ON ARBITRARILY PERFECT IMAGERY WITH A FINITE APERTURE

OPTICAL SCIENCES

THE UNIVERSITY OF ARIZONA
TUCSON, ARIZONA

Technical Report 34

ON ARBITRARILY PERFECT IMAGERY WITH A FINITE APERTURE

B. Roy Frieden

15 January 1969

OPTICAL SCIENCES

THE UNIVERSITY OF ARIZONA
TUCSON, ARIZONA

Technical Report 34

ON ARBITRARILY PERFECT IMAGERY WITH A FINITE APERTURE

B. Roy Frieden

15 January 1969

ABSTRACT

Despite its necessarily finite aperture, an optical system can theoretically be coated to produce arbitrarily perfect imagery over a limited field. When the object is of limited extent, this field can be made the optical conjugate to the object, so that the whole object is imaged with arbitrary precision.

The required pupil coating approximates low-contrast cosine fringes over its central region; toward the aperture edge the frequency and amplitude rapidly accelerate. The maximum occurs as a narrow spike.

The frequency near the central region varies directly with the total extent of the conjugate field and inversely with the required central core width Δ in the point amplitude response. As Δ is made arbitrarily narrow, the point amplitude response approaches the form of a sinc function over the field of view. This function is precisely the point amplitude for a diffraction-limited pupil with a magnified aperture of $1/\Delta$ times the given pupil aperture! The only image property that is not in compliance with this effective aperture magnification is that of total illumination. This is severely reduced from that of the original, uncoated aperture, and is the major restriction on practical use of the derived pupil.

Applications to microscopy and telescopy are discussed.

TABLE OF CONTENTS

INTRODUCTION

The image produced by an optical system can be degraded in various ways, such as by motion of the object or the optical system, and some work has been done at improving such images by processing of the image itself. An image is also degraded by intrinsic characteristics of the lens. Correction of this type of degradation can be done by processing the image, as above, and in theory such image processing can restore the object scene perfectly[1-4]. In practice, however, such a restoration is inaccurate because of measurement errors in the image[5]. Furthermore, much computing time is required, and the restoration is not available until some time after the image is first obtained.

It would be of advantage to be able to make the perfect restoration as the image forms. Conceivably this could be accomplished by coating the lens with a particular transmission pattern. Some years ago Toraldo di Francia[6] worked on a similar problem. He showed that this method might be feasible, but his paper is relatively inaccessible and most optical workers have been unaware of it and its promise.

In this paper, we investigate the coating that would be required to produce arbitrarily perfect imagery. This is done for the usual cases of interest: (a) when the object is of limited extent and radiates either totally coherently or totally incoherently, (b) when the image and object are connected by a convolution, and (c) when the Fraunhofer approximation holds.

Unfortunately, we find that the perfect restoration is achieved only at the expense of a considerable loss of illumination.

PHYSICAL ASSUMPTIONS

Limited field

Suppose an object is radiating either coherently or incoherently. We let o represent the object amplitude for the object radiating coherently, or radiance for the object radiating incoherently. We assume the observer knows that o is of limited extent x_0, so that if x is the distance coordinate,

$$o(x) = 0 \quad \text{if } |x| > x_0/2. \tag{1}$$

If x_0 is small enough, a situation of isoplanatism exists, and image i is formed as a convolution[7]:

$$i(x) = \int_{-x_0/2}^{x_0/2} dx' \, o(x') \, s(x-x'). \tag{2}$$

For simplicity, unit magnification is assumed throughout this paper. In Eq. (2), function s is the optical point response, defined as

$$s(x) = \begin{cases} a(x) & \text{for o coherent} \\ |a(x)|^2 & \text{for o incoherent.} \end{cases} \tag{3}$$

Function a(x) is the point amplitude response for the optics.

A condition of perfect image formation exists when

$$i(x) = C\, o(x) \tag{4a}$$

over the field of view

$$|x| \le x_0/2. \tag{4b}$$

C represents a uniform attenuation over the field of view if $C < 1$, or a uniform gain if $C > 1$. Remember that equality (4a) need be satisfied only over the *limited* field (4b). The ramifications of this situation are discussed next.

With an arbitrary object, the image (2) meets requirement (4a) only if optical response

$$s(x) = C\,\delta(x), \tag{5a}$$

where $\delta(x)$ is the Dirac delta function.[7] Identity (4a) follows from substitution of Eq. (5a) into convolution (2), with use of the "sifting property"[7] of the delta function. Now, by the limited field requirement (4b) and convolution (2), optical response $s(x)$ only contributes to the image at all $|x| \le x_0$. The strong requirement (5a) is now somewhat weakened. It need hold only over the limited field

$$|x| \le x_0. \tag{5b}$$

It will be shown that any finite x_0 allows the problem to be solved. A pupil function exists whose corresponding response $s(x)$ is arbitrarily close to $C\,\delta(x)$ over the field of view $|x| \le x_0$.

State of coherence

Both coherent and incoherent imagery can be simultaneously optimized by one pupil if point amplitude $a(x)$ satisfies Eq. (5). That is,

$$a(x) = K\,\delta(x) \qquad \text{for } |x| \le x_0\ . \tag{6}$$

This is because of identities (2) and (3), and the fact that $[\delta(x)]^2$ must be proportional to $\delta(x)$, as shown below.

Function $\delta(x)$ is defined by two conditions[7]:

$$\delta(x) = 0 \qquad \text{for } x \neq 0 \tag{7a}$$

and

$$\int_{-\varepsilon}^{\varepsilon} dx\,\delta(x) = 1 \qquad \text{for any } \varepsilon > 0. \tag{7b}$$

The square of Eq. (7a) is the same identity in $[\delta(x)]^2$. Hence, $[\delta(x)]^2$ satisfies (7a). Also, since $[\delta(x)]^2 >> \delta(x)$ for $x = 0$, certainly

$$\int_{-\varepsilon}^{\varepsilon} dx[\delta(x)]^2 > 0. \tag{8}$$

Hence, $[\delta(x)]^2$ satisfies requirement (7b) except for a multiplicative constant. Therefore $[\delta(x)]^2$ must itself be proportional to $\delta(x)$. In summary, we may conclude that $\delta(x)$ is infinitely sharp, but $[\delta(x)]^2$ is sharper yet. Or physically, if we can make amplitude $a(x) \simeq \delta(x)$, then $s(x)$ approximates $\delta(x)$ even better.

TORALDO DI FRANCIA'S APPROACH

Before embarking on the derivation, it is useful to review Toraldo di Francia's work[6] on a similar problem. He found that a pupil of concentric, equispaced rings, of alternating phase, makes the first dark ring of s arbitrarily wide, while maintaining a fixed central core width. This design problem resembles requirements (5a) and (5b) if x_o is taken as the outside radius of the first dark ring. Toraldo di Francia was able to systematically widen x_o, keeping the sidelobes within x_o relatively small, by increasing the number of pupil rings to five. However, this technique probably cannot be used to solve the converse problem of arbitrarily reducing the energy in the sidelobes for a *fixed* x_o. This is because the point amplitude contribution from his ℓth pupil ring is the Bessel function $J_o(\ell x/L)$, where $\ell = 1, 2, \ldots, L$. But functions $J_o(\ell x/L)$ are not a complete set, so no superposition of them, in image space, could approximate a delta function. If s(x) were expanded as a series of complete functions (as we do here), then it might approach $\delta(x)$ as the number of series terms is indefinitely increased.

The existence of Toraldo di Francia's pupil shows that the solution found in this paper is not so radical as it seems. Both solutions are similar in being extremely wasteful of illumination, most of which is thrown outside the vield of view, Eq. (5b).

DERIVATION OF REQUIRED PUPIL

We now find the optical pupil function that accomplishes the requirements (6) of perfect imagery. In the Fraunhofer approximation,[7] a(x) is related to the optical pupil function $U(\beta)$ through

$$a(x) = \int_{-\beta_o}^{\beta_o} d\beta\, U(\beta)\, e^{j\beta x}, \quad j = (-1)^{\frac{1}{2}}. \tag{9}$$

Coordinate β is $2\pi/\lambda R$ times the pupil distance coordinate, where λ is the radiation wavelength and R is the Gaussian image distance. β_o defines the pupil extent, a finite number for any real optical system.

We seek the pupil $U(\beta)$ which, through Eq. (9), satisfies design requirements (6). We previously noted that requirements (6) can be satisfied only if a(x) is physically an expansion of complete functions. Hence, the problem is to choose a series of functions for $U(\beta)$ such that its finite Fourier transform, a(x) in Eq. (9), is a series of complete functions. Luckily, such a representation exists.

Over the pupil $|\beta| \le \beta_o$, let amplitude

$$U(\beta) = \sum_{n=0}^{\infty} b_n\, \psi_n(c, \beta x_o/\beta_o), \tag{10a}$$

$$c = \beta_o x_o, \tag{10b}$$

where ψ_n is the nth prolate spheroidal wave function.[8] By Eq. (10b), c is thus the space-bandwidth product for the system design. For brevity, we drop the notation c from the argument of ψ_n. Coefficients b_n are to be found.

Functions $\psi_n(x)$ are normalized "angular functions"[8] $S_{on}(x)$:

$$\psi_n(x) = (\lambda_n/N_n)^{\frac{1}{2}}\, S_{on}(x/x_o). \tag{11a}$$

Normalization coefficients

$$N_n = \int_{-1}^{1} dt[S_{on}(t)]^2 \tag{11b}$$

and the λ_n are known eigenvalues. All quantities in Eqs. (11) are functions of c.

Functions $\psi_n(x)$ have precisely the properties we require above. They are complete[8] over the field $|x| \leq x_o$, i.e.

$$\sum_{n=0}^{\infty} \lambda_n^{-1} \psi_n(0)\ \psi_n(x) = \delta(x) \quad \text{for } |x| \leq x_o. \tag{12a}$$

Also, the finite Fourier transform[8] of ψ_n is again ψ_n.

$$\int_{-\beta_o}^{\beta_o} d\beta\ \psi_n(\beta x_o/\beta_o)e^{j\beta x} = j^n(2\pi\lambda_n\beta_o/x_o)^{\frac{1}{2}}\ \psi_n(x). \tag{12b}$$

By substituting representation (10a) into Eq. (9), switching orders of integration and summation, and using identity (12b), we find

$$a(x) = (2\pi\beta_o/x_o)^{\frac{1}{2}} \sum_{n=0}^{\infty} b_n\ j^n\ \lambda_n^{\frac{1}{2}}\ \psi_n(x). \tag{13}$$

Comparing Eq. (13) with the completeness property (12a), we note that design requirements (6) can indeed be met if each

$$b_n = (x_o/2\pi\beta_o)^{\frac{1}{2}}\ j^{-n}\ \lambda_n^{-3/2}\ \psi_n(0). \tag{14}$$

The problem is solved.

The required pupil for meeting requirements (6) is found by substituting coefficients (14) back into series (10a). Using the fact that[8,9]

$$\psi_n(0) = 0 \quad \text{for } n \text{ odd}, \tag{15}$$

we find a pupil

$$U(\beta) = (x_o/2\pi\beta_o)^{\frac{1}{2}} \sum_{n(\text{even})=0}^{\infty} (-1)^{n/2} \lambda_n^{-3/2} \psi_n(0) \psi_n(\beta x_o/\beta_o). \tag{16a}$$

The resulting point amplitude a(x) obeys

$$a(x) = \sum_{n(\text{even})=0}^{\infty} \lambda_n^{-1} \psi_n(0) \psi_n(x), \tag{16b}$$

which, according to identity (12a), is just $\delta(x)$ within the field of view $|x| \le x_o$. We shall see below that, outside the field of view, a(x) departs maximally from delta function behavior, having incredibly large sidelobes.

PROPERTIES OF THE PUPIL AND POINT AMPLITUDE

By Eq. (16a) and the evenness of ψ_n for n even,[8,9] pupil $U(\beta)$ is purely real and even, but not necessarily positive everywhere. Hence, the pupil contains zones of zero and π phase values. The necessity for π phase changes is in accord with the discovery of Gal'pern[10] that a purely positive pupil cannot produce an arbitrarily narrow central core in $a(x)$.

Eq. (16b) shows that $a(x)$ is even in x. In order for $U(\beta)$ to represent real optics, passive or active, it must be finite everywhere. But, for the infinite number of terms indicated in Eq. (16a), $U(\beta)$ must be singular at least at one point. This follows from observation of Eq. (9): If we demand that $a(x) = K\,\delta(x)$, then $a(0)$ must be singular, and since β_o is finite, $U(\beta)$ must contain a singularity somewhere.

Hence, in order to arrive at a $U(\beta)$ that is everywhere finite, series (16a) must be truncated at finite n. By Eq. (9), the effect upon $a(x)$ is likewise truncation at the same n. Hence, we now define the truncated quantities

$$U_N(\beta) = (x_o/2\pi\beta_o)^{\frac{1}{2}} \sum_{n(\text{even})=0}^{N} (-1)^{n/2}\, \lambda_n^{-3/2}\, \psi_n(0)\, \psi_n(\beta x_o/\beta_o) \tag{17a}$$

and

$$a_N(x) = \sum_{n(\text{even})=0}^{N} \lambda_n^{-1}\, \psi_n(0)\, \psi_n(x). \tag{17b}$$

Both $U_N(\beta)$ and $a_N(x)$ are observed to be even functions.

<u>Active optics</u>

If we physically allow values of $U_N(\beta) > 1$, which assumes an active optical system, then pupil (17a) may be directly used.

Aside from the desirable delta-function properties of $a_N(x)$ (shown below), there is no light loss due to attenuation in the pupil. Such light loss will be shown to be the major limitation to the use of passive optics in this manner. Hence, pupil (17a) is one of the best possible uses for an active optical pupil, once active optics are physically realizable.

Passive optics

In the meantime, we must be content with examination of the role of passive optics in our scheme. Numerical evaluation of Eq. (17a) shows that $U_N(\beta)$ is maximum in absolute value at $\beta = \beta_o$. Therefore, we can define a passive pupil

$$P_N(\beta) = U_N(\beta)/U_N(\beta_o). \tag{18a}$$

For purposes of analyzing $a_N(x)$, it is convenient to examine the normalized point amplitude

$$A_N(x) \equiv a_N(x)/a_N(0). \tag{18b}$$

The efficiency of $P_N(\beta)$ in transmitting light into the central core of $A_N(x)$ may be measured by the Strehl flux ratio.[7] By use of Eqs. (9) and (18a),

$$S \equiv [a_N(0) / 2\beta_o U_N(\beta_o)]^2. \tag{18c}$$

All quantities in Eqs. (18) may be related to known quantities ψ_n and λ_n . through Eqs. (17). Numerical analysis of $P_N(\beta)$, $A_N(x)$, and S was performed on a CDC 6400 computer.

Point amplitude

Figure 1 shows $A_N(x)$ for the case c = 6.25, with N = 40.

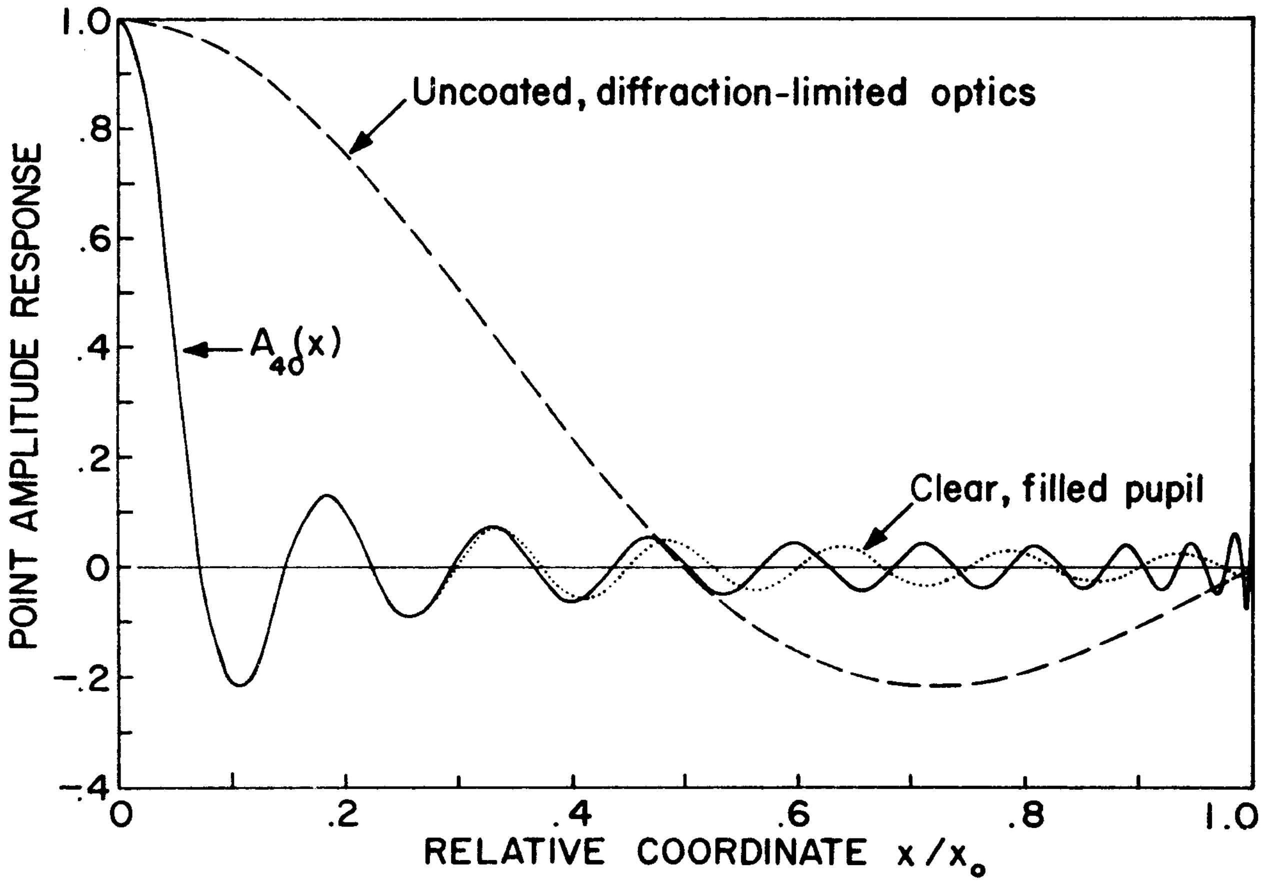

Fig. 1. Point amplitude response $A_{40}(x)$ *for space-bandwidth of c = 6.25. Dashed curve is the point amplitude for uncoated, diffraction-limited optics of same physical aperture as* $A_{40}(x)$. *Comparison of the two central core widths shows an effective aperture magnification of about 6.67 in* $A_{40}(x)$. *The dotted curve is the amplitude response from a clear, filled pupil of physical extent equal to the effective aperture extent of* $A_{40}(x)$. *To the accuracy of the plots, this curve and* $A_{40}(x)$ *are identical through the first two sidelobes.*

This is a typical $A_N(x)$ curve, and certain properties of $A_N(x)$ for general N and c may be drawn from it:

(1) There are exactly N zeros over the entire field $|x| \leq x_o$. (The half-field $x \leq x_o$ is shown.)

(2) Let Δ_N represent the total width of the central core. For $N \gtrsim 20$ we find the amazingly simple result that $\Delta_N = 6.0\ x_o/N$, independent of c. A simple measure of the resolution quality is the ratio of Δ_N to the central core width for uncoated optics of the same aperture β_o. The latter core width is $2\pi/\beta_o$. Thus, we define resolution enhancement

$$\delta_N \equiv \Delta_N \beta_o/2\pi. \tag{19a}$$

From the preceding identity we therefore have

$$\delta_N \simeq 3c/N\pi \tag{19b}$$

for $N \gtrsim 20$. This shows that the core width is made arbitrarily narrow as N is increased, which corroborates Wilkins' observation[11] on ultimate core width. We also see that, for given aperture size β_o and order N, δ_N directly increases with field x_o.

(3) The dashed curve in Fig. 1 is the point amplitude $A_o(x)$ due to uncoated optics of aperture β_o. The enhanced resolution due to $A_{40}(x)$ is immediately apparent. However, a much more important observation is that the *form* of $A_{40}(x)$ is very similar to that of $A_o(x)$, except for being compressed in x. In fact, we find that asymptotically as N increases

$$A_N(x) \rightarrow \mathrm{sinc}(\pi\beta_o'), \tag{20a}$$

with

$$\beta_o' = \beta_o/\delta_N, \tag{20b}$$

except for x in the vicinity of x_o. For example, in Fig. 1 the dotted curve is $\mathrm{sinc}(\pi\beta_o/\delta_{40})$, with $\delta_{40} = 0.149$. We note that the dotted curve agrees perfectly with $A_{40}(x)$, to the accuracy of the plot, for all $0 \le x/x_o \le 0.25$.

Farther out, the difference is mainly an increasing shift of phase between the two curves. Thus, the sidelobes of $A_{40}(x)$ decrease nearly as $1/x$, which is both convenient to know and a good image-forming property. For incoherent illumination the sidelobes fall off as $1/x^2$, an even better situation.

Since we know now that $A_N(x)$ is a sinc function for N large, we can see the manner in which $A_N(x) \rightarrow \delta(x)$ as N increases. The sidelobes do not decrease in amplitude, but rather become increasingly compressed about the origin, so that the total radiant energy in the sidelobes approaches zero. Thus, the image contributions from the sidelobes become decreasingly energetic, and the imagery approaches perfection.

Figure 1 shows that as $x/x_0 \rightarrow 1$, the sidelobes become increasingly compressed. These are the last small oscillations of $A_N(x)$ before a mammoth sidelobe has its beginnings at about $x/x_0 = 0.995$. Before this sidelobe can attain any great magnitude, it is cut off by the field edge $x = x_0$ at about 0.2. Considering the great slope of this sidelobe near $x = x_0$, it is quite important that the value x_0 not be exceeded in the field of view. Otherwise, this sidelobe would swamp the useful $A_N(x)$ distribution for $|x| \leq x_0$ and would cause very imperfect image formation.

In summary, pupil $P_N(\beta)$ causes a point amplitude which, for $|x| < x_0$, acts in almost all respects like the point amplitude from a "magnified" aperture (20b). Moreover, by Eq. (19b), the magnification increases arbitrarily and linearly with N. Thus, neglecting the question of total illumination for the moment, a given small aperture may be made to form images as if it were a much larger and "filled" aperture. This ultimate goal of aperture synthesis[12] is now seen to be a theoretical possibility.

The pupil

Figure 2 shows the pupil $P_{40}(\beta)$ for the case $c = 6.25$. This pupil causes point amplitude $A_{40}(x)$ in Fig. 1.

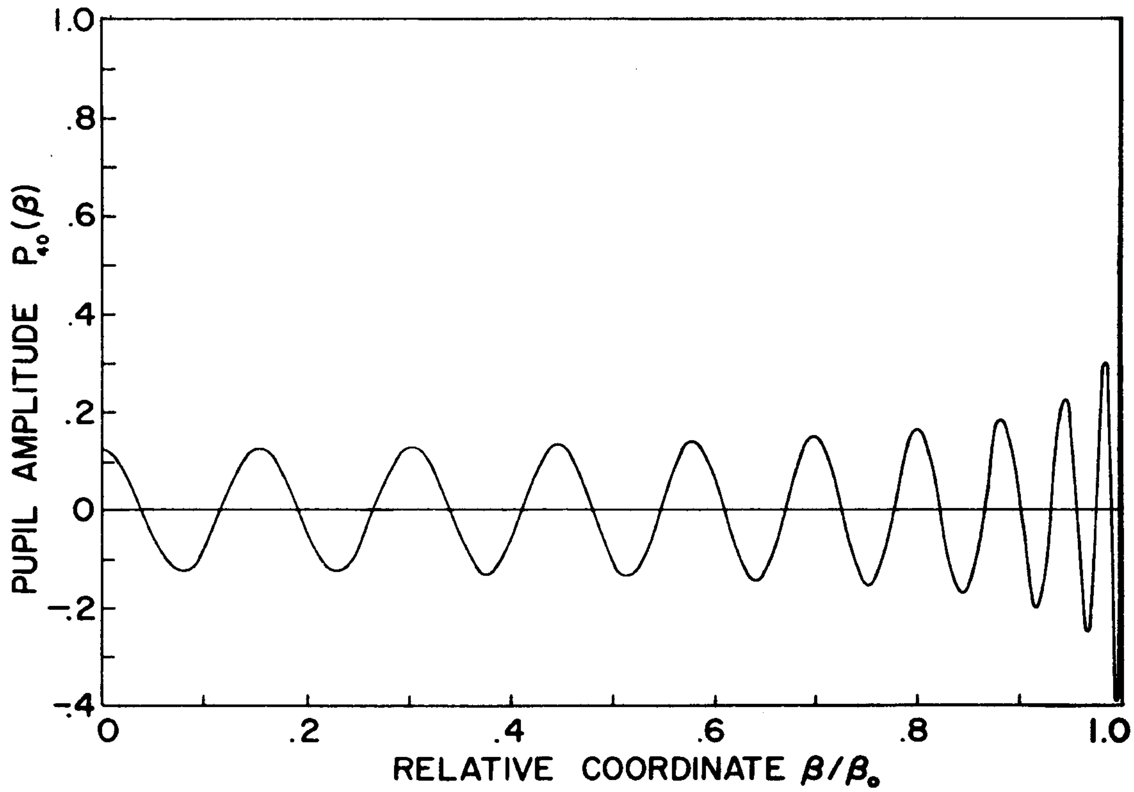

Fig. 2. Pupil function $P_{40}(\beta)$. *For values* $|\beta/\beta_o| \le 0.3$, $P_{40}(\beta)$ *is well approximated by* $\cos(\beta x_o/\delta_{40})$. *Both frequency and amplitude increase rapidly once* (β/β_o) *exceeds about 0.8.*

Pupil $P_{40}(\beta)$ is typical of the general pupil $P_N(\beta)$ with arbitrary N and c, which has the following properties:

(1) There are exactly N zeros to $P_N(\beta)$ over the pupil $|\beta| \le \beta_o$. This results in N/2 regions each of zero and π phase shift.

(2) $P_N(\beta)$ attains its maximum of unity at the pupil edge.

(3) For $|\beta/\beta_o|$ small (for example, less than 0.3 in Fig. 2),

$$P_N(\beta) \propto (-1)^{N/2} \cos(\beta x_o/\delta_N). \tag{21}$$

The proportionality constant seems to be finite in the limit $N \to \infty$, but this has not yet been conclusively established. It might be zero. When $|\beta/\beta_o|$ is more than 0.3, $P_{40}(\beta)$ still resembles a cosine curve, but the frequency and amplitude noticeably increase as $|\beta/\beta_o|$ increases.

One reason that the largest sidelobes fall outside field of view $|x| \leq x_o$ now becomes apparent. Substitution of pupil (21) into Eq. (9) yields amplitude

$$\mathrm{sinc}(c/3)\,(x/x_o - 1/\delta_N) + \mathrm{sinc}(c/3)\,(x/x_o + 1/\delta_N). \qquad (22)$$

With δ_N typically on the order of 0.1, the main sidelobes are centered about $x/x_o = \pm 10$, that is, at 10 times the total field of view. Also, with c on the order of 10, the amplitude contribution (22) near the center of the field is merely 0.06 of the sidelobe amplitude, or 0.0036 of the energy. We shall see that other effects further diminish the total energy in the field of view.

(4) As $\beta/\beta_o \to 1$, $P_N(\beta)$ rapidly becomes a highly oscillatory function. The last oscillations at the pupil edge are quite narrow and steep; e.g. for $P_{40}(\beta)$, the last and greatest oscillation occurs within the confines $0.99 \leq \beta/\beta_o \leq 1.0$.

It might seem that these large oscillations cannot contribute much to $A_N(x)$ because they are so narrow and therefore nonenergetic. This was found to be untrue. A program was written for numerically taking the finite Fourier transform of $P_N(\beta)$ over intervals of $0 \leq |\beta| \leq (\beta_o - \Delta\beta)$, $\Delta\beta$ arbitrary. In particular, $\Delta\beta$ was the width of the last fluctuation in $P_N(\beta)$. Comparison of the resulting point amplitude with $A_N(x)$ showed large differences over the field of view. Use of this program with $\Delta\beta = 0$ showed perfect agreement with analytic formulas (17b) and (18b) for $A_N(x)$, corroborating derivations (9) through (16b) of pupil $P_N(\beta)$.

Strehl flux ratio

Expression (18c) for S was evaluated at values of $2 \le N \le 40$, with $0.25 \le c \le 10.0$. Since the main importance of point amplitude $A_N(x)$ lies in the arbitrary narrowness of its central core, which we have measured by the enhancement ratio δ_N, it is pertinent to find the variation of S with δ_N. Further, we have previously found that δ_N is increased (so that resolution is degraded) by an increase in field x_o. It would be useful to establish the tradeoff among parameters S, δ_N, and x_o, and this we do in Fig. 3. However, instead of x_o we use parameter c, assuming the pupil extent β_o to be fixed.

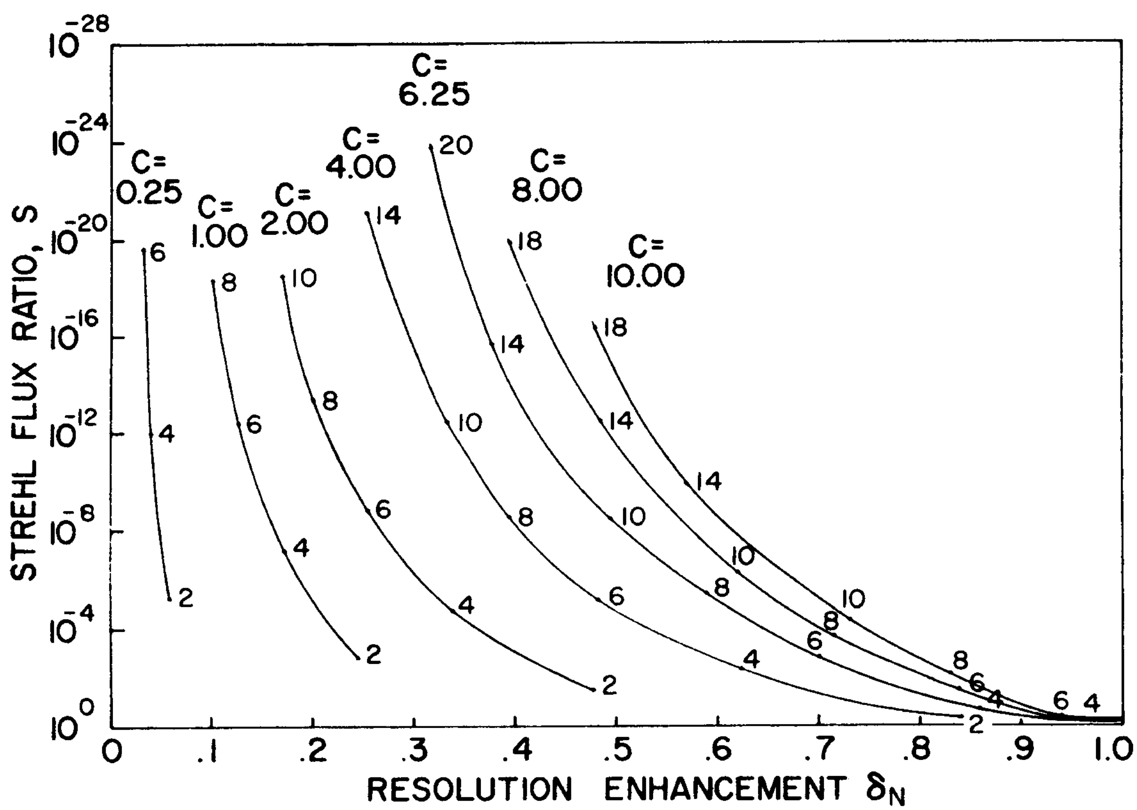

Fig. 3. Tradeoff among Strehl S, resolution enhancement δ_N, and space-bandwidth c, for $0.25 \le c \le 10.0$. Numbered points on the curves designate values of N. The curves show that for a required δ_N, S falls off by orders of magnitude with unit increments in c. Parameter c represents the field extent expressed alternatively as (1) a linear width of 2π times the number of Airy central core widths or (2) angu- extent of π times the number of light wavelengths within the pupil.

Independent of β_o, parameter c is directly related to the field extent as follows. Since the central (or Airy) core for the uncoated pupil β_o has width $\Delta_o = 2\pi/\beta_o$, and since there are x_o/Δ_o Airy cores in distance x_o, parameter c is 2π times the number of Airy cores in the field of view. Thus, an increment in c of 2π corresponds to the addition of an Airy core, i.e. another resolved point, to the field of view.

Figure 3 shows the family of curves S vs. δ_N for differing values of c. Each curve actually consists of a finite spacing of points, since the variation of S with δ_N is found by truncating the series $A_N(x)$ at discrete values of N. However, not all values of N up to 40 were used in the calculation, so we allow for interpolation to intermediate N-values by forming a smooth curve through points (S,δ_N). Each point used to establish the curves S vs. δ_N is indicated by its N-value.

The most important feature of Fig. 3 is the low values of S, once resolution δ_N is required to be less than 0.9. For any δ_N, we note that S monotonically decreases by orders of magnitude with each additional Airy core half-width (i.e., an increase in c of π) in the field of view. The upshot is that the simultaneous existence of superresolution and low sidelobes is paid for with 99.99% and more of the object radiation. Any practical use of pupils $P_N(\beta)$ therefore depends critically on use of a sufficiently (a) small field of view, (b) bright source, (c) long time exposure, and (d) sensitive detector.

A value of S also indicates the required gain in the equivalent active pupil. To maintain a Strehl of 1, the active pupil would have to obey

$$P_N(\beta) \;/\; \sqrt{S}\;. \tag{23}$$

The square root is, of course, a helpful effect. Observing Fig. 3, amplitude gains of 100 and upward would be required.

TWO-DIMENSIONAL RESULTS

A rectangular field of view may be imaged with arbitrary precision by a rectangular pupil

$$P_{MN}(\beta,\gamma) = U_M(\beta)U_N(\gamma) \,/\, U_M(\beta_o)U_N(\gamma_o), \qquad |\beta| \le \beta_o, \quad |\gamma| \le \gamma_o. \tag{24a}$$

The derivation is analogous to the preceding one for one dimension. Function U is the one-dimensional solution (17a), and (β,γ) are rectangular pupil coordinates. There are now two space-bandwidth products, $c_1 = \beta_o x_o$ and $c_2 = \gamma_o y_o$, with $c_1 = c_2$ for the square field case.

By using pupil (24a) and the two-dimensional analogy to Eq. (9), we now find a Strehl flux ratio

$$S' = [a_N(0) \,/\, 2\beta_o U_N(\beta_o)]^4 = S^2, \tag{24b}$$

where S was the one-dimensional result (18c). Equation (24b) assumes a square field of view and the use of $M = N$, which would be required for equal resolution enhancement in the x and y directions.

If rectangular fields of view are characteristic of photographic instruments, circular fields are characteristic of observational instruments. For this case, a circular pupil of the type (17a) must be designed. Since the "hyperspheroidal" functions[9] are the circular analogy to spheroidal functions, a pupil expansion in terms of the former is probably required. Further work on this problem awaits a representation of Dirac δ (of radius) in terms of these functions.

POSSIBLE APPLICATIONS

Despite the extremely small values of S in Fig. 3, and despite exponent 2 in Eq. (24b), we will now show that pupils $P_{NN}(\beta,\gamma)$ may be useful in certain square-field cases. The basis of such application is that a light beam with 1 W of power contains a flux of about 10^{18} photons/sec.

We will not examine the problem of *fabricating* such coatings, which should be considerable.

In the following applications and others, it is useful to know the angular field of view θ subtended by region $|x| \leq x_0$. It is easily shown that

$$\theta \quad = \quad (c/2\pi)(\lambda/r_0) \tag{25}$$

where $2r_0$ is the spatial extent of the square pupil. Hence, for values of $c \lesssim 10$, θ is very small when visible radiation is used.

Laser-illuminated microscope

A continuously radiating laser beam, with wavelength $\lambda = 500$ μm and irradiance $p = 0.25$ mW/mm^2, illuminates a specimen whose intensity distribution is desired. The microscope objective, with a square aperture of area $A = 4$ mm^2 and an f/number $F = 10$, is diffraction-limited. It is to be coated according to $P_{NN}(\beta,\gamma)$ for a resolution enhancement $\delta_N \simeq 0.60$ in both directions. In effect, the numerical aperture is to be nearly doubled. We also require the image to be bright enough to be photographed with an exposure time not exceeding 2 sec.

The central irradiance in the Airy disc for a square aperture is

$$I(0) \quad = \quad pA/(4\lambda^2F^2). \tag{26}$$

For the values given above, we find $I(0)$ = 10 W/mm^2. With λ = 500 μm, 1 W of light contains 2.5×10^{18} photons/sec, so that I(0) represents 2.5×10^{19} photons/mm^{2}sec.

For a typical film, the grain size is about 5.0 μm, and the slowest grain component requires 250 photons/grain for development. Therefore, I(0) represents 0.25×10^{13} times the number of photons needed for development of any one grain in 1 sec, and a Strehl attenuation of $S' \simeq 10^{-13}$ is required.

Fig. 3 showed that the coating $N = 10$ and $c = 8.0$ results in $\delta_N = 0.62$ and $S = 4.74 \times 10^{-7}$. Then, by Eq. (24b), $S' = 2.25 \times 10^{-13}$. Attenuation of our I(0) by this factor results in at least 0.56 developed grains/sec. Hence, an exposure time of 1.8 sec produces a developable photo.

By Eq. (25), angular field θ equals 2.2 arc. min. If the specimen is larger than this, it must be masked off with a 2.2 min (or smaller) aperture. Otherwise, the severe sidelobes of $A_{10}(x)$ for $|x| > x_0$ would obliterate the picture. With mask held fixed, the specimen may be moved about for piecewise observation of its entire image.

Telescope

The limitation on field of view cited above would be an artificial limitation on many uses of the astronomical telescope. For example, the angular field of interest when observing the diameter of Betelgeuse is 0.047 arc sec. Furthermore, if the instrumental field of view initially exceeds the permissible value θ of Eq. (25), it can be reduced to θ in certain cases by suitable baffling.

On the other hand, star irradiances are extremely small compared with the laser illumination cited above, so some other means of providing a large irradiance must be found. Of course, the collecting area of the telescope aperture provides just this effect.

Suppose a typical 6th magnitude star, with an irradiance of 2×10^{-14} W/mm^2, is to be observed with a resolution enhancement $\delta_N < 0.5$. The same light wavelength as above (now representing an average), film characteristics, and f/number are assumed. In addition, suppose the aperture is 10 inches on a side so that $A = 0.65 \times 10^5$ mm^2. Substitution of these parameters into Eq. (26) now produces an $I(0) = 0.31 \times 10^{14}$ photons/mm^{2}sec. This represents 0.31×10^7 times the number of photons required to develop the slowest grain in 1 sec.

Searching Fig. 3 for designs with $\delta_N < 0.5$, we note a value $\delta_N = 0.488$ for the case $c = 3$, $N = 4$, and $S = 4.91 \times 10^{-4}$. The net intensity resulting from use of this pupil coating is therefore $(0.31 \times 10^7)(4.91 \times 10^{-4})^2 = 0.75$ of the required irradiance to develop the slowest grain in 1 sec. Therefore, an exposure time of 1.3 sec results in a developable picture. Use of Eq. (25) discloses an angular field θ of 0.39 arc sec.

Hence, by use of pupil coating $P_{NN}(\beta,\gamma)$ with $c = 3$ and $N = 4$, a diffraction-limited 10-inch telescope is made to act like a diffraction-limited telescope of aperture $10 \div 0.488 = 20.5$ inches, except for light-gathering power. For earthbound telescopes, it would not pay to make $\delta_N < 1$ once aperture size exceeds roughly 10 inches. This is because, for large apertures, image degradation due to atmospheric turbulence is dominant over that due to pupil diffraction. An orbiting astronomical telescope, however, is above the turbulence and hence is limited in image quality only by its own capabilities. Hence, the enhancement of δ_N to 0.5 and less, by use of pupils P_N, is conceivable for a large-aperture orbiting telescope.

TOLERANCE ON PUPIL NOISE, BY S/N CRITERION

Atmospheric turbulence causes random fluctuations of phase and amplitude in the pupil of any earth-based telescope. Consequently, no telescope can be diffraction-limited during actual use. Recalling that coating $P_{NN}(\beta,\gamma)$ was to be applied to a diffraction-limited instrument, we must expect turbulence to degrade the coated telescope's amplitude response from the ideal result $a_{NN}(x,y)$.

In addition, any optical instrument *intrinsically* suffers from random phase errors across the pupil. This is a consequence of tolerances during fabrication, of misalignment, of flexing, etc. During use, vibrations and heat waves further contribute to the phase randomness.

It is therefore of practical interest to compute the effect upon the point amplitude of random amplitude and phase errors in the pupil $P_{NN}(\beta,\gamma)$. Recalling the sensitivity of the amplitude response to removal of a minute fluctuation in transmission $P_N(\beta)$ near the pupil margin, we might well expect a high sensitivity to such errors. This will be borne out.

The analysis is most conveniently done in one dimension. We first treat the case of random errors in pupil phase, alone. Let $\phi(\beta)$ represent the pupil phase, a random number, at each point β. Using the Fraunhofer approximation,[7] the amplitude response is

$$a(x) = \int_{-\beta_o}^{\beta_o} d\beta \; e^{j\phi(\beta)} \; P_N(\beta) \; e^{j\beta x} . \tag{27}$$

Assume $\phi(\beta)$ is small enough ($\leq \pi/6$) that we may expand $\exp(j\phi) = 1 + j\phi$ in Eq. (27). Define error

$$\Delta a(x) = a(x) - a_N(x), \tag{28}$$

the change in amplitude response due to the presence of a $\phi(\beta)$ distribution. Then,

$$\Delta a(x) \simeq j \int_{-\beta_o}^{\beta_o} d\beta \; \phi(\beta) \; P_N(\beta) \; e^{j\beta x}. \tag{29}$$

The mean-square average value of $\Delta a(x)$ across the field is defined as

$$e^2 = \left\langle \int_{-x_o}^{x_o} dx \; |\Delta a(x)|^2 \right\rangle . \tag{30}$$

This is a simple and (as shown below) convenient measure of the randomness introduced in $a_N(x)$ due to the random phase.

Taking expectation $\langle \; \rangle$ within the integral, substituting Eq. (29) for $\Delta a(x)$, and performing the x-integration, we find

$$e^2 = 2x_o \iint_{-\beta_o}^{\beta_o} d\beta \; d\beta' \left\langle \phi(\beta)\phi(\beta') \right\rangle P_N(\beta) \; P_N(\beta') \; \text{sinc} \; x_o(\beta-\beta'). \tag{31}$$

To proceed further, we must specify a functional form for the expectation in Eq. (31). Assume uncorrelated, white noise characteristics for ϕ. Then

$$\left\langle \phi(\beta)\phi(\beta') \right\rangle = \sigma^2 \; \delta(\beta-\beta') \tag{32}$$

where σ is the noise variance. Substitution into Eq. (31), use of Eqs. (17a), (18a), and evaluation of the remaining integral by use of the orthogonality[8] of functions $\psi_n(x)$, results in

$$e^2 = \sigma^2 \; x_o^2 \; \pi^{-1} \; [U_N(\beta_o)]^{-2} \sum_{n(\text{even})=0}^{N} [\psi_n(0)/\lambda_n]^2. \tag{33a}$$

By proceeding with the same analysis, (27) to (33), we can find e_o^2, the response noise due to random phase fluctuations in an *uncoated*, diffraction limited pupil of the same extent β_o and field x_o:

$$e_o^2 = 4 c \sigma_o^2 \tag{33b}$$

We may now compare e^2 and e_o^2.

If $\sigma = \sigma_o$, evaluation of Eqs. (33) shows that $e^2 << e_o^2$; that is, the coated amplitude response fluctuates much less than the uncoated one. But this is merely a consequence of the extremely low light transmission for the coated pupil (see Fig. 3), which attenuates *both* the signal and the noise from the uncoated case. A judgment on sensitivity to pupil noise should instead be based on the response noise relative to the background signal, since the output of the image detector (regardless of type) is always scaled by the average signal level. This signal-to-noise ratio, S/N, also represents the number of distinguishable intensity levels within the central region $|x| \leq x_o$ of the amplitude response. By evaluating S/N for each of the two pupils, the sensitivity to phase fluctuation is measured by a physically pertinent (and, it will be seen, mathematically convenient) factor.

The average signal in the amplitude response for each of the pupils may be found in the same manner as above: For the coated pupil, define the mean-square signal response over the field $|x| \leq x_o$ as

$$S^2 \equiv \left\langle |a_N(x)|^2 \right\rangle = \left\langle \int_{-x_o}^{x_o} dx \; |a_N(x)|^2 \right\rangle . \tag{34}$$

Using

$$a_N(x) = \int_{-\beta_o}^{\beta_o} d\beta \; P_N(\beta) e^{j\beta x} \tag{35}$$

and Eqs. (17a) and (18a), we find

$$S^2 = x_o[U_N(\beta_o)]^{-2} \sum_{n=0}^{N} \lambda_n^{-1} \psi_n(0)^2 . \tag{36a}$$

For the uncoated pupil we find a mean-square signal

$$S_o^{\ 2} \simeq 4\pi\beta_o, \tag{36b}$$

with reasonably small error for values $c \gtrsim \pi$.

The quotient of Eqs. (33a) and (36a) yields

$$(S/N)^2 = \pi \; x_o^{-1}(\sigma\gamma_N)^{-2} \tag{37a}$$

where
$$\gamma_N^{\ 2} = \sum_{n=0}^{N} \lambda_n^{-2} \psi_n(0)^2 \Big/ \sum_{n=0}^{N} \lambda_n^{-1} \psi_n(0)^2 . \tag{37b}$$

The quotient of Eqs. (33b) and (36b) yields

$$(S/N)_o^{\ 2} = \pi \; x_o^{-1} \; \sigma_o^{-2}. \tag{38}$$

Finally, Eqs. (37a) and (38) show that

$$(S/N) = (\sigma_o/\sigma)\gamma_N^{\ -1} \; (S/N)_o. \tag{39}$$

Thus, the sensitivity to noise in the coated pupil depends critically upon numerical factor γ_N. For example, if $\sigma = \sigma_o$ (i.e. both pupils have the same phase fluctuations), a large γ_N would cause the coated response to be severely degraded, in distinguishable levels, below that of the uncoated response. A condition $\sigma = \sigma_o$ arises, e.g., when both pupils are to be exposed to the same atmospheric turbulence.

If σ and σ_o arise chiefly from manufacturing tolerance, they are adjustable to some extent. Then, to keep $(S/N) \geq (S/N)_o$, the tolerance σ on phase must be σ_o/γ_N or less. If $\sigma_o = \pi/2$, corresponding to an rms quarter wave variation, then a value $\gamma_N = 100$ would force a tolerance on σ of 1/40 wave. Evidently, all depends on quantities γ_N.

We computed values of γ_N from its definition (37b). Figure 4 shows γ_N against resolution enhancement δ_N for various values of field parameter c. We see immediately that, for each c value, γ_N rises sharply with small

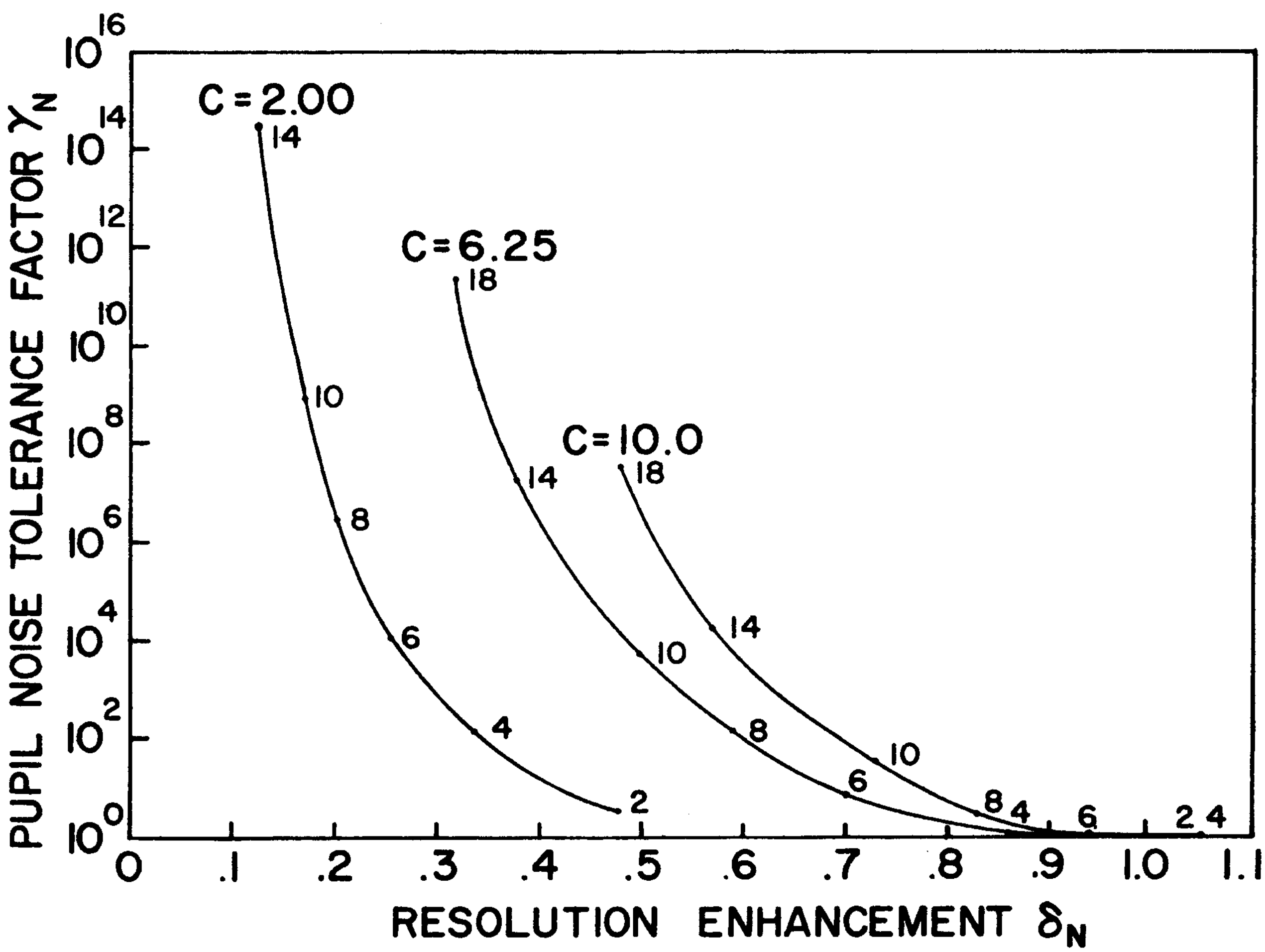

Fig. 4. Pupil noise tolerance factor γ_N against resolution enhancement δ_N, for various field widths c. Numbered points on each curve denote values of N. Each curve shows extreme sensitivity (very high γ_N) to phase at values of δ_N less than a characteristic resolution; for example, for c = 6.25 it is the value $\delta_6 = 0.70$. In effect, resolution enhancement beyond these values requires impossibly narrow tolerance on the rms noise in phase or in log amplitude.

decrements in δ_N once a critical value of δ_N is attained. Hence, for each c there is a limiting resolution determined jointly by a sudden increase in S/N sensitivity to phase fluctuations and the limit of manufacturing precision on phase variation. Thus, pupils $P_N(\beta)$ require extreme manufacturing precision, and extremely clean seeing conditions, to be used with a substantial resolution enhancement.

If the pupil suffers random amplitude variations, without phase variation, then Eq. (39) still holds, and σ and σ_o now refer to rms variations in the log amplitude transmittance. Thus, the tolerances on phase, or on log amplitude, are the same. The case of simultaneous phase and amplitude errors will not be analyzed.

We may conclude, then, that a coating $P_{NN}(\beta,\gamma)$ will improve both the resolution and the S/N of an uncoated pupil only if the coating is applied with a much narrower tolerance on rms phase and amplitude noise than existed for the uncoated pupil. If the pupil noise arises mainly from turbulence or vibration, so that it is unadjustable and equal for the two pupils, then a tradeoff between resolution enhancement and S/N must take place: improvement in one causes a degradation in the other.

DISCUSSION

Although arbitrarily perfect image formation is possible in theory, we have seen it to be limited in practice by (a) severe loss of illumination in the image, (b) a severely narrow field of view, or (c) a severely narrow tolerance on pupil rms noise. In any real design, resolution enhancement must necessarily be compromised by these effects, as indicated in the section on applications.

The situation in image processing is analogous. Although arbitrarily perfect restoration is possible, in practice the net resolution is severely limited by error in the image measurements.[5]

Since neither method of image enhancement is perfect in practice, whereas either is perfect in theory, a combination of the two methods might be advisable. Because of the two degrees of freedom, i.e. both pupil and image manipulation, it should be possible to provide a greater Strehl intensity and less sensitivity to image noise than by either method alone, for a required resolution.

The exotic functions $\psi_n(x)$ have been seen to provide a basis for both band-unlimited restoration[4] and formation of optical images. It is noted here that, in theory, it is even possible to perfectly restore an object scene from a piece of its image, no matter how small! Again, functions $\psi_n(x)$ provide the key. Because an image i(x) is band-limited (for either the coherent or the incoherent case),[8]

$$i(x) = \sum_{n=0}^{\infty} \lambda_n^{-1} \psi_n(x) \int_{-x_0/2}^{x_0/2} dx' \; i(x') \; \psi_n(x') \tag{40}$$

at *all* x. This allows the image at all x, even beyond the edge of the given image piece, to be determined from the image piece alone (which in itself is quite a feat, considering that all phase information is lacking in the given $i(x)$ if incoherent radiation is present). By taking the Fourier transform of both sides of Eq. (40), the image spectrum $I(\omega)$ is known. Finally, $I(\omega)$ is input into the band-unlimited restoration scheme,[4] resulting in perfect restoration of the object. Noise of detection would, at present, rule out the practical use of this scheme, except where the image piece extends over nearly all significantly nonzero image values.

In apposition to the phenomenon of arbitrarily perfect imagery is the finite cutoff[7] in the optical transfer function. The latter seems to rule out the former. Now, the cutoff in the latter follows from its definition as a Fourier transform over the entire point amplitude distribution. However, it is the point amplitude only within a finite interval $|x| \le x_0$ which is optimized in this paper. Therefore, the infinite transform, and its cutoff, are of little relevance to the problem at hand. In fact, a finite transform, which *would* be relevant, cannot cut off. This reasoning seems consistent with our discovery that a tradeoff can be made, to any desired extreme, between good amplitude response within a finite interval, and poor response outside that interval.

We have computed the S/N, before and after coating $P_N(\beta)$ is employed, due to uncorrelated, white noise in the pupil. If the noise is also Gaussian random, the amount of information in the image space is[13]

$$\log\,(1 + S/N) \text{ per sampling point.} \tag{41}$$

From the preceding study of S/N, we see that use of a coating $P_N(\beta)$ will simultaneously improve the information content and the resolution if the noise variance σ in the coated pupil obeys $\sigma \le \sigma_0/\gamma_N$. But, if atmospheric turbulence is the main source of pupil noise, the information must be reduced even though the resolution is improved. This is not a paradox, since information and resolution are independent criteria. In deciding whether or not to use a pupil $P_N(\beta)$, the user may have to decide which of the two criteria he wants emphasized.

We conclude that diffraction is not the major limitation on image quality. As detectors are made more sensitive, and manufacturing tolerances on pupil wave and transmittance errors are narrowed, perfect imagery can be arbitrarily approached.

CURRENT RESEARCH EFFORTS

Inherent in the design procedure of this paper is the unconstrained nature of Strehl S. The user has no control over its outcome for any design. It might seem possible to remedy this situation by defining a new optimization problem for which the user can specify S. However, the solution pupil $U(\beta)$ for this problem would have to obey an added constraint $|U(\beta)| \leq 1$ for all $|\beta| \leq \beta_0$, and there seems to be no mathematical trick for constraining U in this manner during optimization. Moreover, if S is specified by the user, the solution $U(\beta)$ cannot finally be renormalized to 1, for this automatically changes S from the required value.

To sidestep the renormalization problem, it is possible to use the procedure of Rhodes[14] and specify the fraction γ of the total radiant energy that is to lie within $|x| \leq x_0$. Here

$$\gamma = \int_0^{x_0} dx\ |a(x)|^2 \Bigg/ \int_0^{\infty} dx\ |a(x)|^2 . \tag{42}$$

Since a(x) is linear in $U(\beta)$, by Eq. (42) if $U(\beta)$ is renormalized then γ remains fixed. The design problem is now as follows: For a specified ratio γ, find the pupil $U(\beta)$ whose response $a'(x)$ is the best approximation to a truncated delta function (12a), in the mean-square sense and over limited interval $|x| \leq x_0$. After obtaining the solution $U(\beta)$, divide it by the proper constant so that $|U(\beta)| \leq 1$ for all $|\beta| \leq \beta_0$.

Since total illumination is governed jointly by the pupil transmittance and γ, a fixed γ does not alone fix the total image illumination. However, it is expected that the pupil transmittance will not fall below about 0.01 (this was about the lower limit for the design problem of this paper), and thus the absolute image illumination will always exceed 0.01 γ.

This constrained design problem has been solved, essentially as by Rhodes,[14] and the resulting point amplitude and Strehl are being evaluated on the CDC 6400 computer. The algebraic solution shows that the point amplitude cannot arbitrarily approach a delta function, as did the unconstrained solutions (17b). Hence it will be important to discover the ultimate resemblance to delta-function behavior that can be obtained for various fixed values of γ.

ACKNOWLEDGEMENTS

I wish to thank H. A. Brown of the Naval Research Laboratory for providing me with a program deck for generating prolate spheroidal coefficients. All of the numerical work in this paper derives from these numbers.

Acknowledgement is also due my colleague Dr. William Swindell for interesting discussions on the physical interpretation and application of this work.

REFERENCES CITED

1. J. L. Harris, "Diffraction and resolving power," *J. Opt. Soc. Am.* 54(7):931-936, 1964.

2. C. W. Barnes, "Object restoration in a diffraction-limited imaging system," *J. Opt. Soc. Am.* 56(5):575-578, 1966.

3. G. J. Buck and J. J. Gustincic, "Resolution limitations of a finite aperture," *IEEE Trans. on Antennas and Propagation* AP-15(3):376-381, 1967.

4. B. R. Frieden, "Band-unlimited reconstruction of optical objects and spectra," *J. Opt. Soc. Am.* 57(8):1013-1019, 1967.

5. C. K. Rushforth and R. W. Harris, "Restoration, resolution, and noise," *J. Opt. Soc. Am.* 58(4):539-545, 1968.

6. G. Toraldo di Francia, "Super-gain antennas and optical resolving power," *Nuovo cimento* 9(Suppl.):426-438, 1952.

7. Max Born and Emil Wolf, *Principles of Optics; Electromagnatic Theory of Propagation, Interference and Diffraction of Light,* New York, Pergamon Press, 803 pp., 1959.

8. D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty--I," *Bell System Tech. J.* 40:43-63, 1961.

9. Carson Flammer, *Spheroidal Wave Functions,* Stanford, California, Stanford University Press, 220 pp., 1957.

10. D. Yu. Gal'pern, "Apodization," *Optics and Spectroscopy (USSR)* 9:291, 1960.

11. J. E. Wilkins, Jr., "The resolving power of a coated objective II," *J. Opt. Soc. Am.* 40(4):222-224, 1950.

12. See, for example, National Academy of Sciences, *Synthetic-Aperture Optics, vol. 1, Woods Hole [Mass.] Summer Study, 1967.* Copies may be obtained from Defense Documentation Center, Cameron Station, Alexandria, Va.

13. C. E. Shannon and Warren Weaver, *The Mathematical Theory of Communication,* Urbana, Ill., University of Illinois Press, 1964.

14. D. R. Rhodes, "The optimum line source for the best mean-square approximation to a given radiation pattern," *IEEE Trans. on Antennas and Propagation* AP-11:440-446, 1963.

mws