b/f is called the lens "numerical aperture."

For the Airy function, ≈84% of the light energy lies within the Airy disc. But, as b increases further, spherical aberration kicks in, the intensity on the optic axis starts to diminish and a greater percentage of light energy appears beyond r_A. For the values b =.16, .19, .23 and .29μm, used in Fig. (20), Eq. (F11) gives r_A =1.56, 1.3, 1.1 and .86μm respectively. The first two curves appear to reach 0 at these values of r_A, whereas the last two curves deviate somewhat.

One wants b to be as large as possible, to decrease r_A and thus increase resolution, and to let as much light as possible exit the lens. However, as b grows, spherical aberration grows, as seen in Fig. (20): $I(r)$ decreases for $r < r_A$ and more light appears for $r > r_A$, so resolution decreases. A rule of thumb, called the Strehl criterion, suggests increasing b until the maximum intensity, the intensity on the optic axis $I(0)$, is reduced to 80% of the maximum intensity on the optic axis without any spherical aberration. Then, the image is considered still *diffraction limited*, i.e., the image is still essentially the Airy disc. From (F9), we see $I(0) \approx 1/2[1 - (\bar{B})^2/5]$. Thus, the Strehl criterion implies $\bar{B} = 1$. It does seem from Fig. (20) that this is an optimal choice.

For $\bar{B} = 1$, the wavefront (the surface of constant phase), for a ray exiting the lens a distance $\equiv \rho b$ above the optic axis, goes beyond the tangent plane by the distance $R\rho^4\bar{b}^4/21.6 = \rho^4 6\bar{B}/k \approx \lambda\rho^4$, according to Eq.(F4). Thus, the wavefront at the edge of the exit pupil, $\rho = 1$, is about a wavelength in front of the tangent plane. For $\bar{B} > 1$, images are available[85] showing appreciable spherical aberration, for path differences from $1.4\lambda\rho^4$ to $17.5\lambda\rho^4$.

4. Optical Path Calculation

The unfinished business remains of showing that the optical path length, of a ray emerging from the source at the lens focal length, passing through the lens and up to its exit surface, is given by Eq. (F2). For the following discussion, refer to Fig. (21). The focal length of the lens, according to Eq. (4), is $f = nR/2(n-1) = 1.5R$ for $n = 1.5$. Thus, the point source at a is at a distance $R/2$ to the left of the lens surface. We shall follow a ray which leaves the source at angle α to the optic axis.

A simplifying feature, which occurs only for $n = 1.5$, is that the angle of refraction cde also happens to be α. That can be seen as follows. The angle of incidence θ and the angle of refraction θ' are related by Snell's law, $\sin\theta = n\sin\theta'$. By the law of sines applied to the triangle adc, $\sin\alpha/R = \sin(\pi-\theta)/f$, or $\sin\theta = (f/R)\sin\alpha$. This is the same as Snell's law provided $f = nR$, which is only true for n=1.5.

The axial ray, $\alpha = 0$, obviously has optical path length $(R/2) + n2R = 3.5R$. For arbitrary α, the optical path length Φ_0 is ad+$n$de or, as can readily be seen from

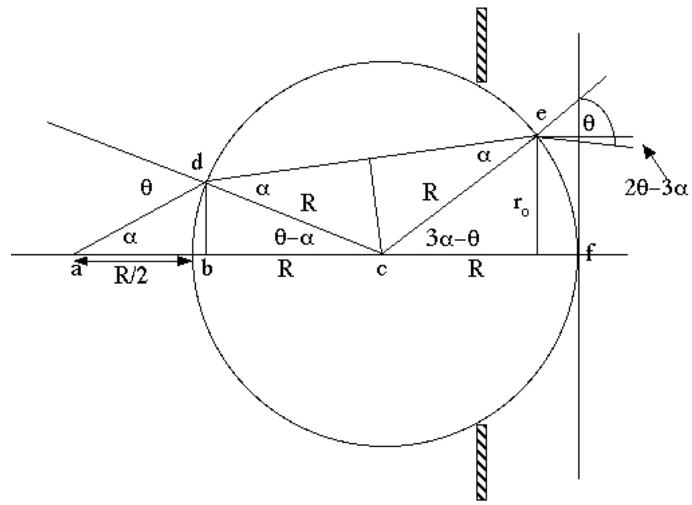

FIG. 21: Optical path length geometry

Fig.(21),

$$\begin{aligned}\Phi_0 &= \frac{R\sin(\theta-\alpha)}{\sin(\alpha)} + (1.5)2R\cos\alpha \\ &= 4.5R\cos\alpha - R\cos\theta \\ &= 4.5R\cos\alpha - R\sqrt{1-(1.5\sin\alpha)^2} \\ &\approx R[3.5 - (9/8)\alpha^2 + (57/128)\alpha^4 + ...] \quad \text{(F12)}\end{aligned}$$

The approximation (F12) is good to about 1% at $\alpha = .6 \approx 34°$. We want to express Φ_0 in terms of the distance r_0 between the optic axis and the exit point e. In terms of α, r_0 is

$$\begin{aligned}r_0 &= R\sin(3\alpha-\theta) \\ &= R\sin 3\alpha\sqrt{1-(1.5\sin\alpha)^2} - R1.5\cos 3\alpha\sin\alpha \\ &\approx R[1.5\alpha - (7/8)\alpha^3 + ...].\end{aligned}$$

This equation can be inverted,

$$\alpha = R^{-1}[(2/3)r_0 + (14/81)r_0^3 + ...]$$

and inserted into Eq. (F12), with the result (F2).

It was mentioned earlier that the angle γ the exiting ray makes with the horizontal is quite small. Here is the argument. From Fig. (21), $\gamma = 2\theta - 3\alpha$. From Snell's law, $\theta - \theta^3/6 \approx (3/2)[\alpha - \alpha^3/6]$, or $\theta \approx (3/2)\alpha + (5/16)\alpha^3$, so $\gamma \approx (5/8)\alpha^3$. Thus, the horizontal distance σ from e to the tangent plane differs from the actual distance $\sigma[1 + (1/2)\gamma^2]$ by a negligibly small amount.

APPENDIX G: EXTENDED OBJECT

Having treated the image of a point source, we shall now consider the image of a uniformly illuminated hole of radius a. The hole models a transparent object such as a spherosome or a polystyrene sphere. We shall suppose that the lens is diffraction limited, i.e., the exit pupil has been chosen so that the image of a point source is the Airy intensity distribution.

But first, we remark that the intensity in this problem is related to that of the complementary case, the image of an opaque disc of radius a. According to *Babinet's principle*, the sum of the light amplitudes for these two cases is the constant light amplitude without either. (This is easy to see from the linearity of the Huyghens-Fresnel-Fraunhofer expression discussed in Appendix B 3). So, where one is light, the other is dark.

Suppose the hole is illuminated with incoherent light, as in ordinary microscopy. If $a < \lambda/4$, the illumination is nonetheless effectively coherent, since any incident plane wave of random phase will have little phase difference across the hole. If $a > \lambda/2$, the illumination may be considered incoherent. This is the case considered here.

1. Incoherent Illumination

If this were geometrical optics, light from each uniformly illuminated point of the object plane would pass through the lens and be focused as an illuminated point on the image plane. The properly scaled image of all these illuminated points would be a uniformly illuminated circle of radius a. We shall call the circumference of this circle the "image circle edge." The new wrinkle is that diffraction surrounds each imaged point with its own Airy disc (assuming that spherical aberration is negligible), so that the image extends beyond the image circle edge. The intensities add so, at a point $\mathbf{r}$ on the image plane, the net intensity is

$$I(r) \sim \frac{1}{\pi}\int_{A_0} dA_0 \left[\frac{J_1(k\tilde{b}|\mathbf{r}-\mathbf{r_0}|)}{|\mathbf{r}-\mathbf{r_0}|}\right]^2, \qquad \text{(G1)}$$

where A_0 is the area of the image circle, and $\tilde{b} \equiv b/f$ is called the numerical aperture.

For $a >> r_A$, where r_A is the Airy radius, the intensity at the center point of the image circle is, by (G1),

$$I(0) \sim \frac{1}{\pi}\int_0^a r_0 dr_0 \int_0^{2\pi} d\phi \left[\frac{J_1(k\tilde{b}r_0)}{r_0}\right]^2 \approx 1.$$

In this equation, the limit a has been extended to ∞ with no appreciable error, since the major contribution is from Airy discs centered within distance r_A of the origin.

As the point of interest moves off center, the intensity remains essentially constant, until at a distance $\approx a - r_A$ from the center, a distance r_A from the image circle edge. Then I starts to decrease, reaching the value $\approx .5$ at the edge. This is because, at the edge, ≈half the Airy discs contribute intensity, compared to the discs which contribute intensity at a point well inside the image circle.

Now, we turn to quantitative analysis of the general case, with no restriction of the relative sizes of a and r_A. We shall calculate the intensity (G1) outside the image circle, at $r = 0$ which is placed a distance z beyond the image circle edge, i.e., the center of the image circle in this coordinate system is at $r = a + z$. The contributing Airy disc centers lie within the image circle, between radius r_0 ($z \leq r_0 \leq 2a + z$) and radius $r_0 + dr_0$, along an arc subtending an angle 2ϕ. The hole circumference $(x-a-z)^2 + y^2 = a^2$ cuts this arc at two points. Setting $x = r_0 \cos\phi$ and $y = r_0 \sin\phi$ in this expression allows one to find $\cos\phi$. Eq. (G1) becomes

$$I_{\text{out}}(z) \sim \frac{2}{\pi}\int_z^{2a+z} dr_0 \cos^{-1}\left[\frac{r_0^2 + z^2 + 2az}{2r_0(a+z)}\right]\frac{J_1^2(k\tilde{b}r_0)}{r_0} \qquad \text{(G2)}$$

For completeness, we put here the comparable expression for the intensity inside the image circle. Again, we calculate the intensity (G1) at $r = 0$, where this new coordinate system origin is a distance z away from the center of the image circle. There are two contributions, one from a circular area of radius $a - z$, the other from the rest of the disc ($a - z \leq r_0 \leq a + z$):

$$\begin{aligned} I_{\text{in}}(z) \; &\sim \; 2\int_0^{a-z} dr_0 \frac{J_1^2(k\tilde{b}r_0)}{r_0} \\ &+ \frac{2}{\pi}\int_{a-z}^{a+z} dr_0 \cos^{-1}\left[\frac{r_0^2 + z^2 - a^2}{2r_0 z}\right]\frac{J_1^2(k\tilde{b}r_0)}{r_0}. \qquad \text{(G3)} \end{aligned}$$

For large a, (G2) becomes

$$I_{\text{out}}(z) \approx \frac{2}{\pi}\int_z^{\infty} dr_0 \cos^{-1}\left[\frac{z}{r_0}\right]\frac{J_1^2(k\tilde{b}r_0)}{r_0}$$

This is a function of $k\tilde{b}z = 3.83(z/r_A)$. Numerical evaluation shows $I_{\text{out}}(z)$ drops from .5 at $z = 0$ to $\approx .05$ at $z = r_A$. While it is somewhat subjective, this suggests that we take the perceived edge of the image of the hole to be located where the intensity is 5% of its maximum value at the center of the image circle. Thus, diffraction increases the radius of a large hole from a to $R \approx a + r_A$.

By changing the variable of integration in (G2) to r_0/a, one sees that the intensity is a function of two variables, z/a and $k\tilde{b}a/3.83 = a/r_A$. For each value of a/r_A, one can numerically solve Eq. (G2) for the value of z/a for which $I(z) = .05I(0)$. This is the ratio R/a, where R is defined as the radius of the image. A graph of R/r_A vs a/r_A is given in Fig. 12, and is discussed in section III H.

---

[1] P. W. van der Pas, Scientiarum Historia **13**, 127 (1971): see [68] for a more detailed discussion.

[2] A transcript of an interesting historical talk about