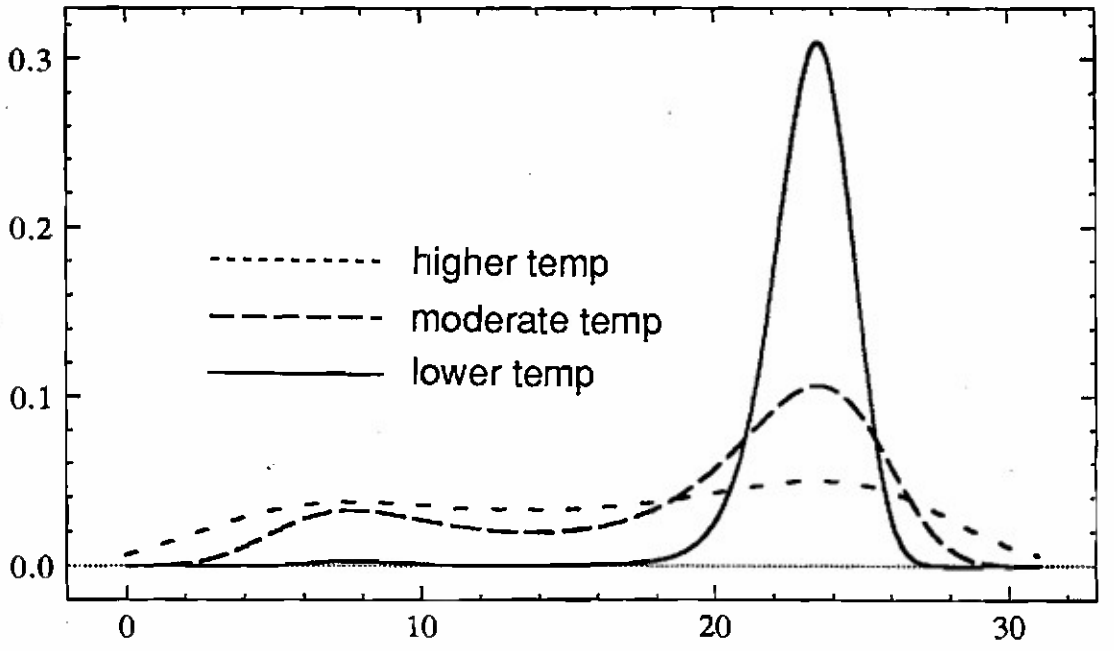
$$N(\beta) = \int e^{-\beta f(r)}\, d\tau,$$

provided $S(\boldsymbol{r},\boldsymbol{r}') = S(\boldsymbol{r}',\boldsymbol{r})$.

A simple merit function



Equilibrium density plots



"Optimal" temperature and step schedules

Heuristic principle :

At every step, keep the occupation density as close to equilibrium as possible.

...need to have measure of distance between distributions, so define one:

$$\mathcal{D}\{\rho_a(\boldsymbol{r}), \rho_b(\boldsymbol{r})\} = \tfrac{1}{2}\int |\rho_a(\boldsymbol{r}) - \rho_b(\boldsymbol{r})|\, d\tau.$$

Measures of equilibration

- Rate of change of equilibrium distribution with temperature

$$\mathcal{S} := \lim_{\Delta\beta\to 0}\left\{\frac{\mathcal{D}\left\{\rho_{eq}(\boldsymbol{r};\beta),\ \rho_{eq}(\boldsymbol{r};\beta+\Delta\beta)\right\}}{\Delta\beta}\right\}.$$

- Reduction *per step* in distance to equilibrium: Suppose $\rho^{[n]}(\boldsymbol{r}) = \rho_{eq}(\boldsymbol{r};\beta)$ and temperature is changed to $\beta+\Delta\beta$ then define

$$\mathcal{R} := \lim_{\Delta\beta\to 0}\ \frac{\mathcal{D}\left\{\rho^{[n+1]}(\boldsymbol{r}),\ \rho_{eq}(\boldsymbol{r};\beta+\Delta\beta)\right\}}{\mathcal{D}\left\{\rho^{[n]}(\boldsymbol{r}),\ \rho_{eq}(\boldsymbol{r};\beta+\Delta\beta)\right\}}.$$

Now can do some algebra to find:

$$\mathcal{S} = \tfrac{1}{2}\int \rho_{eq}(\boldsymbol{r};\beta)\,|f(\boldsymbol{r}) - \langle f \rangle|\,d\tau,$$

where

$$\langle f \rangle := \int \rho_{eq}(\boldsymbol{r};\beta)\, f(\boldsymbol{r})\, d\tau.$$

Also have

$$\mathcal{R} = \frac{\int \rho_{eq}(\boldsymbol{r};\beta)\left|f(\boldsymbol{r}) - \langle f \rangle + \int [f(\boldsymbol{r}') - f(\boldsymbol{r})] S(\boldsymbol{r},\boldsymbol{r}') A(\boldsymbol{r},\boldsymbol{r}') d\tau'\right| d\tau}{\int \rho_{eq}(\boldsymbol{r};\beta)\,|f(\boldsymbol{r}) - \langle f \rangle|\,d\tau}.$$

Key Observations :

- It is possible to evaluate numerically any such entity during the annealing process.
- Can use $\mathcal{S}$ and $\mathcal{R}$ to control algorithm.

Restate heuristic

Keep within a distance of ε of equilibrium:

If β_n is temperature at cycle n, require

$$\mathcal{D}\left\{ \rho^{[n]}(\boldsymbol{r}),\ \rho_{eq}(\boldsymbol{r};\beta_n) \right\} < \varepsilon, \quad \text{for all } n.$$

- Can now deduce that, at each cycle, β should change by no more than:

$$\Delta\beta = \frac{\varepsilon(1-\mathcal{R})}{\mathcal{S}}.$$

NB. Have other checks on whether cooling too quickly that can be used as governors. For example, it is easy to derive the following standard result:

$$\frac{d<f>}{d\beta} = -[<f^2> - <f>^2].$$

Skeleton of simple algorithm

Input : merit function, constraints, and ε.

- Cycle at $\beta = 0$ (ie. infinite temperature) to initialize the statistics $\mathcal{R}$, S, $\langle f \rangle$, etc.
- **repeat**
  - **repeat**
    - generate a random step
    - reflect from linear constraints
  - **until** have feasible system
  - accept or reject current step
  - update statistics
  - increase β by $\frac{\varepsilon(1-\mathcal{R})}{S}$
  - **until** β is sufficiently high.

Key points in context of lens design :

- Efficient evaluation of merit function.
- Explicitly incorporate constraints.
- Transformations of variables & merit fn.

CENTER FOR OPTO-ELECTRONIC SYSTEMS RESEARCH
LENS DESIGN WIT ADAPTIVE SIMULATED ANNEALING

Lens Design with Adaptive Simulated Annealing

Andrew Jones
Institute of Optics
University of Rochester

Goal:

To locate with maximum certainty the global minimum of the merit function for a lens system.

Merit function:

RMS spot size, integrated over the pupil, field and color:

$$\Phi = \left[\int_{h} \int_{\lambda} \int_{pupil} \int (\varepsilon_x^2 + \varepsilon_y^2)\, dy\, dx\, d\lambda\, dh \right]^{1/2}$$

Variables:

Curvature,
Thickness,
Entrance pupil location,
etc.

Region of interest (constraints on independent variables):

- Simple inequality constraints:

  $c_{min} \leq c \leq c_{max}$

  $t_{min} \leq t \leq t_{max}$

- Feasibility constraints:

  Non-negative edge thicknesses

  $-(\text{semi-aperture})^{-1} \leq c \leq (\text{semi-aperture})^{-1}$

- Equality constraints:

  $u_k' = -1/(2\ F\#)$

Annealing can be applied in a simple fashion:

- generate curvatures, thicknesses, etc. randomly
- discard systems which violate constraints
- evaluate remaining systems

This leads to problems at high T, when we are trying to sample nearly uniformly over the region of interest -- almost all generated systems violate constraints and the annealing process becomes very inefficient.

Improve efficiency:

- transformation of independent variables
- reflection of step vectors from constraints

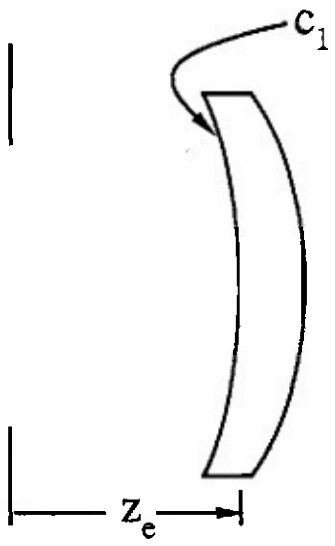
Example of transformation:

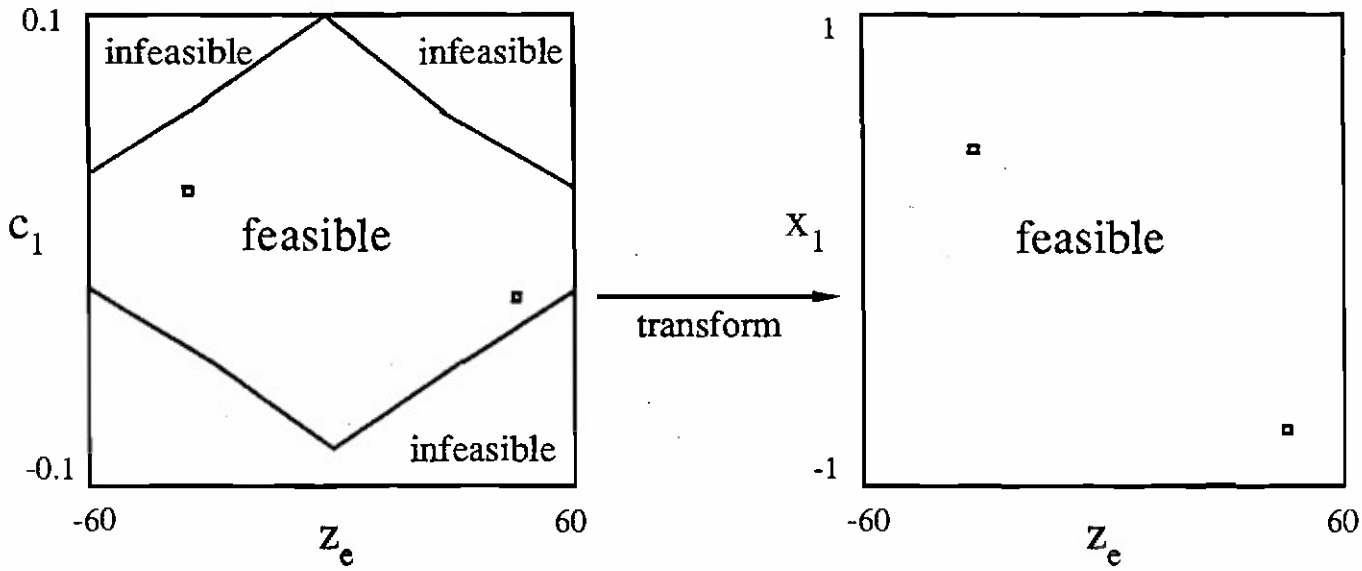
Instead of using curvatures as variables, use transformed variables **x**:

$$c = c_{min} + (x + 1)(c_{max} - c_{min}) / 2 \qquad -1 \leq x \leq 1$$

where c_{min} and c_{max} are the minimum and maximum feasible values of the curvature. The constraints are now "combined" with the independent variables, and the region of interest is now an N-dimensional "box."

Example: F/8 Landscape lens, 15° hfov, 160 mm efl

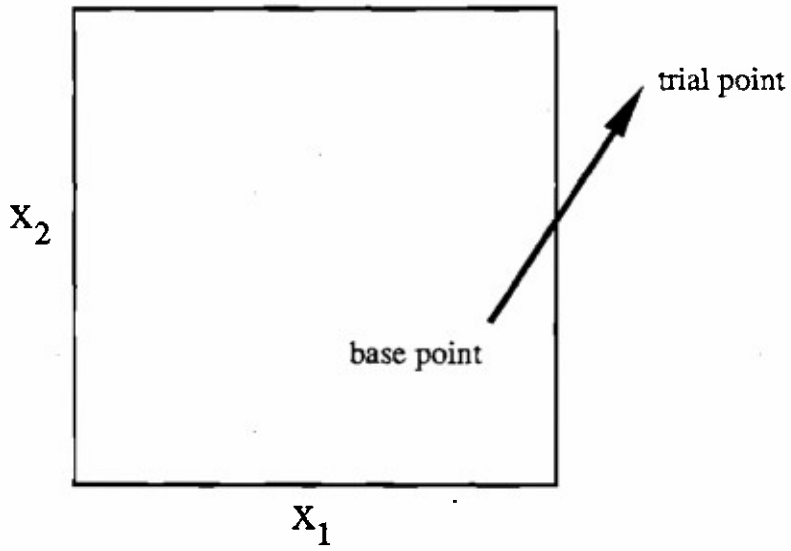




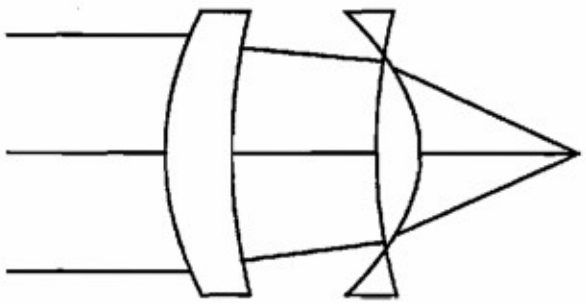
Reflection:

Even after transformation, some infeasible systems may be generated:

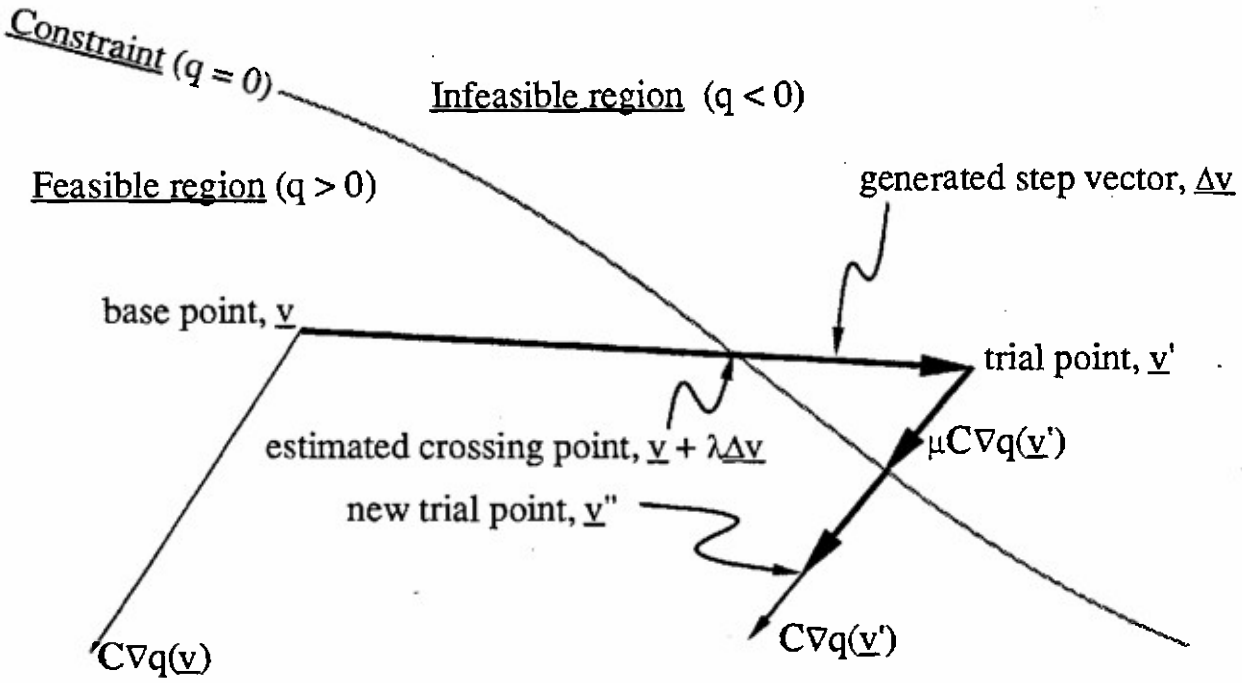
- trial point with $x > 1$ or $x < -1$



- constraint violation at last surface (dependent variable)



Instead of discarding the infeasible trial point, we perform a reflection about the constraint which is violated:



The function q(w) gives an estimate of the "feasibility" of a point w:

q(w) > 0 if w is feasible
q(w) < 0 if w is infeasible
q(w) = 0 if w is on the constraint

The larger the magnitude of q(w), the "farther" w is from the constraint.

The reflection should be performed so that the equilibrium distribution at $\beta = 0$ is uniform.

Reflection algorithm:

1. Estimate the point where the step vector crosses the constraint, $\underline{v} + \lambda\underline{\Delta v}$
2. Choose the value of μ such that $\underline{v'} + \mu C \nabla q(\underline{v'})$ lies on the linear approximation to the constraint
3. Set the new trial point $\underline{v''}$ equal to $\underline{v'} + 2\mu C \nabla q(\underline{v'})$

where

$\mathbf{C}$ = covariance matrix of step distribution

$\underline{v}$ = base point

$\underline{\Delta v}$ = generated step vector

$\underline{v'} = \underline{v} + \underline{\Delta v}$ = trial point (infeasible)

$\underline{v''}$ = new trial point

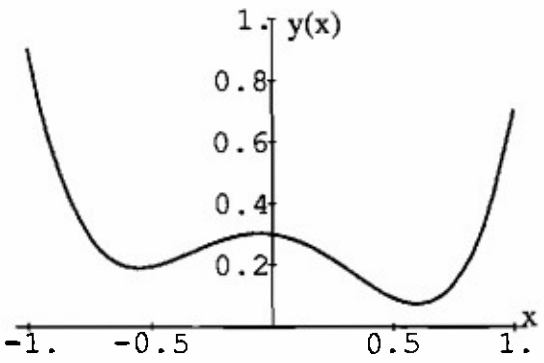
If the new trial point $\underline{v''}$ is also infeasible, it is discarded and a new step vector is generated randomly.

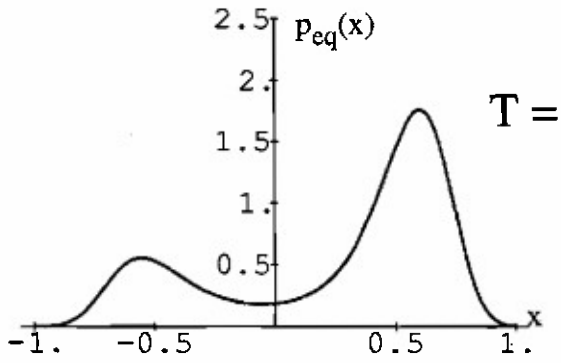
Transformations:

Monotonic nonlinear transformations of the independent variables and the merit function can change greatly the behavior of the algorithm.

Example:

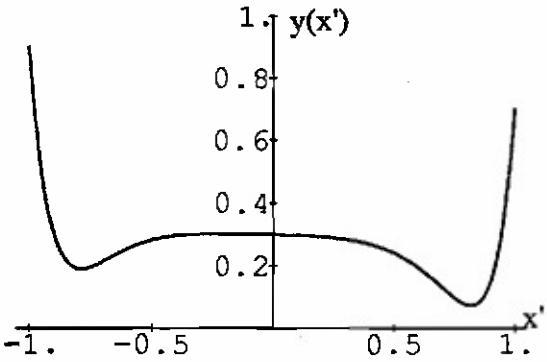
$y(x) = 1.5\,x^4 - x^2 - 0.1\,x + 0.3 \qquad -1 \le x \le 1$

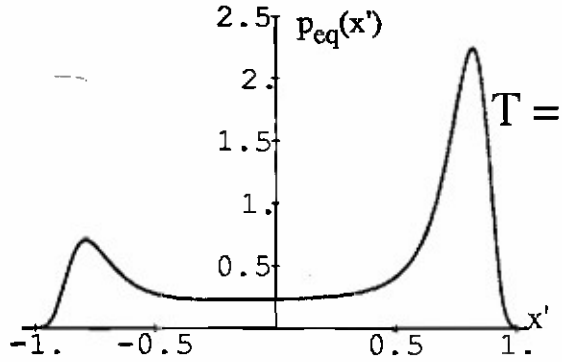




T = 0.1

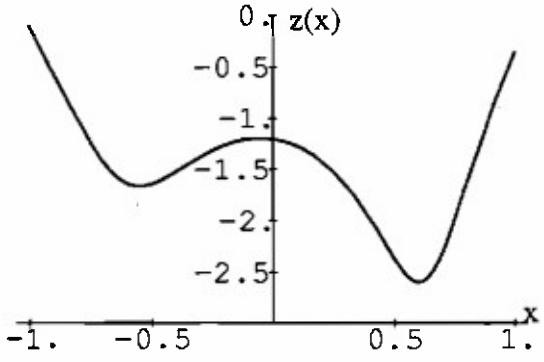
Transform: $x = 0.8\,(x')^3 + 0.2\,x'$

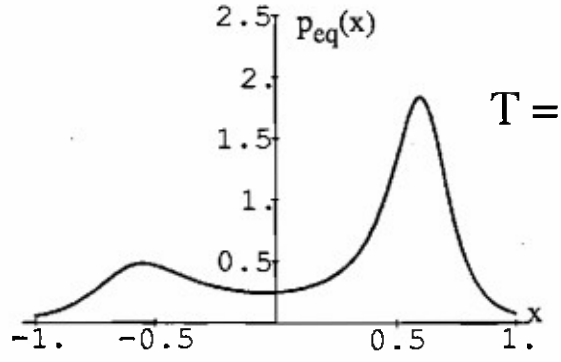




T = 0.1

Transform: $z(x) = \log y(x)$

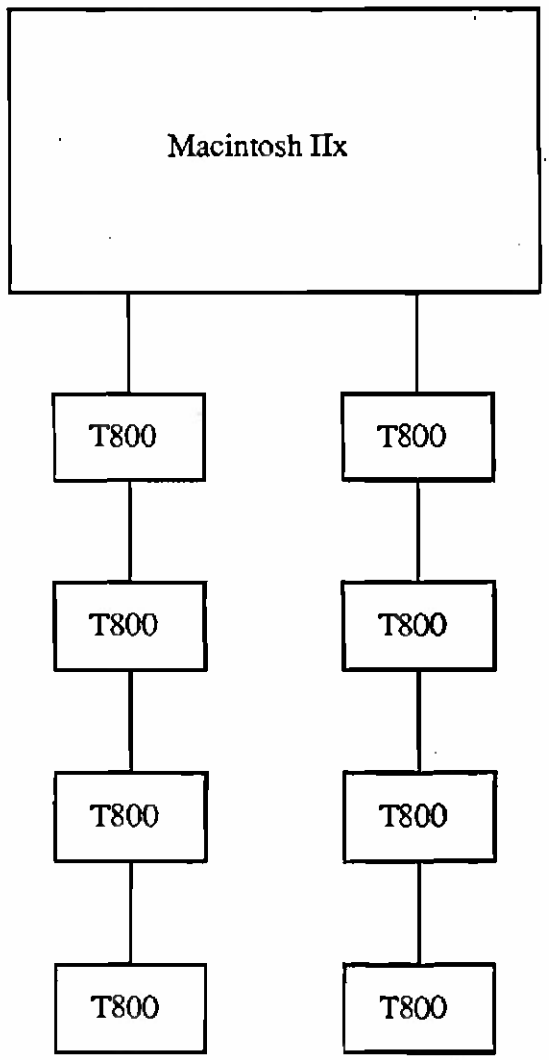




T = 0.7

Adaptive simulated annealing on parallel processors:

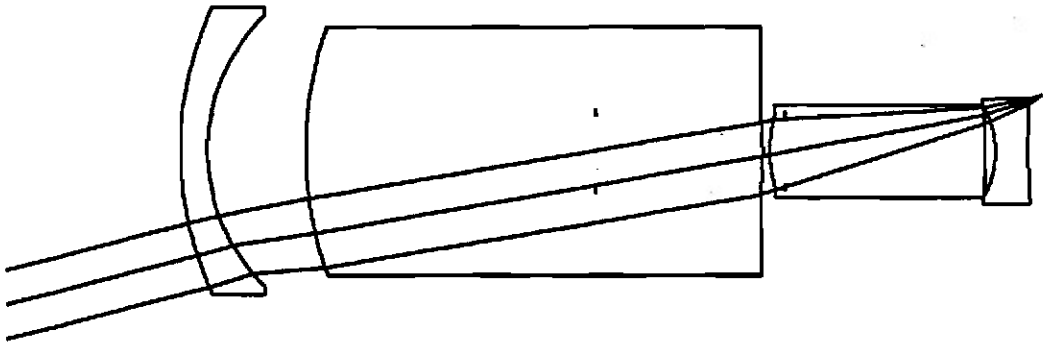
Adaptive simulated annealing depends heavily upon the statistics of previously generated trial points in order to determine the temperature schedule and step distribution. By using parallel processors, we can improve greatly the quality of the statistics collected.



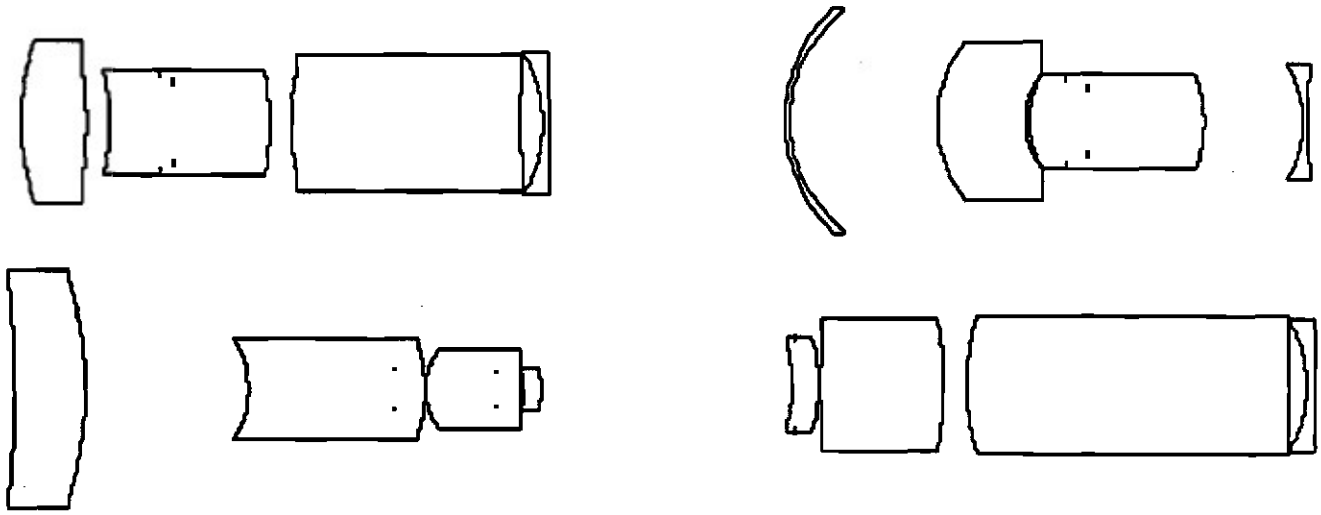
Using eight transputers, we can generate and evaluate for each base point groups of eight trial points until an accepted point is found. The trial points and merit function values are used in formulating statistics. The annealing process is also much more efficient.

Experiences with adaptive simulated annealing

- algorithm has been developed using "monochromatic quartet" problem from ILDC '90:
  - many local minima
  - minima differ significantly
  - most minima are near constraints
  - 15 dimensions



F/3, 100 mm efl, hfov 15°



- algorithm seems to work well on polychromatic systems if proper glass choices are made *a priori*

Possible future modifications:

- run downhill optimizer from selected starting points
- develop method for optimizing glass combinations

REFERENCES

Towards global optimization with adaptive simulated annealing

G. W. Forbes and Andrew E. W. Jones

The Institute of Optics
University of Rochester
Rochester, NY 14627

Abstract

The structure of the simulated annealing algorithm is presented and its rationale is discussed. A unifying heuristic is then introduced which serves as a guide in the design of all of the sub-components of the algorithm. Simply put, this heuristic principle states that, at every cycle in the algorithm, the occupation density should be kept as close as possible to the equilibrium distribution. This heuristic has been used as a guide to develop novel step generation and temperature control methods intended to improve the efficiency of the simulated annealing algorithm. The resulting algorithm has been used in attempts to locate good solutions for one of the lens design problems associated with this conference, viz. the "monochromatic quartet", and a sample of the results is presented.

1 Global optimization in the context of lens design

Whatever the context, optimization algorithms relate to problems that take the following form: Given some configuration space with coordinates $r = (x_1, x_2, \ldots, x_n)$ and a merit function written as $f(r)$, find the point $r_{\min}$ where $f(r)$ takes it lowest value. That is, find the global minimum. In many cases, there is also a set of auxiliary constraints that must be met so the problem statement becomes: Find the global minimum of the merit function within the region defined by $E_j(r) = 0,\ j = 1, 2, \ldots, p$ and $I_j(r) > 0,\ j = 1, 2, \ldots, q$. This region is referred to in what follows as the *region of interest*. In the context of lens design, the coordinates are just the system parameters in one form or another (e.g. refractive indices, thicknesses, curvatures or radii, etc.), the inequality constraints may take the form of the requirement for positive edge thicknesses or specifying maximum glass thicknesses on axis, and an example of an equality constraint is the requirement of a specific focal length.

It is evident that such a problem is easy to state, however, its solution is another matter — in fact, this problem of global optimization is generally intractable. This is readily appreciated by considering even the simplest case illustrated schematically in Fig. 1 where a smooth one dimensional merit function is represented. Even if the region of interest is finite (i.e. there are inequality constraints of the form: $x - x_0 > 0$, and $x_1 - x > 0$), it is clear that by making the steep dip in the function narrower, it is possible to make the problem of locating the global minimum arbitrarily difficult.

In practice, as a consequence of certain bounds associated with realistic problems, the task of locating the global minimum is feasible, in principle. In most cases, however, the effort required to find the minimum is truly formidable. For example, imagine that the problem illustrated in Fig. 1 represents a lens design problem where the variable is, say, a thickness limited to the interval 0 to 100mm. It is clear that it is possible to sample the merit function on a regular grid where the sample spacing is taken to be comparable to the manufacturing tolerance on such a thickness, presumably somewhere between 10^{-1}mm and 10^{-3}mm. In this way, with something like a thousand samples (and certainly no more than one hundred thousand), it is possible to guarantee that all minima of any practical significance will be found.

Assuming that the cost of each evaluation of the merit function is typical of lens design problems, this computation is hardly a challenge at all — even for a personal computer. Consequently, the global minimum can be guaranteed in such a case. However this problem is not representative of real design problems; the key issue is the dimensionality.

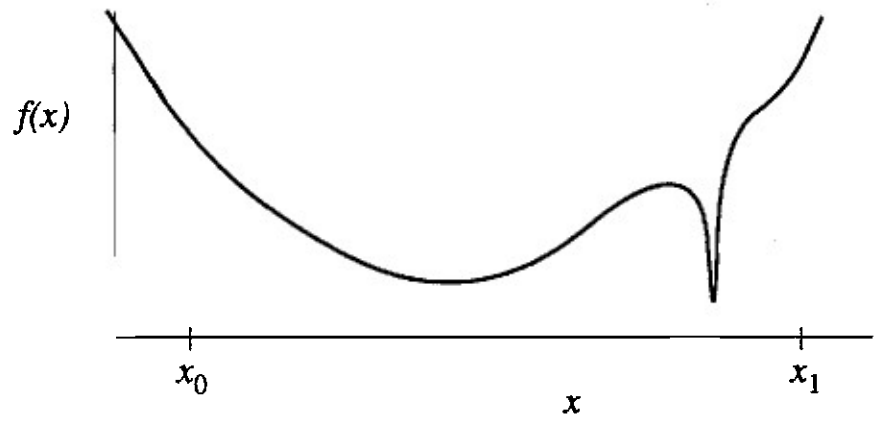


Fig. 1 A simple one dimensional merit function.

Consider the case of the "monochromatic quartet" (one of the lens design problems associated with this conference). This system has four elements which means that there are eight surfaces (each of which has an associated curvature and thickness) and, since the refractive indices are fixed, there is a total of sixteen system parameters. In fact one degree of freedom is lost since the focal length is required to take a specified value so the problem is seen to be fifteen-dimensional. This, of course, is relatively modest since the refractive indices are fixed, all surfaces are required to be spheres, and there are only four elements. It is clear that many problems in lens design are of a significantly higher dimensionality. Now consider the prospect of sampling on a regular grid in this fifteen-dimensional space. Suppose that, just to get started, a grid is laid out with only ten samples in each dimension. (It was suggested above that possibly a thousand points or so would be more typical for a reasonable guarantee.) In fifteen dimensions, this coarse grid requires sampling at 10^{15} points in the configuration space. Supposing that the merit function can be evaluated by tracing just a few rays, there are of the order of 10^{16} - 10^{17} ray-surfaces to be traced here. Currently, supercomputers can trace no more than about 10^{6} ray surfaces per second so this job would require at least 10^{11} seconds. This is about ten thousand years — and we've hardly scratched the surface with such coarse sampling. Increasing the sampling to just thirty points in each dimension pushes the supercomputer computation time for this modest lens design problem up by another seven orders of magnitude to far beyond the age of the universe! This is a formidable global optimization task and there can be no guarantees.

Of course, the lack of a guarantee need not proscribe a search. It is clear, however, that some drastic simplification of the problem is necessary. The most common approach is to reduce the demands placed on the algorithm and design it to seek any local minimum. Most optimization algorithms used in lens design simply seek a local minimum in the neighborhood of some specified starting point by proceeding downhill until some flat terrain is found. This is a relatively straightforward process and there is a variety of algorithms designed for this purpose. In essence, the tough component of the design problem now becomes finding a good starting point for such a blinkered algorithm to polish up. With the prospect of individuals having workstations on their desk (more than likely sitting idle each night), a more ambitious, automated search of the configuration space is becoming a realistic option. In many practical problems, it seems reasonable to expect to find some good local minima.

This particular option is what we have in mind when we speak of a move towards global optimization. The goal of our research in this area is the specification of an algorithm that, speaking loosely, has the maximum chance of finding the best minima for any given amount of computational effort. It is expected that the chance of success increases with increasing effort, so when such an algorithm is invoked the user must provide an indication (even if only a relative one) as to how much effort is to be expended in the current search. The algorithm would explore the region of interest in the configuration space in search of good minima and, *given sufficient computational effort*, would undoubtedly succeed in discovering minima that are comparable to (if not better than) those uncovered using the conventional methods. (Notice that such an algorithm should not require a user-specified starting point — just the specification of a region to explore.) The question that immediately comes to mind is: How much computational effort is required in the case of modest lens design problems? In 1990 terms, the answer is expected to lie somewhere between a PC-day and a Cray-month.

A related, but more easily answered question is: How much success can be had in the context of lens design with a reasonable expenditure of computational effort? By "reasonable", it is meant that the algorithm should terminate within a realistic period of time, say a couple of hours or even a couple of days, when running on a medium sized computer, say one whose cost is of the order of a typical annual salary of an engineer. At present, the price may be around \$10K - \$100K which includes machines capable MFLOPS and multiprocessing. It is clear that what is "reasonable" by this definition is sure to be significantly different in just five years. Consequently, the answer to this question is forever changing and we must make do with a rough cut at the answer for 1990. To this end, in Section 4, we present the results for the design of a monochromatic quartet derived by one particular global optimization algorithm.

2 Introduction to simulated annealing

The development of global search algorithms has proceeded, in some cases, by making analogies with models varying from biological evolution with pairwise reproduction in populations subject to random mutation, through to thermodynamics and the configuration of an alloy during cooling. On account of its simplicity and elegance, we have chosen the general approach of the so-called simulated annealing algorithm[1] as the underlying philosophy upon which to base our investigation.

The simulated annealing algorithm is essentially a prescription for a partially random walk in the configuration space. At each cycle of the algorithm, a random step is generated and, upon evaluation of the merit function at the associated trial point, a decision is made as to whether to accept that step or not. A more specific description of the algorithm involves two entities. The first parameter, referred to as the *temperature* and written here as T, is used in determining whether a proposed step is to be accepted. In particular, if the merit function is reduced as a result of taking the step, the step is always accepted. On the other hand, if the merit function is increased by an amount Δf, the step is accepted with a probability equal to $e^{-\Delta f/T}$. (One way to realize the process of accepting a step with probability p is to generate a random number, say u, uniformly distributed between zero and one and then accept the step if $u < p$.) That is, T simply sets a scale on the size of the allowed increases in the merit function value for any single step. The probability of accepting a step from the point with coordinates $\mathbf{r}$ to the point $\mathbf{r}'$, say $A(\mathbf{r}, \mathbf{r}')$, is now seen to be given by

$$A(\mathbf{r}, \mathbf{r}') = \begin{cases} 1, & f(\mathbf{r}') < f(\mathbf{r}), \\ e^{-\beta[f(\mathbf{r}')-f(\mathbf{r})]}, & f(\mathbf{r}') > f(\mathbf{r}), \end{cases} \tag{2.1}$$

where β is defined by $\beta := 1/T$. The second entity is a probability density that characterizes the distribution of the random steps generated at each cycle of the algorithm: given that the current point in the walk about the configuration space has coordinates $\mathbf{r}$, the probability that a generated step lands within a volume element $d\tau' = dx'_1\, dx'_2 \ldots dx'_n$ at some

point r' is written as $S(r, r')\,d\tau'$. Notice that, on account of this definition, it follows that $\int S(r, r')\,d\tau' = 1$ for all r. In these terms, the heart of the simulated annealing algorithm takes the form of repeatedly carrying out the following two operations:

- Generate a random step from the current point r with distribution $S(r, r')$.
- `If` $f(r') < f(r)$ `then` Accept the step (i.e. r is replaced by the value of r'),
  `else` Accept the step with probability $e^{-\beta[f(r')-f(r)]}$.

Cycling through these two operations essentially generates a tour about the configuration space and, to understand the rationale behind this algorithm, it is important to characterize the manner in which such a tour occupies the space.

To this end, consider defining an occupation density as follows: First imagine a set-up consisting of a host of computers all carrying out this guided tour — each one starting in the same way (either from a specific point in the space or from a point in the region of interest chosen at random in the same way on each machine) but with the different computers starting their random number generators with independent seeds. Suppose that all the computers execute m cycles of the algorithm given above and now define the occupation density $\rho^{[m]}(r)$ in the configuration space to characterize the distribution of the final positions arrived at by each of the computers. That is, $\rho^{[m]}(r)\,d\tau$ is just the fraction of the points within the volume $d\tau$ at point r and it follows that $\int \rho^{[m]}(r)\,d\tau = 1$, for all m. According to this definition, $\rho^{[0]}(r)$ characterizes the distribution of the start points, $\rho^{[1]}(r)$ characterizes the distribution of the points after one cycle, etc.

It is now straightforward to derive an explicit expression for $\rho^{[m+1]}(r)$ in terms of $\rho^{[m]}(r)$:

$$\rho^{[m+1]}(r) = \rho^{[m]}(r) + arrivals - departures, \tag{2.2a}$$

where

$$arrivals = \int \rho^{[m]}(r')\, S(r', r)\, A(r', r)\, d\tau',$$

$$departures = \int \rho^{[m]}(r)\, S(r, r')\, A(r, r')\, d\tau'. \tag{2.2b}$$

With this expression, it is possible to investigate the evolution of the occupation density and explore the effect of different step distributions , etc. For any fixed step distribution and temperature, the most important and striking result is that, as m becomes large, this occupation density tends to a fixed distribution. This fixed distribution, called the *equilibrium distribution* and written here as $\rho_{eq}(r)$, can be determined by simply taking $\rho^{[m+1]}(r) = \rho^{[m]}(r) = \rho_{eq}(r)$ in Eq.(2.2). Provided the step distribution is chosen to be symmetric [i.e. $S(r, r') = S(r', r)$] and T is non-zero, the resulting equation is readily seen to be satisfied by an equilibrium distribution of the form

$$\rho_{eq}(r;\beta) = \tfrac{1}{N(\beta)}\, e^{-\beta f(r)}, \tag{2.3}$$

where the temperature dependence is now made explicit for the equilibrium density and $N(\beta)$ is simply a normalization factor given by

$$N(\beta) = \int e^{-\beta f(r)}\, d\tau. \tag{2.4}$$

(As may be expected, there are, in fact, other weak conditions on the form of the merit function and step distribution to attain this equilibrium form — e.g. that they be not too pathological — but they are of little significance here.)

Notice that the equilibrium distribution is not only independent of the step distribution but it is also independent of the initial occupation density. It is now clear that, as may be expected for such a partially random walk, the choice of start point is of no consequence. Also notice that if β is set equal to zero (i.e. the temperature is infinite so all steps are

accepted) the equilibrium density corresponds to uniform occupation of the configuration space. However, when β is non-zero the equilibrium density is larger where the merit function is lower and it is easy to see that, as β becomes large, the equilibrium density corresponds to having all points clustered around the global minimum. The rationale behind simulated annealing now becomes clear: start at high temperature (where the equilibrium distribution is essentially uniform) and cycle while gradually reducing the temperature. If the temperature is reduced to zero *infinitely slowly*, the occupation density is always identical to the associated equilibrium density and we are guaranteed to end up at the global minimum. In practice, of course, the temperature must be reduced faster than this and, as a result, the global minimum can not be guaranteed.

It is now clear that there are two crucial aspects in any implementation of the simulated annealing algorithm. The first is to determine how to control the temperature — there are surely stages when it would be advantageous to reduce the temperature relatively slowly. Similarly, the optimal step distribution must be determined and it can be expected that this will change with temperature. In earlier work, the temperature control (both the initial value and the reduction schedule) and the step generation components of the algorithm have often been simplistic and governed by a multiplicity of unrelated *ad hoc* prescriptions. For example, the temperature may be reduced by a specific fraction every N steps, where N is predetermined, and the steps chosen to uniformly fill a rectangular parallelepiped aligned to the coordinate axes and scaled such that roughly 50% of the trial steps are accepted.[2] Others have proposed coupling the parameter that is referred to here as temperature directly to the merit function value.[3] This radical proposal may prove to be effective, but it breaks the principal link in the analogy with the physical annealing process (i.e. the equilibrium distribution discussed above has no relevance to this, so-called, generalized simulated annealing algorithm). We have taken a different path and adopted a unifying, simple heuristic principle that serves as a guide in the design of every aspect of the algorithm and allows for harmony in a completely adaptive algorithm. The resulting algorithm, which we refer to as *adaptive simulated annealing*, is also designed such that its performance is unchanged by linear transformations of the coordinates.

3 Adaptive simulated annealing

It is intuitive that there are temperature schedules and step generation procedures that are in some sense optimal for annealing. For example, it is now apparent that the parameter referred to as temperature determines the region of spread of the occupation density in the configuration space which, in turn, can be expected to set the scale for the step distribution. At any particular temperature, it is evident that if the generated steps are too small, the occupation density will be slow to change and therefore equilibration will also be slowed. Conversely, if the generated steps are too big, the vast majority will be rejected so that, once again, equilibration will be slowed. Similarly, there will be times when it will be expedient to reduce the temperature more rapidly. The key issue here is: Just how is the worth of a proposed prescription for step generation or temperature control measured? Ideally, of course, the probability of winding up at the global minimum should be maximized (i.e. the higher this probability, the better the prescription). However, we have found criteria like this to be unworkable so we have tried a simpler alternative.

As a means to define a measure of the effectiveness of any proposed annealing algorithm, we have introduced the following **heuristic principle**: *At every stage of the annealing process, the occupation density should be kept as close as possible to the equilibrium density*. That is, the smaller the maximum distance from equilibrium, the better. On a moment's reflection, it can be appreciated that the traditional "staircase" temperature schedule and the associated concept of cycling at each fixed temperature value "until equilibrium is attained" must be rejected here. In order to apply this heuristic, however, it is necessary to introduce a measure of the distance between two distributions. There are a number of reasonable ways to do this and we have adopted a measure of distance such that the distance between two occupation densities, say

$\rho_a(r)$ and $\rho_b(r)$, is given by

$$\mathcal{D}\{\rho_a(r),\ \rho_b(r)\} = \tfrac{1}{2}\int |\rho_a(r) - \rho_b(r)|\, d\tau. \tag{3.1}$$

With this definition, since all densities are everywhere positive and integrate to one by definition, it follows that distance is always between zero (complete coincidence) and one (nothing in common). If $\rho_b(r)$ is taken to be the desired equilibrium density, this distance can be loosely interpreted as a measure of the fraction of the cases for which the walk characterized by $\rho_a(r)$ is "out of place".

Consider a system where the temperature is changed by a small amount ("small" is given a more definite meaning in a moment) at every step. If at the m'th cycle the reciprocal temperature is β_m, the heuristic can now be stated more specifically as:

$$\mathcal{D}\left\{ \rho^{[m]}(r),\ \rho_{eq}(r;\beta_m) \right\} < \varepsilon, \quad \text{for all } m, \tag{3.2}$$

where ε is some fixed parameter to be specified as input to the algorithm. It can be expected that smaller values of ε will require more computation for any given annealing problem. That is, ε is a relative measure of how exhaustively the algorithm will explore the configuration space.

With these definitions and the idea of small temperature changes at each step, it seems natural to introduce a measure of the rate of change of the equilibrium distribution with change in temperature, i.e. a "sensitivity", as follows:

$$\mathcal{S} := \lim_{\Delta\beta \to 0} \left\{ \frac{\mathcal{D}\left\{\rho_{eq}(r;\beta),\ \rho_{eq}(r;\beta+\Delta\beta)\right\}}{\Delta\beta} \right\}, \tag{3.3}$$

It can be anticipated that it will be appropriate to cool more slowly when the value of this sensitivity parameter $\mathcal{S}$ becomes larger. To quantify such an expectation, it is necessary to also have some measure of how much each cycle reduces the distance to equilibrium. Accordingly, define a reduction factor $\mathcal{R}$ by

$$\mathcal{R} := \lim_{\Delta\beta \to 0} \frac{\mathcal{D}\left\{\rho^{[m+1]}(r),\ \rho_{eq}(r;\beta+\Delta\beta)\right\}}{\mathcal{D}\left\{\rho^{[m]}(r),\ \rho_{eq}(r;\beta+\Delta\beta)\right\}}, \tag{3.4}$$

where it is to be assumed that $\rho^{[m]}(r)$ is equal to $\rho_{eq}(r;\beta)$ and the reciprocal temperature for the current step is equal to $\beta+\Delta\beta$. In fact, $\rho^{[m]}(r)$ will not be equal to $\rho_{eq}(r;\beta)$ in practice, however, it seems reasonable to expect that $\rho^{[m]}(r) - \rho_{eq}(r;\beta+\Delta\beta)$ is roughly equal to $\gamma\{\ \rho_{eq}(r;\beta) - \rho_{eq}(r;\beta+\Delta\beta)\ \}$, for some constant γ (presumably greater than one). With this more realistic form for $\rho^{[m]}(r)$, i.e. $\rho^{[m]}(r) = \gamma\ \rho_{eq}(r;\beta) + (1-\gamma)\ \rho_{eq}(r;\beta+\Delta\beta)$, it is readily shown that, as a consequence of homogeneity, the resulting value of the reduction factor is identical to the simpler form defined above, for any non-zero value of γ.

If each cycle of the annealing process is successfully drawing the occupation density toward the equilibrium density, the reduction factor defined above must take values between zero (instant equilibration) and one (infinite time to equilibrate). It is now proposed that the step distribution should be chosen to minimize the value of $\mathcal{R}$ at each stage of the process. That is, equilibration should be as rapid as possible — in keeping with the heuristic and the desire to reduce the temperature as quickly as possible. This means that there is now a well defined criterion for assessing step generation algorithms. Furthermore, the reduction factor, the sensitivity, and the statement of the heuristic given in Eq.(3.2) can be used together to deduce that, at each cycle, the reciprocal temperature should be increased by no more than

$$\Delta\beta = \frac{\varepsilon\,(1-\mathcal{R})}{\mathcal{S}}. \tag{3.5}$$

This follows simply from the fact that the equilibrium distribution moves a distance $\mathcal{S}\Delta\beta$ when the reciprocal temperature is changed by $\Delta\beta$ and that, assuming the distance to equilibrium was about ε, one cycle of the algorithm will reduce this distance by a factor of $\mathcal{R}$. The skeleton of an adaptive algorithm is now apparent: start with β set equal to zero, generate steps chosen to minimize $\mathcal{R}$ and, at each cycle, change the temperature by the value specified in Eq.(3.5). There is only one missing piece at this stage: a means to estimate $\mathcal{R}$ and $\mathcal{S}$ "on the fly".

It is a simple exercise to expand the form defining $\mathcal{S}$ in Eq.(3.3) as a Taylor series in $\Delta\beta$, use the definition of distance given in Eq.(3.1), and the form of the equilibrium distribution specified in Eq.(2.3) in order to find the following exact result:

$$\mathcal{S} = \tfrac{1}{2}\int \rho_{eq}(r;\beta)\left|f(r)-<f>\right| d\tau, \tag{3.6}$$

where $<f>$ is defined by

$$<f> := \int \rho_{eq}(r;\beta)\, f(r)\, d\tau. \tag{3.7}$$

Similarly, Eq.(3.4) can be reduced to

$$\mathcal{R} = \frac{\int \rho_{eq}(r;\beta)\left|f(r)-<f>+\int[f(r')-f(r)]S(r,r')A(r,r')d\tau'\right| d\tau}{\int \rho_{eq}(r;\beta)\left|f(r)-<f>\right| d\tau}. \tag{3.8}$$

The key observation here is that each of these three expressions can be estimated during the annealing process although this requires some explanation.

Notice that all but one of the integrals appearing in these equations take the form of an integral over the configuration space of a function of the value of the merit function multiplied by the equilibrium density. Integrals like this can be readily estimated if the walk is *ergodic*: So far, the occupation and equilibrium densities have been defined by reference to a host of independent runs of the algorithm. Now consider a single computer executing one such walk about the configuration space at fixed temperature. The walk is said to be ergodic if the occupation density defined in terms of the recorded history of the current point for this one system tends to the same equilibrium density defined above. It is expected that [under weak conditions on the merit function and step distribution like those brushed aside following Eq.(2.4)] this will be so here — all that has happened is that the observations from many separate runs have been replaced by many observations from the one run and remember that the random aspect scrambles any information about the past for each such walk. That is, where the walk will end up after N more steps is essentially independent of the current point provided N is large enough. This being so, an integral like the one appearing in the definition of $<f>$ is readily evaluated by simply averaging over the current value of the merit function at each cycle of the walk. Since the temperature is gradually changing, it is necessary to average only over a set of the most recent values. This can be done conveniently by using exponentially decaying weights in the averaging process. A similar approach can be taken to the estimation of $\mathcal{R}$ and $\mathcal{S}$ which, as indicated above, form the basis of the adaptive control in our simulated annealing algorithm.

A significant simplification in the step generation process follows if the covariance matrix of the generated steps is required to be proportional to the covariance of either the accepted steps or the equilibrium distribution. This virtually ensures that the algorithm is invariant under linear changes in the coordinate system. There is a number of ways to achieve this and we have chosen to use an n-dimensional Gaussian distribution whose covariance is coupled to the accepted steps. This is realized by relying on the central limit theorem and simply forming new steps by making random linear combinations of previously accepted steps which leaves only the overall scale factor free. This scale factor is controlled in a fashion derived from the requirement that $\mathcal{R}$ be minimized. The remaining details of the adaptive simulated annealing algorithm are to be presented elsewhere. In closing, it is mentioned that, in terms of the statistics gathering [especially for

the innermost integral in Eq.(3.8)] and reducing the overall run time, there are significant advantages if the algorithm is implemented in a multiprocessing environment so that, from each current point, multiple random steps can be investigated at the one time.

4 Preliminary results for the "monochromatic quartet"

In the implementation of any global optimization process for lens design, there are several important issues that must be dealt with at the outset. First, since many thousands — perhaps millions — of optical systems will be assessed in the process, it is vital that the evaluation of the merit function be done as efficiently as possible. For a problem like the monochromatic quartet, using Gaussian quadrature[4] is surely the most effective approach. Second, for random walks in the configuration space like the type proposed here, it is crucial to explicitly enforce as many of the constraints as possible. For example, choosing a set of curvatures and thicknesses at random is virtually guaranteed to give an impractical system where edge thicknesses may be negative or there may be such a high degree of vignetting that essentially no light gets through. Accordingly, we have chosen variables which guarantee that almost all of the generated systems are feasible. For instance, to avoid vignetting problems, the surface curvatures may be defined with respect to the (first order) unvignetted aperture radius, say A, in the associated plane, etc. In the interest of efficiency and fair sampling, it is also crucial to devise a (linear coordinate transformation independent) procedure for reflecting wayward steps off violated constraints to force new trial points to lie always within the region of interest. For instance, if a generated step causes an element to have negative thickness on axis, it is appropriate to add a component proportional to the gradient of the constraint (with respect to the natural metric associated with the accepted steps) in order to reflect this step back into the region of interest.

Another important point to realize is that non-linear transformations of the variables (e.g. using surface radii in place of the curvatures) fundamentally changes the optimization problem — a totally random walk in the configuration space will have an entirely different occupation density for different sets of coordinates. Furthermore, using geographic terminology, the merit function terrain may be smooth and rolling with wide valleys surrounding the lowest minima in one set of coordinates, while this same problem may have steep curved gulleys with the lowest minima hidden as "sink holes up in the hills" when viewed using another set of coordinates. Clearly, a global optimizer would have a significantly better chance of finding the global minimum with the former coordinate set. Furthermore, since the global minimum of $f(r)$ is also the global minimum of any function of the form $M[f(r)]$ provided M is a monotonic, increasing function, there is another set of transformations of a particular problem that must be considered. It is crucial to realize that the equilibrium densities for the annealing algorithm are fundamentally different when a non-linear transformation of this type is applied to the merit function. For example, is it better to optimize using a mean square criterion, a root mean square criterion or some other transformation of the merit function?

We have not had the opportunity to explore coordinate transformations or merit function transformations in any systematic way since we are still in the process of fine-tuning the free parameters that remain in the adaptive simulated annealing algorithm. Furthermore, we have not explored the option of incorporating a regular downhill optimizer which may be invoked to start at any promising intermediate configurations. Ideally all these considerations should be combined into the one investigation since, no doubt, there are interdependencies here. Having done all this, it would then perhaps be clear whether it is best to have a small number of extensive runs (with the value of ε taken to be as small as possible while remaining within the constraints imposed by a limited computational budget) or many quicker, independent runs.

We have implemented adaptive simulated annealing on an accelerated Macintosh II. The computer is capable of multiprocessing since it incorporates eight "transputer" processors (made by Inmos) giving us a throughput comparable to

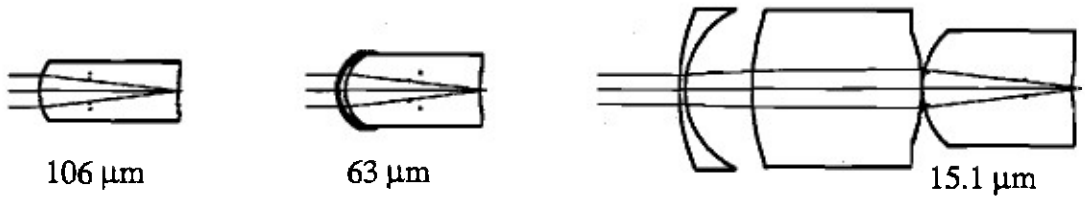


Fig. 2 The best one, two, and three element configurations found by annealing are presented with the associated spot sizes (measured as in the statement of the monochromatic quartet problem). The entrance pupils and stops are shown as short dashes here and in Fig. 3.

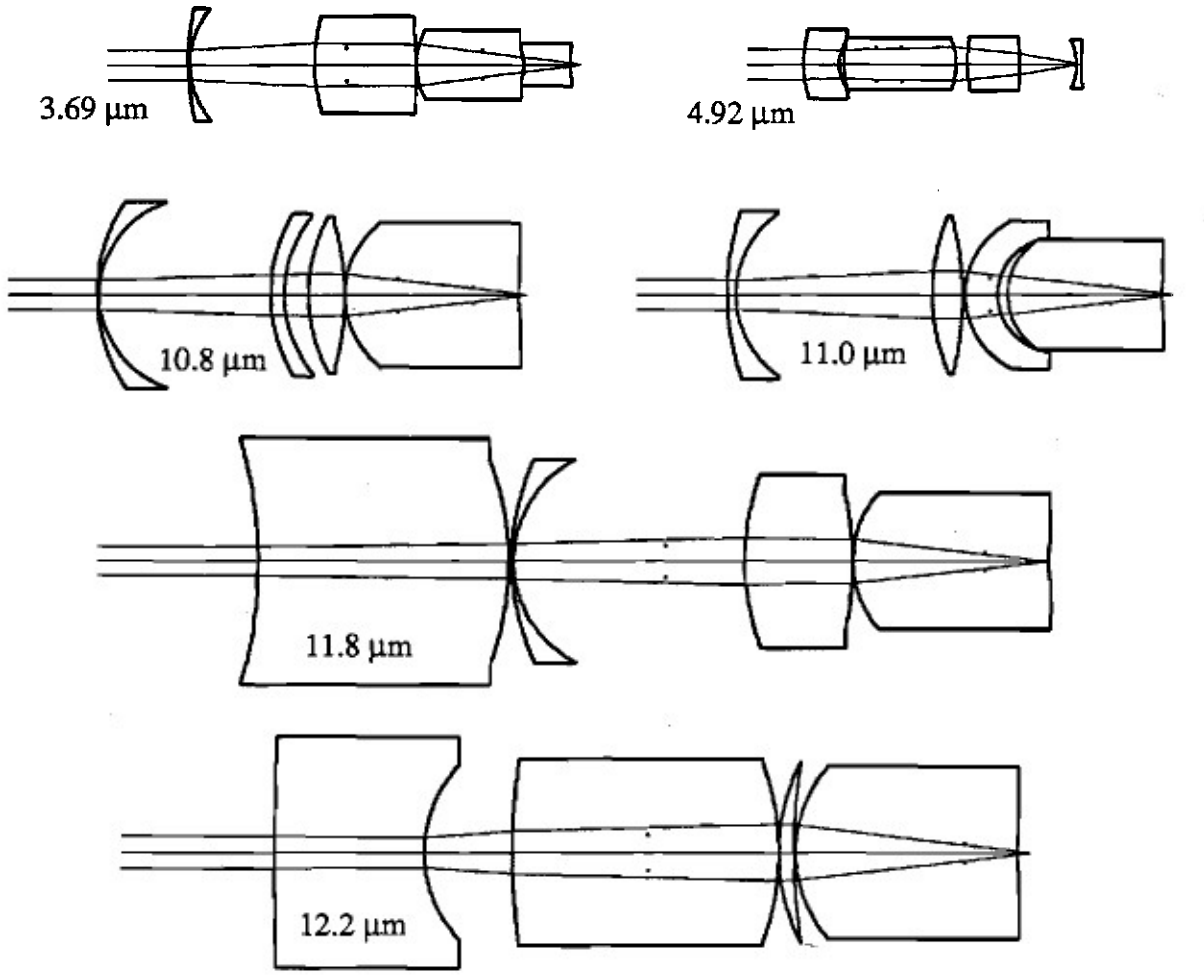


Fig. 3 A sample of some of the better monochromatic quartets arrived at by adaptive simulated annealing.

the newly released RISC workstations that are available from a variety of computer firms. The option of parallel processing suits this application well. In order to interpret the results of the algorithm in the case of the monochromatic quartet, we began with the same problem except the number of elements was first reduced to one and then increased to two and three. That is, the focal length, f-number, field angle, merit function, etc were all unchanged. We executed a number of runs on each of these simpler problems and fairly consistently came up with the three best systems shown in Fig. 2. These results were quite encouraging and provided us with a scale for the results in the four element case: a quartet with a spot size in excess of 15μm can be regarded as a poor solution. But what is the mark of a good solution? A sample of the results we found for the monochromatic quartet are presented in Fig. 3. These were evaluated in overnight runs where the cooling parameter ε was set so that each night would allow up to about six separate attempts at the problem — a new run starting when the preceding run had reduced the temperature to a level that the associated changes in the merit function were no longer significant. There was a wide variety of solutions with many interesting forms giving spot sizes in excess of 15μm. It turns out that the better solutions presented here are within a factor of two of the best solutions reported at the conference and this is an encouraging result considering the fact that there is much left for us to investigate in this context.

5 Concluding remarks

There are many aspects of the annealing process which we are yet to explore fully in the context of lens design. Most important among these is the investigation of the significance of specific transformations of the lens design merit functions and coordinates (which include alternative schemes for enforcing such things as focal length constraints). In a broader context, we are yet to consider the incorporation of downhill optimizers. Even so, the early results are very encouraging. The three element system found reliably by adaptive simulated annealing was confirmed in a personal communication from David Shafer (the creator of the monochromatic quartet problem) to coincide with the best solution known to him. A couple of the monochromatic quartet solutions are within about a factor of two (of the 2μm spot size) of the best systems reported at this meeting. The variety of different systems with spot sizes around 10-20μm that was found by annealing suggests that lens design problems are sure to offer an attractive challenge to any advocates of particular global optimization algorithms. There is a wide variety of algorithms that have been proposed for global optimization and there are many aspects to be addressed in their successful implementation in any given context. We would be delighted if, as in years gone by, the field of lens design becomes the arena for the development of state-of-the-art optimization algorithms.

Acknowledgements

This work was supported in part by the University Research Initiative of the U.S. Army Research Office.

References

1. David Vanderbilt and Steven G. Louie, "A Monte Carlo Simulated Annealing Approach to Optimization over Continuous Variables", J. Comp. Phys. **56**, 259-271, 1984.
2. See for example: A. Corana, et. al., "Minimizing Multimodal Functions of Continuous Variables with the Simulated Annealing Algorithm", ACM Trans. Math. Softw. **13**, 262-280, 1987.
3. See for example: V. K. Viswanathan, et. al., "An attempt to develop an 'intelligent' lens design program", Proceedings of SPIE **554**, 10-17, 1985.
4. G. W. Forbes, "Optical system assessment for design: numerical ray tracing in the Gaussian pupil", J. Opt. Soc. Am. **A 5**, 1943-1956, 1988 and **A 6**, 1123, 1989. Only about twenty rays were necessary in the assessment of quartet solutions: something like four rays on-axis and nine at the off-axis field points is certainly adequate.

LIST OF ATTENDEES

WORKSHOP ON GLOBAL OPTIMIZATION WITH ADAPTIVE SIMULATED ANNEALING

NAME	ORGANIZATION	PHONE NO.
RON ANTOS	U of R	(716)275-4179
ANDREW JONES	U of R	(716)275-6205
KAIWAN SAFAVI	C2NVEO	(703)664-6665
JOHN HALL	C2NVEO	(703)664-6665
BAHRAM ZANDI	C2NVEO	(703)664-1935
GEOFF SAWYER	C2NVEO	(703)664-6066
TOM COLANDENE	C2NVEO	(703)664-6665
QUOCHIEN VUONG	C2NVEO	(703)664-4127
GREG SALAMO	C2NVEO	(703)664-1432
GREG FORBES	U of R	(716)275-7227
FRED SEMENDY	C2NVEO	(703)664-1935
MARTIN LAHART	C2NVEO	(703)664-5207
DUC NGUYEN	C2NVEO	(703)664-5635
DANH PHUNG	C2NVEO	(703)664-5207
ROBERT SPANDE	C2NVEO	(703)664-6665