

Chapter 16

Diffraction and Optical Resolution

First Edition

Hamzeh Khanpour

hamzeh.khanpour@agh.edu.pl

Faculty of Physics and Applied Computer Science

AGH University of Krakow

Physics 2

Physics for Technical and Engineering Students -
Volume II

Electricity, Magnetism, Waves, Optics, Relativity,
Quantum Physics, and Nuclear Physics

1 September 2026

Contents

Preface	iii
Conventions and notation	iv
16 Diffraction and Optical Resolution	1
16.1 Motivation and roadmap	1
16.2 From wave superposition to diffraction	2
16.2.1 Huygens–Fresnel viewpoint	2
16.3 Fraunhofer and Fresnel diffraction	3
16.4 Single-slit diffraction: geometry and minima	4
16.5 Single-slit intensity distribution	6
16.5.1 Secondary maxima	7
16.6 Central maximum width and screen geometry	8
16.7 Diffraction by obstacles and Babinet’s principle	9
16.8 Circular apertures and the Airy pattern	10
16.8.1 Camera design: diffraction, f-number and pixel sampling	11
16.9 The Rayleigh criterion and optical resolution	12
16.9.1 Linear separation at a known distance	13
16.10 Microscopes, numerical aperture and diffraction-limited detail	14
16.11 Two finite slits: interference inside a diffraction envelope	16
16.12 From two slits to many: the diffraction grating	17
16.12.1 Allowed diffraction orders	18
16.13 Angular dispersion of a diffraction grating	19
16.13.1 Finite arrays: line shape and width	20
16.14 Resolving power of a grating	21
16.14.1 Deriving resolving power from peak width	22
16.15 Order overlap and free spectral range	23
16.16 X-ray diffraction and Bragg’s law	24
16.16.1 Reading an X-ray diffraction scan	25
16.17 Laboratory investigation of single-slit diffraction	26
16.17.1 Experimental procedure	26
16.17.2 Uncertainty propagation	27
16.17.3 Estimating a slit width from several diffraction minima	28
16.17.4 Uncertainty propagation in slit metrology	29
16.17.5 Computational fitting of the full intensity profile	29
16.18 Computational exploration of diffraction patterns	30
16.18.1 A reproducible finite-array calculation	31
16.19 Main chapter summary	31
16.20 Advanced Topics for Technical and Engineering Students	33
16.20.1 From the scalar wave equation to the diffraction integral	33
16.20.2 Complex amplitude and the Fourier-transform viewpoint	34
16.20.3 Two-dimensional rectangular apertures	35
16.20.4 Circular pupils and the Bessel-function Airy pattern	35
16.20.5 Point-spread functions and convolution	36
16.20.6 Abbe’s spatial-frequency picture of microscopy	36
16.20.7 Coherent and incoherent imaging	37

16.20.8	Optical transfer function and modulation transfer function	38
16.20.9	Detector sampling, Nyquist reasoning and aliasing	39
16.20.10	Advanced synthesis example: imaging trade-offs for a space telescope .	40
16.20.11	Apodization, windowing and side-lobe control	40
16.20.12	Gaussian beams and diffraction-limited divergence	41
16.20.13	Wavefront aberrations and the Strehl ratio	42
16.20.14	Fresnel diffraction, Fresnel zones and zone plates	43
16.20.15	Fourier-transform rules as optical design rules	44
16.20.16	The 4f optical system and spatial filtering	44
16.20.17	Numerical Fraunhofer diffraction with the FFT	45
16.20.18	FFT sampling: connecting numerical pixels to physical angles	46
16.20.19	Advanced mini-project: numerical aperture-to-image pipeline	47
16.20.20	Spectrometer design: dispersion, slit width and detector sampling . . .	47
16.20.21	Advanced synthesis example: designing a compact grating spectrometer	48
16.20.22	Reciprocal lattices and the deeper meaning of Bragg diffraction	49
16.20.23	Crystal structure factors and systematic absences	50
16.20.24	Powder diffraction and line broadening	51
16.20.25	Electron diffraction and matter waves	51
16.20.26	Diffraction and information: what a finite pupil removes	52
16.20.27	Advanced section summary	52
16.21	Original Practice Problems	53
16.22	Selected Solutions to Representative Problems	56

Preface

A wave transmitted through a finite aperture does not continue in a single sharply defined direction. Fields from different parts of the aperture arrive at an observation point with different phases, and their superposition produces an angular intensity pattern. This is diffraction. The phrase “bending around edges” can be a useful visual cue, but phase and superposition provide the physical explanation.

For technical and engineering students, diffraction is both a wave phenomenon and a design constraint. Aperture diameter limits angular resolution in telescopes and cameras; numerical aperture sets microscopic detail scales; grating geometry determines spectral dispersion and resolving power. The same interference principles let X rays probe crystal structure. Understanding these applications requires attention to illumination, geometry, approximations and what the detector actually measures.

The core develops single-slit minima and intensity, circular apertures and Rayleigh resolution, finite-width double slits, gratings, and Bragg diffraction. Worked examples connect each model to a calculation or design decision. Laboratory analysis and computational exploration then test the assumptions against measured or simulated patterns. The optional enrichment section connects these results through Fourier propagation, pupil functions, image contrast, sampling, Gaussian beams and reciprocal-space scattering.

The treatment assumes calculus-based introductory physics, wave superposition and the geometrical optics of Chapter 14. Chapter 15 supplies the language of phase and coherence. Additional mathematics is developed as needed in the enrichment section. The practice problems range from conceptual reasoning to computation and instrument design; selected solutions demonstrate setup, units and physical interpretation.

The Guideline

When approaching a diffraction problem, identify four things before using a formula: the wavelength λ , the aperture geometry and characteristic size, the observation regime, and the quantity actually requested. A condition for a *minimum*, a condition for a *maximum*, and a *resolution criterion* are not interchangeable even when they contain similar symbols.

Pedagogical boundary

The main chapter treats scalar diffraction of monochromatic or nearly monochromatic light under standard introductory assumptions, with special emphasis on Fraunhofer patterns. Vector diffraction, full electromagnetic boundary-value solutions, rigorous coherence theory, aberration theory, and modern super-resolution microscopy are beyond the main scope. Selected aspects appear only in the advanced enrichment section, where their assumptions are stated explicitly.

Conventions and notation

Use the wavelength *in the propagation medium*, $\lambda = \lambda_0/n$, in aperture phase differences; λ_0 is the vacuum wavelength and n the refractive index. Numerical examples use air or vacuum unless stated otherwise. Microscope formulas written with $\text{NA} = n \sin \alpha$ use λ_0 , since NA already includes the object-space refractive index.

The observation angle θ is measured from the forward aperture normal. For a flat screen normal to that axis, $y = L \tan \theta$; only at small angles does $y \approx L\theta \approx L \sin \theta$. A spectrometer camera can have a different axis, which is specified when needed.

Symbol	Meaning	Typical unit
λ, λ_0	medium and vacuum wavelengths	m
k	wave number, $2\pi/\lambda$	rad m ⁻¹
a	full slit width; controls the diffraction envelope	m
d	center-to-center slit or grating spacing	m
D	circular-aperture diameter	m
L, y	aperture-to-screen distance and transverse coordinate	m
θ	observation angle from the aperture normal	rad
β	slit phase variable, $\pi a \sin \theta / \lambda$	dimensionless
γ	array phase variable, $\pi d \sin \theta / \lambda$	dimensionless
$I(\theta), I_0$	intensity and locally defined reference intensity	W m ⁻²
N	number of illuminated grating periods	dimensionless
N_f	f-number, f/D , for focal length f	dimensionless
R	spectral resolving power, $\lambda/\Delta\lambda$	dimensionless
NA	$n \sin \alpha$; α is the object-side collection half-angle	dimensionless
θ_R, δ_R	angular and object-space Rayleigh separations	rad, m
d_{hkl}	spacing of a specified crystal-plane family	m
θ_B	Bragg angle measured from the crystal planes	rad or degree

The integer m labels grating orders $0, \pm 1, \dots$; for single-slit minima it excludes zero. Bragg order is denoted by n locally. In symmetric reflection geometry the reported detector deflection is $2\theta_B$. The symbols d and d_{hkl} distinguish grating spacing from crystal-plane spacing when both contexts are nearby; R denotes source distance only where explicitly defined.

Caution

Use radians in small-angle approximations and angular derivatives. Convert degree measurements before using $\sin \theta \approx \theta$, propagating angular uncertainty, or inserting a peak width into the Scherrer relation. An intensity ratio is dimensionless; a screen distance, angular width and spectral width are different physical quantities.

Chapter 16

Diffraction and Optical Resolution

16.1 Motivation and roadmap

The ray model of geometrical optics is extraordinarily successful when optical components are large compared with the wavelength and when one is interested in image positions rather than fine spatial structure. Yet no lens, mirror, telescope, microscope or camera has infinite aperture. The moment a wave is truncated by a finite opening, the transmitted field acquires angular spread. That spread is not a manufacturing defect; it is a consequence of wave superposition.

A useful way to organize the chapter is to move from one aperture to many apertures and then from pattern formation to information content:

1. a single slit shows how a finite aperture creates a broad intensity pattern;
2. a circular aperture gives the Airy pattern that limits ordinary imaging systems;
3. two finite-width slits show interference fringes modulated by a diffraction envelope;
4. many equally spaced slits create narrow principal maxima and therefore form a diffraction grating;
5. the narrowness and angular separation of grating lines determine spectral resolving power;
6. periodic atomic planes act as a three-dimensional diffraction structure for short wavelengths, leading to Bragg diffraction.

Learning goals

After studying the core chapter, the student should be able to:

- identify the aperture, wavelength, coherence assumptions and propagation regime;
- derive single-slit minima and the sinc-squared intensity distribution;
- distinguish angular width from screen width and test the small-angle approximation;
- apply circular-pupil and microscope resolution criteria with the correct wavelength convention;
- explain finite-slit envelopes, missing orders and grating peak narrowing;
- derive grating dispersion and resolving power and assess detector sampling;
- infer crystal spacing from Bragg angles and distinguish peak position from peak intensity;
- fit diffraction measurements, propagate uncertainty and check a numerical model.

The optional enrichment develops Fourier optics, transfer functions and reciprocal space.

Key message

A smaller aperture does not produce a sharper beam. For a fixed wavelength, decreasing the transverse size of an aperture increases the angular spread of the transmitted wave. Spatial confinement in the aperture plane and angular confinement in the far field are competing requirements.

16.2 From wave superposition to diffraction

The physical starting point is the superposition principle. If several coherent wave contributions reach the same point, the electric fields add before intensity is calculated. For two contributions this idea produced the interference fringes studied in the previous chapter. For a finite aperture, the number of contributing points is effectively continuous.

A useful scalar model is obtained by dividing a slit into many narrow strips. Each strip transmits a small part of the incident wave. At an observation point far from the aperture, contributions from neighboring strips travel nearly parallel paths but differ slightly in path length. The accumulated phase difference across the full aperture determines whether the contributions reinforce or cancel.

The word *diffraction* is therefore best understood as interference associated with a spatially extended source or aperture. The distinction between “interference” and “diffraction” is partly conventional. In many introductory contexts, interference refers to a small number of well-separated coherent sources, while diffraction refers to interference among contributions from different parts of one or more finite apertures. Mathematically, both are applications of wave superposition.

Definition 16.1: Diffraction

Diffraction is the redistribution of wave amplitude in direction or position caused by coherent superposition of contributions transmitted by, reflected from, or scattered by a structure of finite transverse extent. In optics, diffraction becomes especially important when aperture dimensions are not overwhelmingly larger than the wavelength or when high angular resolution is required.

16.2.1 Huygens–Fresnel viewpoint

Huygens’ construction says that each point on a wavefront can be regarded as the source of a secondary wavelet. Fresnel’s refinement adds phase and superposition. In a modern scalar treatment, the field at an observation point is built by integrating complex amplitudes over the open aperture. The elementary single-slit formula developed later is a direct one-dimensional example of this integral viewpoint.

The central conceptual point is that an aperture does not simply remove part of a wave; it also defines the spatial region over which coherent contributions remain. The shape and size of that region encode the far-field pattern.

Guided checks

1. Why is it misleading to describe diffraction only as light “turning a corner”?
2. If two apertures have the same shape but one is scaled to twice every linear dimension, which one should produce the narrower far-field pattern at the same wavelength?
3. Does diffraction require that the aperture dimension be smaller than the wavelength? Explain.
4. Why must amplitudes, rather than intensities, be added before interference is evaluated?

16.3 Fraunhofer and Fresnel diffraction

Diffraction patterns depend on the distance between the aperture and the observation plane. Two limiting regimes are especially useful.

In *Fresnel diffraction*, the source or screen is close enough that wavefront curvature across the aperture matters. Different parts of the aperture not only acquire different phases but also contribute with geometry that varies appreciably across the opening. Fresnel patterns therefore depend explicitly on propagation distance and can change shape as the screen is moved.

In *Fraunhofer diffraction*, the incident and outgoing waves are treated as approximately planar across the aperture. Experimentally, this can be achieved either by placing the source and screen very far away or, more conveniently, by using lenses to form the far-field angular pattern in a focal plane. The elementary formulas in this chapter are predominantly Fraunhofer formulas.

A common dimensionless indicator is the Fresnel number

$$N_F \sim \frac{a^2}{\lambda L}, \quad (16.1)$$

where this chapter uses the *full slit width* a and aperture-to-screen distance L . With this convention we take $N_F = a^2/(\lambda L)$; a convention based on half-width gives a value four times smaller. For a centered slit under plane-wave illumination, the neglected quadratic propagation phase changes by approximately $\pi a^2/(4\lambda L)$ between center and edge. Thus $N_F \ll 1$ is a useful sufficient far-field estimate, whereas values of order unity require a Fresnel treatment at the desired accuracy. A finite-distance source adds incident wavefront curvature, which must also be negligible or corrected by a collimator.

The far-field approximation and the small-angle approximation answer different questions. The first controls phase curvature over the aperture; the second replaces trigonometric functions of the observation angle. A narrow slit may have a far-field pattern at angles large enough that $y = L\theta$ is inaccurate. Conversely, small observation angles do not by themselves guarantee Fraunhofer propagation.

Caution

The condition $N_F \ll 1$ is a regime estimate, not a universal sharp boundary. In a laboratory, the quality of the Fraunhofer approximation depends on the accuracy required, the illumination geometry and the observable being measured. A lens can produce a far-field pattern at a finite physical distance, so “Fraunhofer” does not literally mean that the detector must be kilometers away.

Example 16.1: Checking whether a direct screen is approximately in the far field

A laser of wavelength 633 nm illuminates a slit of width 0.20 mm. The screen is 2.0 m away. Estimate the Fresnel number using the slit width as the characteristic size and comment on the regime.

Solution. With $a = 2.0 \times 10^{-4}$ m,

$$N_F \sim \frac{a^2}{\lambda L} = \frac{(2.0 \times 10^{-4})^2}{(633 \times 10^{-9})(2.0)} \approx 3.2 \times 10^{-2}. \quad (16.2)$$

This value is much smaller than unity, so a Fraunhofer model is a good first approximation. If high-precision comparison with measured side-lobe positions were required, one would still check the full experimental geometry.

16.4 Single-slit diffraction: geometry and minima

Consider a monochromatic plane wave incident normally on a long slit of width a . The slit is long in the direction perpendicular to the page, so the problem is effectively one-dimensional. Let the observation direction make angle θ with the optical axis.

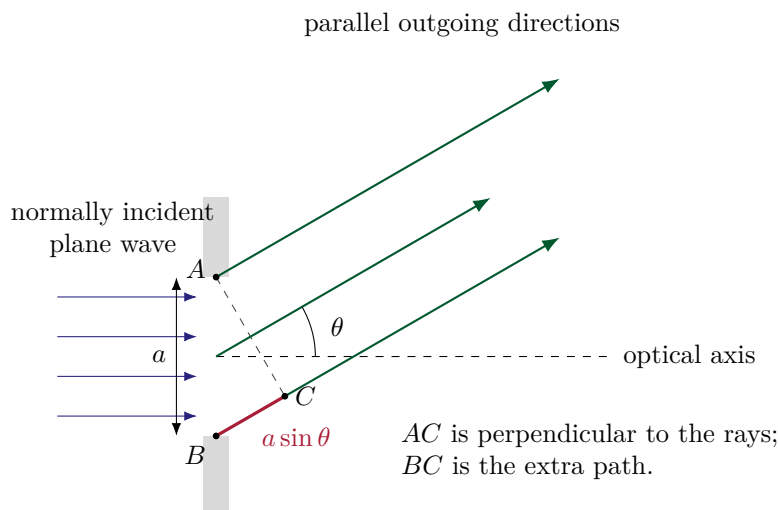


Figure 16.1: Far-field path geometry for a slit of full width a . Parallel outgoing rays define a common wavefront AC ; projecting AB onto their direction gives $BC = a \sin \theta$. For the first minimum, corresponding points separated by $a/2$ differ by half this path.

Figure 16.1 makes the path projection explicit.

The path difference between rays from the top and bottom edges is

$$\Delta = a \sin \theta. \quad (16.3)$$

To find the first minimum, divide the slit into two equal halves. If the total edge-to-edge path difference is one wavelength, then corresponding points separated by $a/2$ have path difference $\lambda/2$. Their contributions arrive π out of phase and cancel pairwise. Therefore the first minimum occurs when

$$a \sin \theta = \lambda. \quad (16.4)$$

For the m th positive minimum divide the slit into $2m$ equal strips. Pair each odd strip with the adjacent even strip. Corresponding points in a pair are separated by $a/(2m)$, so when $a \sin \theta = m\lambda$ their path difference is $\lambda/2$. Uniform illumination makes their amplitudes equal, and every pair cancels. Negative angles follow by symmetry. The top and bottom edges themselves are in phase at these angles; their agreement does not prevent cancellation of the *whole aperture integral*. The minima satisfy

$$a \sin \theta_m = m\lambda, \quad m = \pm 1, \pm 2, \pm 3, \dots \quad (16.5)$$

There is no $m = 0$ minimum. The forward direction $\theta = 0$ is the central maximum.

Equation (16.5) immediately displays the inverse relation between aperture width and angular spread. For the first minimum,

$$\sin \theta_1 = \frac{\lambda}{a}. \quad (16.6)$$

A smaller a requires a larger θ_1 .

Method

For a single-slit minimum problem:

1. identify the physical slit width a ;
2. identify which dark fringe is being used, hence the nonzero integer m ;
3. use $a \sin \theta_m = m\lambda$ without approximating unless needed;
4. if a screen coordinate is given, combine the diffraction condition with $y = L \tan \theta$;
5. use $y \approx Lm\lambda/a$ only after checking that $|\theta| \ll 1$.

Example 16.2: Finding slit width from the first minimum

A 532 nm laser produces a first diffraction minimum at $\theta_1 = 0.305^\circ$ on either side of the central axis. Determine the slit width.

Solution. For the first minimum, $m = 1$, so

$$a = \frac{\lambda}{\sin \theta_1}. \quad (16.7)$$

With $\lambda = 532 \times 10^{-9} \text{ m}$ and $\theta_1 = 0.305^\circ$,

$$a \approx \frac{532 \times 10^{-9}}{\sin(0.305^\circ)} \approx 1.00 \times 10^{-4} \text{ m} = 0.100 \text{ mm}. \quad (16.8)$$

The result is physically plausible: a slit roughly two hundred wavelengths wide gives a diffraction angle of a few milliradians.

Example 16.3: Exact screen position versus small-angle approximation

A slit of width $20 \mu\text{m}$ is illuminated by 600 nm light. A screen is 0.80 m away. Find the position of the first minimum using the exact geometry and compare with the small-angle result.

Solution. The first minimum satisfies

$$\sin \theta_1 = \frac{600 \times 10^{-9}}{20 \times 10^{-6}} = 0.0300, \quad (16.9)$$

so $\theta_1 \approx 1.719^\circ$. The exact screen coordinate is

$$y_1 = L \tan \theta_1 = (0.80) \tan(1.719^\circ) \approx 2.401 \times 10^{-2} \text{ m}. \quad (16.10)$$

The small-angle result is

$$y_1 \approx L \frac{\lambda}{a} = (0.80)(0.0300) = 2.400 \times 10^{-2} \text{ m}. \quad (16.11)$$

The difference is negligible here because the angle is only about 0.03 rad.

Guided checks

1. If the slit width is doubled while wavelength is fixed, what happens to the angular width of the central maximum?
2. Why is $m = 0$ excluded from the single-slit minimum condition?
3. A red and a blue laser illuminate the same slit. Which one produces the wider pattern?

4. Does increasing the screen distance change the angular positions of the minima? Does it change their linear positions on the screen?

16.5 Single-slit intensity distribution

The positions of dark fringes are important, but a complete description requires the intensity between the minima. The central maximum is much brighter than the secondary maxima, and the side lobes decrease rapidly. This behavior emerges naturally when the contributions from all parts of the slit are added as complex amplitudes.

To add sinusoidal fields efficiently, represent the physical electric field by $\text{Re}[\mathcal{E}e^{-i\omega t}]$. The complex number $\mathcal{E}$ stores amplitude and phase; Euler's identity $e^{i\phi} = \cos\phi + i\sin\phi$ is simply a compact way of keeping both components of a phasor. A detector averages over many optical cycles, giving intensity proportional to $|\mathcal{E}|^2$, not to the amplitude itself.

Let x measure position across the slit from $-a/2$ to $+a/2$. For a distant observation angle θ , a point at x contributes an additional phase approximately

$$\phi(x) = -kx \sin \theta. \quad (16.12)$$

The minus sign follows because points with positive x are closer to a positive-angle observer. A consistent opposite phasor convention gives the same intensity. Assume uniform illumination and one common polarization so that the amplitude contribution per unit width is constant. The complex field amplitude is proportional to

$$\mathcal{E}(\theta) \propto \int_{-a/2}^{a/2} e^{-ikx \sin \theta} dx. \quad (16.13)$$

Integrating,

$$\mathcal{E}(\theta) \propto \left[\frac{e^{-ikx \sin \theta}}{-ik \sin \theta} \right]_{-a/2}^{a/2} \quad (16.14)$$

$$= \frac{2 \sin (ka \sin \theta/2)}{k \sin \theta}. \quad (16.15)$$

Using $k = 2\pi/\lambda$ and the definition of β ,

$$\beta = \frac{\pi a \sin \theta}{\lambda}, \quad (16.16)$$

we can normalize the field to its on-axis value and write

$$\frac{\mathcal{E}(\theta)}{\mathcal{E}(0)} = \frac{\sin \beta}{\beta}. \quad (16.17)$$

Intensity is proportional to the squared magnitude of the field amplitude, so

$$I(\theta) = I_0 \left(\frac{\sin \beta}{\beta} \right)^2, \quad \beta = \frac{\pi a \sin \theta}{\lambda}. \quad (16.18)$$

The limiting value at $\beta = 0$ is finite because $\sin \beta/\beta \rightarrow 1$ as $\beta \rightarrow 0$.

The zeros occur when $\sin \beta = 0$ with $\beta \neq 0$, namely $\beta = m\pi$, which reproduces $a \sin \theta = m\lambda$. Figure 16.2 plots the actual function, including its weak side lobes.

Peak intensity and transmitted power. Normalizing every pattern to I_0 hides the change in brightness when the slit is adjusted. At fixed incident field, slit height and observation geometry, the on-axis amplitude is proportional to a , so $I_0 \propto a^2$. The angular width is proportional to $1/a$, while the integrated transmitted power per unit slit height is proportional to a . Indeed, in paraxial screen coordinates,

$$\int_{-\infty}^{\infty} I(y) dy = I_0 \frac{\lambda L}{a} \int_{-\infty}^{\infty} \left(\frac{\sin \pi u}{\pi u} \right)^2 du = I_0 \frac{\lambda L}{a},$$

where $u = ay/(\lambda L)$ and the dimensionless integral is one. This is consistent with transmitted power proportional to width. Diffraction redistributes energy; it does not create extra energy in the central peak. The normalized sinc-squared law neglects slowly varying obliquity and propagation-distance factors, so retaining $\sin \theta$ does not make it a complete electromagnetic model at grazing angles.

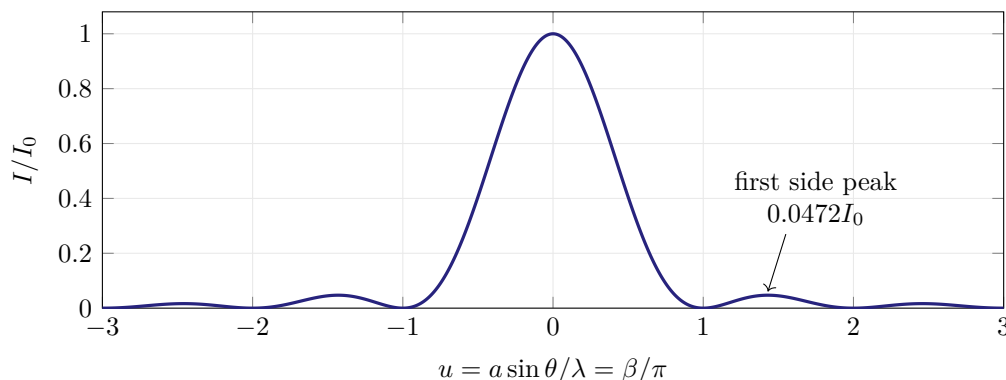


Figure 16.2: Calculated single-slit intensity with $I_0 = I(0)$. Nonzero integer u gives a zero, and the central lobe occupies $-1 < u < 1$. The ordinate is linear and every curve point is evaluated from the sinc-squared law.

16.5.1 Secondary maxima

The secondary maxima are not exactly halfway between neighboring minima. Differentiating the intensity, or equivalently the amplitude factor away from zeros, gives the condition

$$\tan \beta = \beta, \quad (16.19)$$

excluding $\beta = 0$. The first positive solution is approximately $\beta \approx 4.493$, between π and 2π . At this point the intensity is only about 4.7% of the central intensity. This large contrast is one reason the central lobe dominates visually.

Caution

The formula $a \sin \theta = m\lambda$ gives *minima*, not secondary maxima. The side maxima require the transcendental condition $\tan \beta = \beta$ if their exact angular positions are needed. Approximating them as halfway between minima is often acceptable only for quick sketches, not for precision calculations.

Example 16.4: Relative intensity at a specified angle

A slit of width $80\text{ }\mu\text{m}$ is illuminated by light of wavelength 500 nm . Find the intensity relative to the central intensity at $\theta = 0.50^\circ$.

Solution. First compute

$$\beta = \frac{\pi a \sin \theta}{\lambda} = \frac{\pi(80 \times 10^{-6}) \sin(0.50^\circ)}{500 \times 10^{-9}} \approx 4.386. \quad (16.20)$$

Then

$$\frac{I}{I_0} = \left(\frac{\sin 4.386}{4.386} \right)^2 \approx 4.66 \times 10^{-2}. \quad (16.21)$$

Thus the intensity is about 4.7% of the central value. The angle is close to the first secondary maximum, consistent with the known first side-lobe scale.

Key message

The single-slit pattern is a squared sinc function. The aperture width sets the angular scale, while the detailed side-lobe structure follows from coherent addition across the entire aperture. A narrower aperture makes the main lobe broader; it does not simply reduce the total transmitted light.

16.6 Central maximum width and screen geometry

The central maximum extends between the first minima on the two sides of the optical axis. Its exact angular width is

$$\Delta\theta_{\text{central}} = 2 \arcsin \left(\frac{\lambda}{a} \right), \quad (16.22)$$

within the scalar aperture model and for $\lambda \leq a$. At $\lambda = a$ the predicted first zeros are at grazing directions, not at finite positions on a planar screen. For the finite screen width below, $\lambda < a$ is required. Wavelength-scale apertures also require care because the assumed uniform transmitted field and scalar boundary treatment can fail. In the common small-angle regime,

$$\Delta\theta_{\text{central}} \approx \frac{2\lambda}{a}. \quad (16.23)$$

On a screen a distance L away, the exact central width is

$$w_{\text{central}} = 2L \tan \left[\arcsin \left(\frac{\lambda}{a} \right) \right], \quad (16.24)$$

and the small-angle result is

$$w_{\text{central}} \approx \frac{2L\lambda}{a}. \quad (16.25)$$

These formulas provide a practical method for measuring an unknown slit width or wavelength. Because w_{central} depends linearly on L and λ but inversely on a , experimental design can increase sensitivity by increasing the propagation distance or choosing a narrower slit. However, making the slit very narrow also reduces transmitted optical power and may make alignment more difficult.

Example 16.5: Designing a measurable central maximum

A laboratory has a 650 nm diode laser and a screen that can be placed 1.50 m from the slit. Choose a slit width that gives a central maximum approximately 20 mm wide.

Solution. Use the small-angle design relation

$$a \approx \frac{2L\lambda}{w_{\text{central}}}. \quad (16.26)$$

Thus

$$a \approx \frac{2(1.50)(650 \times 10^{-9})}{20 \times 10^{-3}} \approx 9.75 \times 10^{-5} \text{ m} = 97.5 \mu\text{m}. \quad (16.27)$$

A commercially available slit near 0.10 mm would therefore be suitable. The first-minimum angle would be about $\lambda/a \approx 6.7 \times 10^{-3}$ rad, so the small-angle approximation is fully adequate.

Guided checks

1. What happens to the central maximum width if the screen distance is doubled?
2. Why does the central maximum have twice the angular separation between the axis and the first minimum?
3. If the wavelength is increased by 10%, by approximately what percentage does the small-angle central width change?
4. In an experiment, why might one avoid making the slit arbitrarily narrow even though that makes the pattern easier to measure spatially?

16.7 Diffraction by obstacles and Babinet's principle

Diffraction is not restricted to openings. An opaque obstacle also removes part of a wavefront and therefore modifies the field downstream. A striking example is diffraction by a thin wire or a human hair. The far-field pattern has dark and bright bands whose angular scale is related to the obstacle width.

Babinet's principle provides a useful qualitative connection. For complementary screens—one transparent where the other is opaque—the diffracted fields away from the exact forward beam are closely related. Thus a narrow opaque object can produce a diffraction pattern with the same fringe spacing as a slit of comparable width. This is why a laser and a hair can form a simple classroom method for estimating hair diameter.

The exact statement of Babinet's principle belongs to scalar diffraction theory and requires care with the unobstructed incident field. For introductory use, the practical message is enough: complementary aperture and obstacle geometries have the same far-field angular structure apart from the direct beam and overall field combination.

Example 16.6: Estimating the diameter of a hair from diffraction

A red laser with wavelength 632.8 nm illuminates a human hair. On a screen 2.00 m away, the distance between the first minima on opposite sides of the central maximum is 31.5 mm. Estimate the hair diameter using the complementary-slit approximation.

Solution. The half-width to the first minimum is

$$y_1 = 15.75 \text{ mm}. \quad (16.28)$$

The angle is small, so $\theta_1 \approx y_1/L = 7.875 \times 10^{-3}$ rad. The effective width is then

$$a \approx \frac{\lambda}{\theta_1} = \frac{632.8 \times 10^{-9}}{7.875 \times 10^{-3}} \approx 8.04 \times 10^{-5} \text{ m} = 80 \mu\text{m}. \quad (16.29)$$

This is a plausible human-hair diameter. The dominant experimental uncertainties are likely to come from measuring the fringe positions, keeping the hair straight and knowing the screen distance accurately.

Several experimental details deserve attention:

- the object should be nearly perpendicular to the beam so its projected width is known;
- the illuminated segment should be reasonably uniform in diameter;
- the laser spot must cover the object and provide adequate spatial coherence;
- the screen distance must be large enough for the chosen Fraunhofer treatment;
- the central reference position should be determined symmetrically from both sides of the pattern.

16.8 Circular apertures and the Airy pattern

Most optical systems do not have long rectangular slits. Camera lenses, telescope pupils and microscope objectives are approximately circular. The diffraction pattern of a uniformly illuminated circular aperture is rotationally symmetric: a bright central disk is surrounded by alternating dark and bright rings. The central disk is called the Airy disk, and the full pattern is the Airy pattern.

Figure 16.3 distinguishes the disk radius from the complete ring pattern.

A complete derivation requires a two-dimensional aperture integral and leads to a Bessel function. The main result needed here is that the first dark ring occurs at

$$\sin \theta_1 \approx 1.22 \frac{\lambda}{D}. \quad (16.30)$$

For small angles,

$$\theta_1 \approx 1.22 \frac{\lambda}{D}. \quad (16.31)$$

The factor 1.22 is specific to a circular aperture with uniform amplitude across the pupil. It comes from the first zero of the relevant Bessel-function combination.

If a lens of focal length f forms the far-field pattern in its focal plane, the radius of the first dark ring is

$$r_1 \approx f\theta_1 \approx 1.22 \frac{\lambda f}{D}. \quad (16.32)$$

Defining the f-number $N_f = f/D$,

$$r_1 \approx 1.22 \lambda N_f, \quad d_{\text{Airy}} \approx 2.44 \lambda N_f. \quad (16.33)$$

Here d_{Airy} is the diameter of the central disk out to the first dark ring.

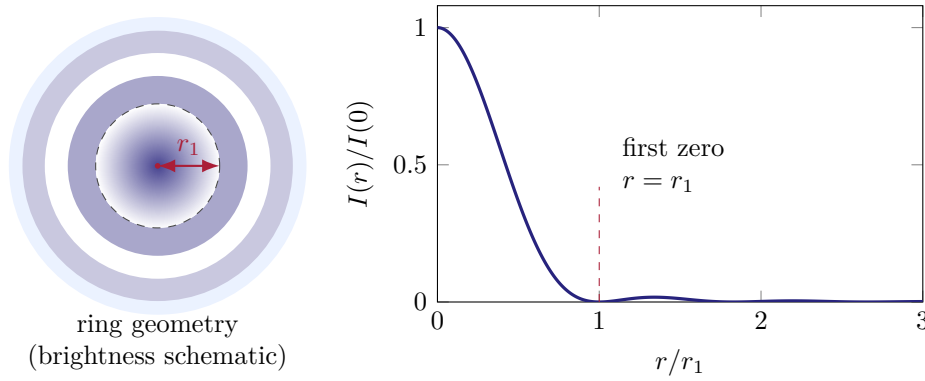


Figure 16.3: Airy pattern of a uniformly illuminated clear circular pupil. Left: ring geometry with a true center-to-first-zero radius; ring brightness is schematic. Right: the calculated radial cut $[2J_1(u)/u]^2$, with $u = 3.8317 r/r_1$ in paraxial focal coordinates.

Example 16.7: Airy-disk diameter on a camera sensor

A camera lens is used at $f/8$ with green light of wavelength 550 nm. Estimate the diffraction-limited Airy-disk diameter at the sensor.

Solution. The f-number is $N_f = 8$. Using Eq. (16.33),

$$d_{\text{Airy}} \approx 2.44\lambda N_f = 2.44(550 \times 10^{-9})(8) \approx 1.07 \times 10^{-5} \text{ m} = 10.7 \mu\text{m}. \quad (16.34)$$

A sensor with pixel pitch much smaller than this value samples the diffraction pattern finely, but the optical image itself cannot contain arbitrarily sharp detail merely because the pixels are small.

Caution

The Airy-disk size is a diffraction contribution to image blur, not the only contribution. Real lenses also have aberrations, defocus, motion blur, atmospheric effects, finite source size and detector sampling. Stopping a lens down can reduce aberrations while increasing diffraction, so practical image quality often has an intermediate optimum aperture.

16.8.1 Camera design: diffraction, f-number and pixel sampling

For a circular lens aperture, the diameter of the Airy disk to the first minimum in the focal plane is

$$d_{\text{Airy}} \approx 2.44\lambda N_f, \quad N_f = \frac{f}{D}. \quad (16.35)$$

This form is especially useful for cameras because photographers and optical engineers commonly specify aperture by f-number rather than by physical pupil diameter.

Equation (16.35) explains an important design trade-off. A smaller aperture corresponds to a larger f-number and therefore a larger diffraction spot. At the same time, stopping down can reduce several geometrical aberrations, increase depth of field, and make focus less sensitive. Real lenses therefore often achieve their best image sharpness at an intermediate f-number rather than at either extreme.

The detector introduces another length scale: the pixel pitch p . Three regimes are worth distinguishing qualitatively:

1. $d_{\text{Airy}} \ll p$: the detector undersamples the optical detail; pixels, rather than diffraction, can dominate recorded sharpness;

2. $d_{\text{Airy}} \sim$ a few p : optical and detector sampling scales are sensibly matched;
3. $d_{\text{Airy}} \gg p$: the diffraction blur extends over many pixels, so smaller pixels alone cannot recover the lost optical detail.

These are qualitative size comparisons. In particular, two pixels across the *Airy diameter* do not satisfy a general Nyquist criterion. For an ideal incoherent circular pupil in air at image-space f-number N_f , the intensity cutoff is $f_c \approx 1/(\lambda N_f)$. Recording the entire optical bandwidth without aliasing therefore requires

$$p \leq \frac{\lambda N_f}{2}, \quad \frac{d_{\text{Airy}}}{p} \gtrsim 4.88. \quad (16.36)$$

The advanced section derives this from the transfer function. Pixel integration, optical aberrations and the low contrast near cutoff affect practical performance, but do not justify confusing first-zero diameter with the shortest transmitted period.

Example 16.8: At what f-number does diffraction span two pixels?

A monochrome sensor has pixel pitch $p = 3.0 \mu\text{m}$. At $\lambda = 550 \text{ nm}$, estimate the f-number for which the Airy diameter equals two pixel pitches.

Solution. Set

$$2.44\lambda N_f = 2p. \quad (16.37)$$

Then

$$N_f = \frac{2p}{2.44\lambda} = \frac{6.0 \mu\text{m}}{2.44(0.550 \mu\text{m})} \approx 4.47. \quad (16.38)$$

Thus an f-number around f/4.5 gives an Airy diameter of about two pixels at 550 nm. Full-band Nyquist sampling for this ideal model would instead require $N_f \gtrsim 2p/\lambda \approx 10.9$, illustrating why the two tests differ. This does not mean that f/4.5 is universally “optimal”: color filters, aberrations, demosaicing, focus, object contrast and the desired depth of field all influence real image quality.

Guided checks

1. Why does increasing focal length at fixed f-number leave the Airy diameter in the focal plane proportional to the same f-number expression?
2. If a camera changes from f/4 to f/16, by what factor does the diffraction-limited Airy diameter change at fixed wavelength?
3. Why is “more megapixels” not equivalent to “better optical resolution” once pixels are already fine enough to sample the delivered image?

16.9 The Rayleigh criterion and optical resolution

Diffraction becomes especially important when an imaging system must distinguish two nearby point-like objects. In this section the two sources have equal brightness and are mutually incoherent, so their time-averaged intensities add. Coherent sources require addition of complex fields and can produce a different image depending on relative phase. In geometrical optics, a perfect lens could map each point object to a point image. With a finite circular aperture, each point is instead mapped to an Airy pattern. Two nearby objects therefore produce overlapping intensity distributions.

Lord Rayleigh proposed a practical criterion for deciding when two identical point sources are just resolved. Under the criterion, the central maximum of one Airy pattern lies at the first minimum of

the other. For a circular aperture of diameter D , the corresponding angular separation is

$$\theta_R \approx 1.22 \frac{\lambda}{D} \quad (16.39)$$

for small angles.

Figure 16.4 compares calculated intensity profiles on the two sides of this benchmark.

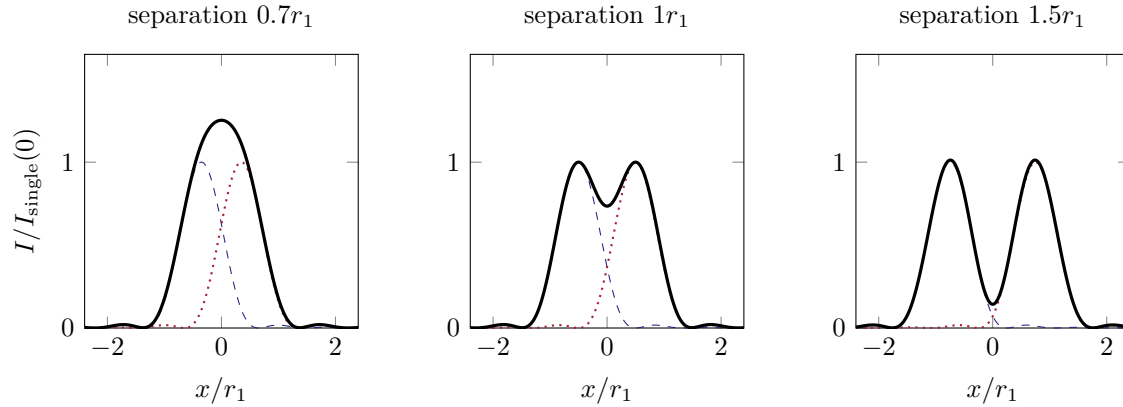


Figure 16.4: Calculated image cuts for two equal mutually incoherent point sources. Dashed blue and dotted red curves are the individual Airy intensities; black is their unscaled sum. At separation r_1 (center panel), each source center lies at the other source's first zero, and the midpoint intensity is approximately 0.735 of a single-source peak. The transition in shape is continuous.

The Rayleigh criterion should not be interpreted as a law stating that information suddenly disappears below θ_R . It is a conventional benchmark. Modern fitting methods can infer separations smaller than a nominal Rayleigh value when the signal model is known and the signal-to-noise ratio is high, while poor signal-to-noise can make objects separated by more than θ_R difficult to distinguish. Nevertheless, Eq. (16.39) remains a very useful engineering scale for diffraction-limited imaging.

Key message

Resolution is not the same as magnification. A system can magnify a blurred image without revealing new spatial information. Increasing aperture diameter reduces diffraction blur and can increase resolvable detail; increasing magnification alone does not.

16.9.1 Linear separation at a known distance

If two distant objects are separated transversely by distance s and lie approximately the same distance R from the aperture, then for small angles

$$\theta \approx \frac{s}{R}. \quad (16.40)$$

The Rayleigh condition becomes

$$s_R \approx 1.22 \frac{\lambda R}{D}. \quad (16.41)$$

This form is useful for astronomy, remote sensing and long-range imaging.

Example 16.9: Diffraction-limited telescope resolution

A telescope with a circular objective of diameter $D = 0.25$ m observes at wavelength $\lambda = 550$ nm. Estimate its Rayleigh angular resolution in radians and arcseconds.

Solution. The diffraction-limited angle is

$$\theta_R \approx 1.22 \frac{550 \times 10^{-9}}{0.25} \approx 2.68 \times 10^{-6} \text{ rad.} \quad (16.42)$$

Since $1 \text{ rad} \approx 206265 \text{ arcseconds}$,

$$\theta_R \approx (2.68 \times 10^{-6})(206265) \approx 0.55 \text{ arcsec.} \quad (16.43)$$

This is an ideal diffraction limit. Ground-based observations may be limited more strongly by atmospheric seeing unless adaptive optics or exceptionally stable conditions are available.

Example 16.10: Aperture required to separate two satellite features

A visible-light imaging system operates at $\lambda = 600$ nm and must just resolve two ideal point-like features separated by 0.50 m at a distance of 400 km. Estimate the required diffraction-limited circular aperture diameter.

Solution. The angular separation is

$$\theta \approx \frac{s}{R} = \frac{0.50}{4.00 \times 10^5} = 1.25 \times 10^{-6} \text{ rad.} \quad (16.44)$$

Set this equal to the Rayleigh angle:

$$D \approx 1.22 \frac{\lambda}{\theta} = 1.22 \frac{600 \times 10^{-9}}{1.25 \times 10^{-6}} \approx 0.586 \text{ m.} \quad (16.45)$$

An aperture of roughly 0.6 m is required under the idealized assumptions. Real remote-sensing performance also depends on motion, atmospheric path, contrast, detector sampling and image reconstruction.

Guided checks

1. Why does increasing telescope diameter improve angular resolution?
2. If a telescope observes at infrared wavelength twice as long as a visible wavelength, what happens to the ideal diffraction angle for the same aperture?
3. Can a larger detector pixel count compensate completely for an aperture that is too small? Explain.
4. Why is Rayleigh's criterion better described as a conventional benchmark than an absolute information boundary?

16.10 Microscopes, numerical aperture and diffraction-limited detail

Figure 16.5 defines the object-side collection cone.

A microscope works in a different geometric regime from a telescope, but the central issue is the same: a finite objective accepts only a limited range of propagation directions. The relevant quantity

is the numerical aperture

$$\text{NA} = n \sin \alpha, \quad (16.46)$$

where n is the refractive index on the object side and α is the half-angle of the cone collected by the objective.

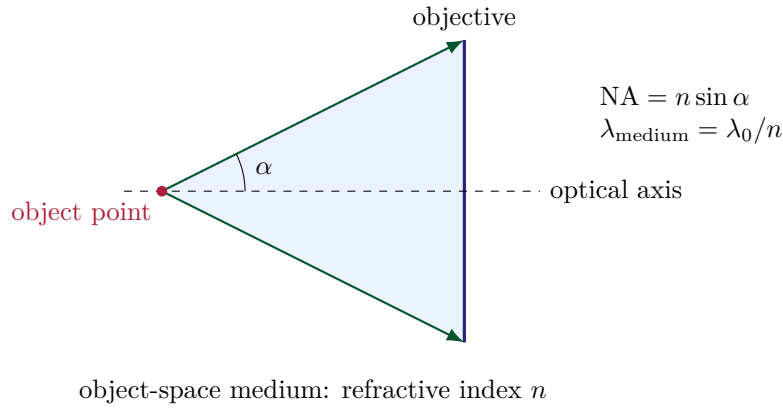


Figure 16.5: Object-side collection cone of a microscope objective. The aperture angle is measured at the specimen. The index n is included once in NA, so $0.61\lambda_0/\text{NA}$ uses vacuum wavelength. Geometry is schematic.

For two equal, mutually incoherent point emitters imaged by an ideal circular pupil, a commonly used scalar Rayleigh lateral-resolution scale is

$$\delta_R \approx 0.61 \frac{\lambda_0}{\text{NA}}. \quad (16.47)$$

Here λ_0 is the vacuum wavelength, while n is already included in NA. Another widely used criterion associated with Abbe's treatment of periodic structure is

$$\delta_{\text{Abbe}} \approx \frac{\lambda_0}{2\text{NA}}. \quad (16.48)$$

The Abbe expression assumes an appropriate illumination bandwidth, such as a transmitted-light condenser with numerical aperture comparable to that of the objective. With an on-axis coherent plane wave alone, the smallest transmitted object period instead scales as λ_0/NA . The advanced discussion derives this distinction; the factor of two must not be introduced independently of illumination. The Rayleigh expression concerns two point emitters, whereas a periodic-object cutoff is a contrast-transfer statement. Both improve with shorter vacuum wavelength and larger numerical aperture. At very high NA, vector polarization effects modify the detailed scalar point-spread function.

Immersion objectives exploit the factor n . By filling the space between specimen and objective with a medium of refractive index greater than unity, the system can collect larger transverse wave-vector components than the same angular cone in air.

Example 16.11: Resolution scale of a microscope objective

A microscope uses green light of vacuum wavelength 550 nm and an oil-immersion objective with numerical aperture 1.30. Estimate the Rayleigh lateral resolution.

Solution.

$$\delta_R \approx 0.61 \frac{550 \text{ nm}}{1.30} \approx 258 \text{ nm}. \quad (16.49)$$

The value should be interpreted as an ideal diffraction-limited scale for conventional imaging, not as a guarantee that all structures separated by 258 nm will be equally easy to distinguish.

Caution

Do not mix the telescope formula $1.22\lambda/D$ directly with the microscope formula $0.61\lambda_0/\text{NA}$ without identifying the geometry. They express closely related diffraction physics but use different system parameters and map angular spread into object-space separation differently.

16.11 Two finite slits: interference inside a diffraction envelope

The ideal Young double-slit formula often treats each slit as infinitesimally narrow. Real slits have finite width. Each slit therefore produces its own single-slit diffraction pattern, while the two slit centers produce interference fringes. The observed intensity is the product of these effects.

For two identical slits of width a and center separation d , define

$$\beta = \frac{\pi a \sin \theta}{\lambda}, \quad \gamma = \frac{\pi d \sin \theta}{\lambda}. \quad (16.50)$$

With I_0 chosen as the central peak intensity, the pattern may be written

$$I(\theta) = I_0 \left(\frac{\sin \beta}{\beta} \right)^2 \cos^2 \gamma. \quad (16.51)$$

The factor $\cos^2 \gamma$ produces closely spaced interference fringes. The factor $(\sin \beta / \beta)^2$ provides the broad diffraction envelope.

The interference maxima are approximately located by

$$d \sin \theta = m\lambda, \quad (16.52)$$

while the diffraction minima satisfy

$$a \sin \theta = p\lambda, \quad p = \pm 1, \pm 2, \dots \quad (16.53)$$

If a nominal constructive order falls exactly at a diffraction minimum, its intensity vanishes. This is called a missing order. The constructive orders of the cosine factor need not be exact maxima of the product: away from zeros, differentiation with respect to β gives

$$\cot \beta - \frac{1}{\beta} - \frac{d}{a} \tan \left(\frac{d}{a} \beta \right) = 0.$$

When d/a is large, the envelope varies slowly over most fringes and the shift is small. Near an envelope zero, weak residual local peaks can remain between nominal orders. Consequently, counting the orders inside an envelope is not the same as counting every local extremum of the exact curve in Figure 16.6.

Example 16.12: Counting constructive orders inside the central envelope

Two identical slits have width $a = 0.040$ mm and center separation $d = 0.20$ mm. How many nominal constructive orders of the two-slit interference factor lie within the central single-slit diffraction envelope, excluding the dark boundaries?

Solution. The central diffraction envelope extends between

$$|\sin \theta| < \frac{\lambda}{a}. \quad (16.54)$$

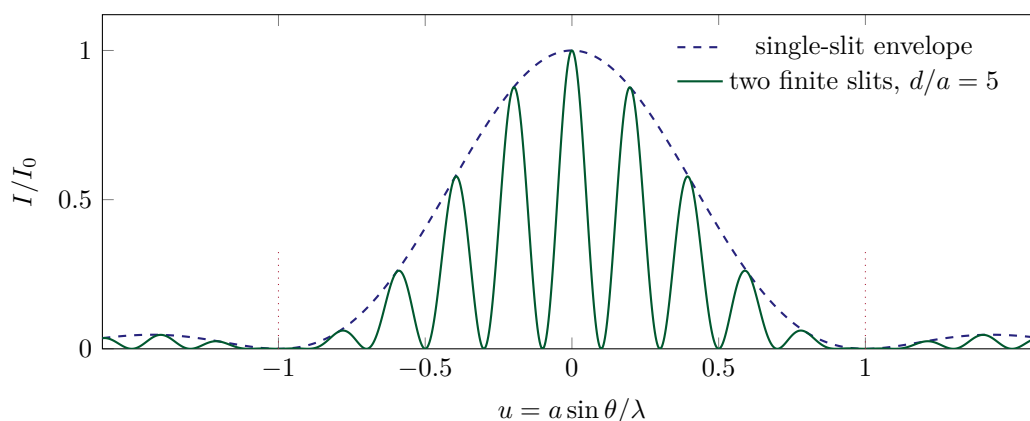


Figure 16.6: Calculated two-slit intensity for $d/a = 5$, using the full product of the sinc-squared envelope and the cosine-squared interference factor. The nominal orders $m = \pm 5$ coincide with $u = \pm 1$ and are suppressed. Tiny local peaks near envelope boundaries are distinct from nominal constructive orders.

Interference maxima satisfy

$$\sin \theta_m = \frac{m\lambda}{d}. \quad (16.55)$$

To lie inside the central envelope,

$$\left| \frac{m\lambda}{d} \right| < \frac{\lambda}{a} \Rightarrow |m| < \frac{d}{a} = 5. \quad (16.56)$$

Thus $m = -4, -3, -2, -1, 0, 1, 2, 3, 4$, giving nine nominal constructive orders inside the open interval. The $m = \pm 5$ interference maxima coincide with the first diffraction minima and are missing.

Guided checks

1. Which parameter controls the broad envelope, a or d ?
2. Which parameter controls the fine fringe spacing?
3. What happens to the number of interference fringes under the central envelope if d/a increases?
4. Why can an interference maximum predicted by $d \sin \theta = m\lambda$ fail to appear experimentally?

16.12 From two slits to many: the diffraction grating

A diffraction grating contains a large number of equally spaced transmitting slits, reflecting grooves or equivalent periodic scattering elements. The essential advantage of many coherent elements is that principal maxima become very narrow. Narrow maxima allow nearby wavelengths to be separated angularly with much greater precision than in a two-slit system.

The geometry in Figure 16.7 separates the slit width from the period.

For normal incidence and grating spacing d , contributions from adjacent slits have path difference

$$\Delta = d \sin \theta. \quad (16.57)$$

Constructive interference occurs when this path difference is an integer number of wavelengths:

$$d \sin \theta_m = m\lambda, \quad m = 0, \pm 1, \pm 2, \dots \quad (16.58)$$

The central order $m = 0$ contains all wavelengths at $\theta = 0$ for normal incidence. Nonzero orders disperse wavelengths because the angle depends on λ .

The grating spacing is often specified through a line density g in lines per unit length. If g is expressed in lines per meter,

$$d = \frac{1}{g}. \quad (16.59)$$

For example, 1200 lines per millimeter corresponds to

$$d = \frac{1 \text{ mm}}{1200} = 833 \text{ nm}. \quad (16.60)$$

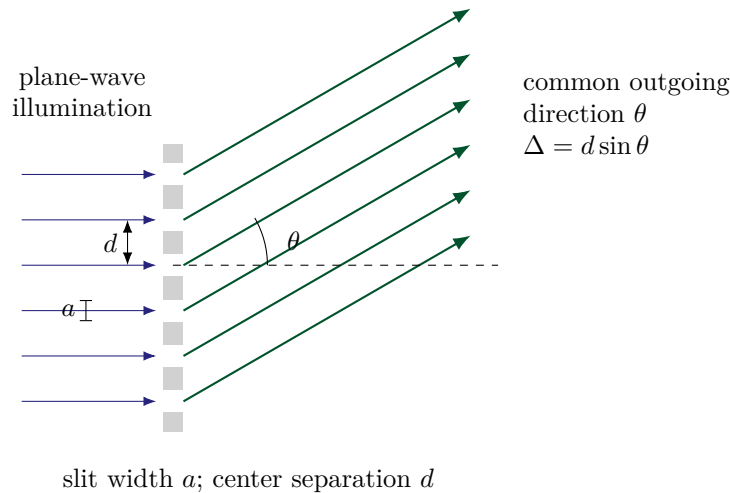


Figure 16.7: Normally illuminated transmission grating. Open slits are drawn explicitly and outgoing rays are parallel in the far-field approximation. The phase step between neighboring centers is $2\pi d \sin \theta / \lambda$; the slit width separately sets the envelope.

16.12.1 Allowed diffraction orders

Because $|\sin \theta| \leq 1$, not every integer order is physically allowed. Equation (16.58) requires

$$|m|\lambda \leq d. \quad (16.61)$$

The largest algebraically allowed order magnitude is the greatest integer satisfying this inequality. Equality corresponds to a grazing ray, which does not reach a finite screen parallel to the grating. For a collected propagating order away from grazing, use the strict inequality. A geometrically allowed order may also be suppressed by the single-element envelope or have negligible efficiency. High line-density gratings have small d , which restricts their available orders.

Example 16.13: Angles and allowed orders of a grating

A transmission grating has 600 lines per millimeter and is illuminated normally by 500 nm light. Find the first- and second-order angles and determine the largest allowed order.

Solution. The spacing is

$$d = \frac{10^{-3} \text{ m}}{600} = 1.667 \times 10^{-6} \text{ m}. \quad (16.62)$$

For $m = 1$,

$$\sin \theta_1 = \frac{500 \times 10^{-9}}{1.667 \times 10^{-6}} = 0.300, \quad (16.63)$$

so $\theta_1 \approx 17.46^\circ$. For $m = 2$,

$$\sin \theta_2 = 0.600, \quad (16.64)$$

so $\theta_2 \approx 36.87^\circ$. The order condition gives

$$|m| \leq \frac{d}{\lambda} \approx 3.33, \quad (16.65)$$

so $m = \pm 3$ is the largest nonzero allowed order. The third order occurs at $\sin \theta_3 = 0.900$, or $\theta_3 \approx 64.2^\circ$.

Key message

A grating does not separate wavelengths because different colors travel at different speeds in vacuum. It separates them because the phase-matching condition $d \sin \theta = m\lambda$ sends different wavelengths to different angles.

16.13 Angular dispersion of a diffraction grating

Spectroscopy requires more than simply producing a spectrum. One wants nearby wavelengths to appear at measurably different positions. The rate at which diffraction angle changes with wavelength is the angular dispersion.

Differentiate the grating equation at fixed order m :

$$d \cos \theta \, d\theta = m \, d\lambda. \quad (16.66)$$

Therefore

$$\mathcal{D} \equiv \frac{d\theta}{d\lambda} = \frac{m}{d \cos \theta}. \quad (16.67)$$

Higher order m , smaller spacing d , and operation at larger $|\theta|$ all increase angular dispersion, although practical design also must consider efficiency, aberrations, order overlap and detector geometry.

Figure 16.8 distinguishes the camera axis from the grating normal.

The detector mapping requires its own geometry. If a planar projection is specified by $x = f \tan \theta$, measured relative to the grating normal, then differentiation gives

$$\frac{dx}{d\lambda} = \frac{fm}{d \cos^3 \theta}. \quad (16.68)$$

This is exact differentiation of the stated projection model, not a universal law for camera lenses at large field angles.

More commonly, a spectrometer camera is pointed along the central diffracted ray at angle θ_0 . Its local coordinate is $x = f \tan(\theta - \theta_0)$, so at the central wavelength

$$\left. \frac{dx}{d\lambda} \right|_{\theta_0} = \frac{fm}{d \cos \theta_0}. \quad (16.69)$$

Only the small field angle *relative to the camera axis* is paraxial; θ_0 itself can be large. The camera-axis convention must accompany every wavelength-per-pixel calculation.

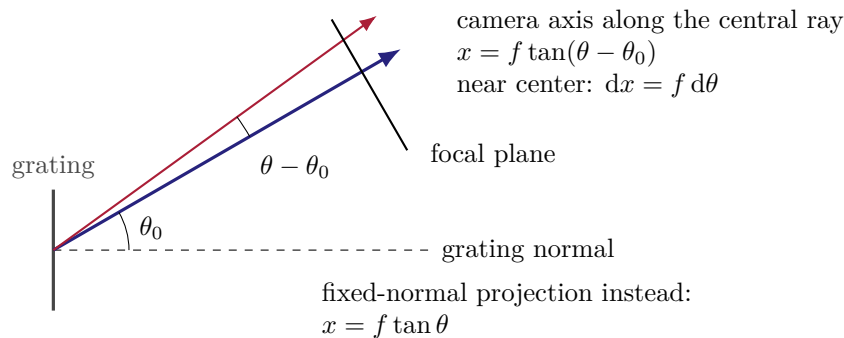


Figure 16.8: Angular coordinate conventions for a spectrometer camera, shown schematically. A camera centered on the chosen diffracted ray measures a small local field angle even when the grating angle is large. The equations define the assumed mapping; real camera distortion may require calibration.

Example 16.14: Angular separation of two nearby spectral lines

A grating with 1200 lines per millimeter is used in first order near $\lambda = 600 \text{ nm}$. Estimate the angular separation of two lines separated by 0.20 nm .

Solution. The spacing is

$$d = \frac{10^{-3}}{1200} = 8.333 \times 10^{-7} \text{ m}. \quad (16.70)$$

The first-order angle satisfies

$$\sin \theta = \frac{600 \times 10^{-9}}{8.333 \times 10^{-7}} = 0.720, \quad (16.71)$$

so $\theta \approx 46.05^\circ$ and $\cos \theta \approx 0.694$. The angular dispersion is

$$\mathcal{D} = \frac{1}{d \cos \theta} \approx 1.73 \times 10^6 \text{ rad m}^{-1}. \quad (16.72)$$

For $\Delta\lambda = 0.20 \text{ nm} = 2.0 \times 10^{-10} \text{ m}$,

$$\Delta\theta \approx \mathcal{D}\Delta\lambda \approx 3.46 \times 10^{-4} \text{ rad} \approx 0.0198^\circ. \quad (16.73)$$

This is the angular separation of the line centers. Whether they are actually resolved also depends on the width of each diffracted line, which leads to resolving power.

Caution

Large dispersion does not automatically imply high resolving power. Dispersion measures separation of line centers per wavelength interval. Resolution compares that separation with the finite width of the spectral lines produced by the instrument. A spectrum can be strongly spread out yet still have broad overlapping features.

16.13.1 Finite arrays: line shape and width

The ideal grating equation gives the principal-maximum angles but not the line shape. For N identical narrow slits with adjacent phase difference 2γ , the complex amplitude is a finite geometric series:

$$S_N = 1 + e^{i2\gamma} + e^{i4\gamma} + \dots + e^{i2(N-1)\gamma}. \quad (16.74)$$

Using the finite-series formula,

$$S_N = e^{i(N-1)\gamma} \frac{\sin(N\gamma)}{\sin \gamma}. \quad (16.75)$$

Therefore the array factor is

$$|S_N|^2 = \left[\frac{\sin(N\gamma)}{\sin \gamma} \right]^2. \quad (16.76)$$

For an N -slit grating with finite slit width, a convenient normalized model is

$$\frac{I(\theta)}{I_0} = \left(\frac{\sin \beta}{\beta} \right)^2 \left[\frac{\sin(N\gamma)}{N \sin \gamma} \right]^2, \quad (16.77)$$

with

$$\beta = \frac{\pi a \sin \theta}{\lambda}, \quad \gamma = \frac{\pi d \sin \theta}{\lambda}. \quad (16.78)$$

Here $I_0 = I(0)$ is the central intensity for the complete array. At $\gamma = m\pi$, the normalized array factor has limiting value one; the envelope can still suppress that order. At fixed incident illumination the unnormalized on-axis intensity grows as N^2 , while the integrated transmitted power grows only as N : increasing the number of slits concentrates the power into narrower peaks.

At a principal maximum, $\gamma = m\pi$. For $N > 1$, the nearest zeros of the array factor occur exactly in phase when

$$N\gamma = Nm\pi \pm \pi, \quad (16.79)$$

so the phase half-width is approximately

$$\Delta\gamma = \frac{\pi}{N}. \quad (16.80)$$

Since

$$\gamma = \frac{\pi d \sin \theta}{\lambda}, \quad (16.81)$$

small variations give

$$\Delta\gamma \approx \frac{\pi d \cos \theta}{\lambda} \Delta\theta, \quad (16.82)$$

leading to the angular half-width scale

$$\Delta\theta \sim \frac{\lambda}{Nd \cos \theta}. \quad (16.83)$$

This is the distance from a peak center to its first zero, *not* its FWHM or its full zero-to-zero width. The latter is approximately twice this value. Linearizing in angle requires a narrow peak away from grazing. The $1/N$ narrowing is the reason that many illuminated grooves improve resolution; Figure 16.9 makes the distinction between fixed peak position and changing width explicit.

16.14 Resolving power of a grating

A finite grating contains a finite number N of illuminated grooves or slits. Consequently its principal maxima have nonzero angular width. Increasing N makes the peaks narrower. This is the diffraction origin of grating resolving power.

For an ideal grating under standard assumptions, applying the Rayleigh criterion to two nearby wavelengths yields

$$R \equiv \frac{\lambda}{\Delta\lambda} = |m|N. \quad (16.84)$$

Here $\Delta\lambda$ is the minimum wavelength separation according to the ideal Rayleigh criterion, $|m|$ is the order magnitude and N is the number of coherently illuminated grating periods.

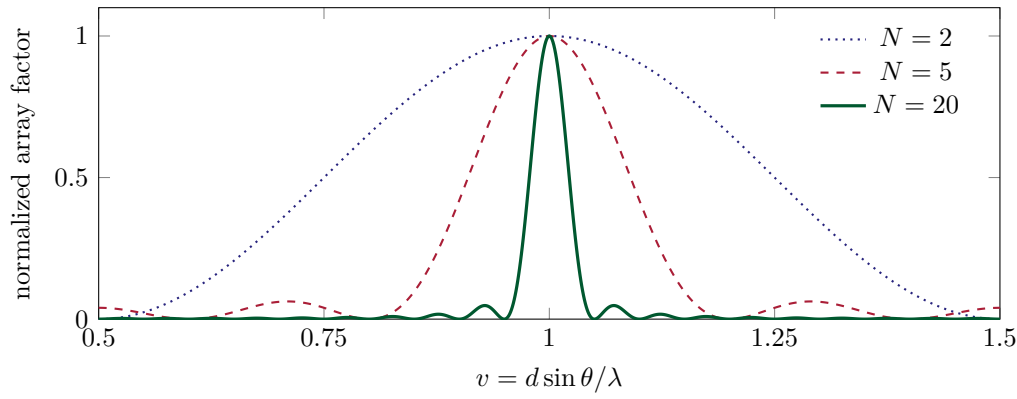


Figure 16.9: Calculated array factors near the first principal order for $N = 2, 5, 20$. Every curve peaks at $v = 1$, while its nearest zeros are at $v = 1 \pm 1/N$. The single-slit envelope has been divided out so that the effect of array length is isolated.

16.14.1 Deriving resolving power from peak width

Two wavelengths separated by $\Delta\lambda$ in the same order have angular separation

$$\Delta\theta_\lambda \approx \frac{m}{d \cos \theta} \Delta\lambda. \quad (16.85)$$

The diffraction width of one grating line is of order

$$\Delta\theta_{\text{line}} \sim \frac{\lambda}{Nd \cos \theta}. \quad (16.86)$$

Equating these scales under a Rayleigh-type criterion gives

$$\frac{m}{d \cos \theta} \Delta\lambda \sim \frac{\lambda}{Nd \cos \theta}, \quad (16.87)$$

so

$$\frac{\lambda}{\Delta\lambda} \sim |m|N. \quad (16.88)$$

For uniformly illuminated identical elements, a narrow unsuppressed order, $|\Delta\lambda| \ll \lambda$ and the center-to-first-zero Rayleigh convention, the leading result has coefficient one:

$$R = |m|N. \quad (16.89)$$

Key message

The grating equation determines where a spectral line appears. The finite illuminated grating size determines how narrow that line can be. Spectral resolution is therefore a diffraction problem in the dimension parallel to the grating dispersion direction.

Equation (16.84) has a simple design message. A high-resolution spectrograph needs a large illuminated groove count and can gain further resolving power by operating in higher order. In real instruments, slit width, optical aberrations, detector sampling, finite source size, grating imperfections and signal-to-noise may reduce the achieved resolution.

If the illuminated grating width along the dispersion direction is W , then approximately

$$N \approx \frac{W}{d}, \quad (16.90)$$

for normal incidence and full illumination of that width. Thus larger beam diameter can directly improve the ideal resolving power by illuminating more grooves.

Example 16.15: Can a grating resolve the sodium doublet?

A spectrometer is to resolve two lines near 589.3 nm separated by 0.60 nm in first order. What minimum number of illuminated grooves is required by the ideal grating resolving-power formula?

Solution. The required resolving power is

$$R = \frac{\lambda}{\Delta\lambda} \approx \frac{589.3}{0.60} \approx 982. \quad (16.91)$$

In first order, $m = 1$, so

$$N_{\min} = \left\lceil \frac{589.3}{0.60} \right\rceil = 983. \quad (16.92)$$

Thus about one thousand coherently illuminated grooves are required in the idealized model. A practical instrument would normally include margin because the entrance slit and optical system also broaden spectral features.

Example 16.16: Illuminated width required for a target resolving power

A grating has 1800 lines per millimeter and is used in second order at 250 nm under normal incidence. Estimate the illuminated width required for an ideal resolving power $R = 1.0 \times 10^5$.

Solution. The required groove count is

$$N = \frac{R}{|m|} = \frac{1.0 \times 10^5}{2} = 5.0 \times 10^4. \quad (16.93)$$

With 1800 grooves per millimeter, the illuminated width is

$$W = \frac{N}{1800 \text{ mm}^{-1}} \approx 27.8 \text{ mm}. \quad (16.94)$$

The existence check is essential: $d = 556 \text{ nm}$ and $2\lambda/d = 0.900$, giving $\theta_2 \approx 64.2^\circ$. Thus second order is accessible at the specified ultraviolet wavelength. About 28 mm of illuminated width along the grating meets the ideal groove-count requirement; the same grating would not support second order at 500 nm under normal incidence.

Guided checks

1. If the number of illuminated grooves doubles, what happens to ideal resolving power?
2. At fixed N , why can a higher diffraction order provide greater resolving power?
3. Why does an entrance slit that is too wide degrade practical spectrometer resolution even if $|m|N$ is large?
4. What is the conceptual difference between $d\theta/d\lambda$ and $\lambda/\Delta\lambda$?

16.15 Order overlap and free spectral range

A grating angle can satisfy the constructive-interference condition for more than one pair (m, λ) . For example, second-order light of wavelength λ appears at the same angle as first-order light of wavelength 2λ because

$$2\lambda = 1(2\lambda). \quad (16.95)$$

This order overlap is a central issue in real spectrometers.

A useful scale is the free spectral range: the wavelength interval that can be observed in a given order before neighboring orders overlap. Around order m and wavelength λ , a common high-order approximation is

$$\Delta\lambda_{\text{FSR}} \approx \frac{\lambda}{m}. \quad (16.96)$$

The exact interval depends on how the boundaries are defined. The inverse dependence on order is the key trade-off: high order improves resolving power but reduces the non-overlapping spectral bandwidth. Echelle spectrographs address this by using cross-dispersion to separate orders in a second detector direction.

Note

Resolution and free spectral range pull instrument design in different directions. High diffraction order can increase $R = |m|N$, but neighboring orders become closer in wavelength coverage. Spectrometer design therefore balances groove density, order, beam size, detector dimensions, optical efficiency and bandwidth.

16.16 X-ray diffraction and Bragg's law

Visible-light wavelengths are hundreds of nanometers, much larger than atomic spacings. To probe crystal structures with wave diffraction, one needs wavelengths comparable to the separation of atomic planes, typically fractions of a nanometer. X rays naturally occupy this scale.

A crystal can be idealized as a set of parallel planes separated by distance d_{hkl} . Consider radiation incident at a glancing angle θ_B relative to the planes. A ray scattered from a deeper plane travels an extra distance equal to two segments, each of length $d_{hkl} \sin \theta_B$. The total path difference is

$$\Delta = 2d_{hkl} \sin \theta_B. \quad (16.97)$$

Constructive interference occurs when this equals an integer number n of wavelengths:

$$2d_{hkl} \sin \theta_B = n\lambda, \quad n = 1, 2, 3, \dots \quad (16.98)$$

This is Bragg's law. The planes are a geometrical bookkeeping device for elastic scattering by atoms; they are not literal partially reflecting mirrors. Figure 16.10 shows the two extra path segments and defines the measured scattering angle.

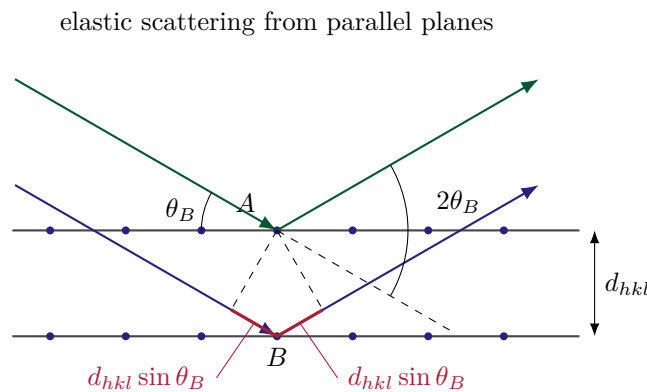


Figure 16.10: Bragg path construction. Perpendicular dashed projections from A identify the two red extra path segments for scattering at B . Their sum is $2d_{hkl} \sin \theta_B$. The detector deflection $2\theta_B$ is measured from the continuation of the incident propagation direction.

A common experimental convention reports the detector angle $2\theta_B$, the angle between the incident and diffracted beam directions. Confusing θ_B with $2\theta_B$ is a frequent source of factor-of-two errors.

Example 16.17: Crystal-plane spacing from an X-ray peak

Monochromatic X rays of wavelength 0.154 nm produce a first-order Bragg peak at a scattering angle $2\theta_B = 44.0^\circ$. Find the plane spacing.

Solution. The Bragg angle is half the reported scattering angle:

$$\theta_B = 22.0^\circ. \quad (16.99)$$

For first order, $n = 1$, so

$$d_{hkl} = \frac{\lambda}{2 \sin \theta_B} = \frac{0.154 \text{ nm}}{2 \sin 22.0^\circ} \approx 0.206 \text{ nm}. \quad (16.100)$$

The sub-nanometer scale is consistent with atomic lattice spacings.

Example 16.18: Which Bragg orders are possible?

A crystal has plane spacing $d = 0.250 \text{ nm}$ and is illuminated by X rays of wavelength 0.100 nm. Which Bragg orders are possible in principle?

Solution. Bragg's law requires

$$\sin \theta_B = \frac{n\lambda}{2d} = \frac{n(0.100)}{0.500} = 0.200n. \quad (16.101)$$

The condition $\sin \theta_B \leq 1$ gives $n \leq 5$. Thus orders $n = 1, 2, 3, 4, 5$ are geometrically possible, with the fifth at $\theta_B = 90^\circ$. In a real crystal, the intensity of a nominally allowed order also depends on the crystal structure factor, atomic form factors and experimental geometry; geometry alone does not guarantee a strong measured peak.

Caution

Bragg's law gives the geometric condition for constructive interference from a family of planes. It does not predict peak intensity by itself. Peak strength depends on which atoms occupy the unit cell, their scattering factors, thermal motion, multiplicity, polarization and instrument response. Some geometrically allowed reflections can be systematically weak or absent.

Guided checks

1. Why are X rays, rather than visible light, suitable for ordinary crystal-lattice diffraction?
2. If d_{hkl} decreases while λ and order are fixed, does the Bragg angle increase or decrease?
3. Why is a measured 2θ value not the angle that should be inserted directly into $2d \sin \theta = n\lambda$?
4. Does satisfying Bragg's law guarantee that a reflection will have large intensity?

16.16.1 Reading an X-ray diffraction scan

A powder X-ray diffraction pattern is usually displayed as intensity versus 2θ . The peak position is related to lattice-plane spacing by Bragg's law, while the peak intensity depends on additional information including which atoms occupy the unit cell, their scattering strengths, multiplicities and experimental geometry. Peak width can also carry information, but it is influenced by both sample and instrument.

A disciplined first interpretation therefore separates three observables:

Peak position: primarily constrains spacings and lattice parameters after the reflection is indexed.

Peak intensity: carries structural and texture information and should not be interpreted from Bragg's law alone.

Peak width: can reflect finite coherent domain size, strain, instrumental resolution and wavelength distribution.

Example 16.19: A fractional lattice-spacing change from a peak shift

A first-order peak measured with fixed wavelength moves from $2\theta = 40.00^\circ$ to $2\theta = 40.20^\circ$. Estimate the fractional change in the corresponding plane spacing using differential reasoning.

Solution. Since

$$d = \frac{\lambda}{2 \sin \theta}, \quad (16.102)$$

taking a differential gives

$$\frac{dd_{hkl}}{d_{hkl}} = -\cot \theta_B d\theta_B. \quad (16.103)$$

The initial Bragg angle is $\theta = 20.00^\circ$. The change in the reported 2θ is 0.20° , so $d\theta = 0.10^\circ = 1.745 \times 10^{-3}$ rad. Therefore

$$\frac{\Delta d}{d} \approx -\cot 20.00^\circ (1.745 \times 10^{-3}) \approx -4.80 \times 10^{-3}. \quad (16.104)$$

The spacing decreases by approximately 0.48%. The negative sign is physically sensible: at fixed wavelength, a larger Bragg angle requires a smaller spacing.

16.17 Laboratory investigation of single-slit diffraction

A simple diffraction laboratory can test several central ideas with inexpensive equipment: a low-power visible laser, a slit of known or adjustable width, a screen, a ruler or camera, and a measured propagation distance.

16.17.1 Experimental procedure

A robust procedure is:

1. Mount the laser so its beam is approximately normal to the slit.
2. Place the screen at a measured distance L large enough that the central pattern is several millimeters wide.
3. Mark or record the central maximum and several dark minima on both sides.
4. Use the symmetry of positive and negative orders to reduce sensitivity to locating the exact optical axis.
5. For each order m , define a half-separation y_m from the center using corresponding minima on opposite sides.
6. Convert to angle using $\theta_m = \arctan(y_m/L)$ rather than immediately assuming small angles.
7. Fit $\sin \theta_m$ versus m . The slope should be λ/a .

Using multiple minima is preferable to calculating a from only one fringe because a fit can average random reading errors and expose systematic deviations. Figure 16.11 defines L and the symmetric

half-separations. Calibration errors in L and λ are shared by every fringe and do not shrink as if each order supplied a new independent calibration. If the same estimated center is subtracted from every position, their errors are correlated; fitting signed positions with a free center parameter avoids concealing that correlation. Curved residuals suggest an inadequate angle approximation, whereas left-right asymmetry calls for an alignment check.

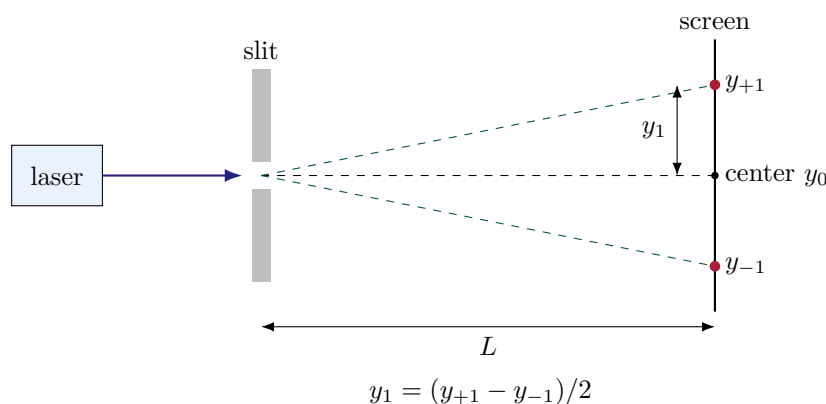


Figure 16.11: Single-slit measurement geometry. Dashed construction lines identify the first minima, not bright rays. Symmetric minima define their half-separation without requiring the center to be read separately. A free-center fit to signed positions is an alternative when several orders are available.

16.17.2 Uncertainty propagation

Suppose the small-angle formula is used for the first minimum,

$$a \approx \frac{\lambda L}{y_1}. \quad (16.105)$$

For independent small relative uncertainties, an approximate fractional uncertainty is

$$\left(\frac{\sigma_a}{a}\right)^2 \approx \left(\frac{\sigma_\lambda}{\lambda}\right)^2 + \left(\frac{\sigma_L}{L}\right)^2 + \left(\frac{\sigma_{y_1}}{y_1}\right)^2. \quad (16.106)$$

In a typical teaching laboratory, wavelength uncertainty may be small compared with the uncertainty in identifying the dark-fringe position. Increasing L increases y_1 and can reduce the fractional impact of a fixed millimeter-scale position error, provided the setup remains stable and sufficiently dark.

Example 16.20: Uncertainty in a slit-width measurement

A 635.0 ± 1.0 nm laser is used with $L = 1.500 \pm 0.003$ m. The first-minimum displacement is $y_1 = 9.50 \pm 0.10$ mm. Estimate the slit width and its standard uncertainty using the small-angle model.

Solution. The width is

$$a \approx \frac{\lambda L}{y_1} = \frac{(635.0 \times 10^{-9})(1.500)}{9.50 \times 10^{-3}} \approx 1.003 \times 10^{-4} \text{ m}. \quad (16.107)$$

The relative uncertainty is

$$\frac{\sigma_a}{a} \approx \left[\left(\frac{1.0}{635.0}\right)^2 + \left(\frac{0.003}{1.500}\right)^2 + \left(\frac{0.10}{9.50}\right)^2 \right]^{1/2} \quad (16.108)$$

$$\approx 0.0108. \quad (16.109)$$

Thus

$$\sigma_a \approx 1.1 \mu\text{m}, \quad (16.110)$$

so a reasonable result is

$$a = (100.3 \pm 1.1) \mu\text{m}. \quad (16.111)$$

The fringe-position uncertainty dominates the error budget.

16.17.3 Estimating a slit width from several diffraction minima

For a slit of width a , the minima obey

$$a \sin \theta_m = m\lambda. \quad (16.112)$$

If a screen is a distance L away and the pattern is paraxial,

$$y_m \approx L\theta_m \approx \frac{mL\lambda}{a}. \quad (16.113)$$

A single measured minimum can therefore determine a , but a better experiment uses several positive and negative orders. Plotting y_m versus m should produce a straight line with slope

$$s = \frac{L\lambda}{a}, \quad (16.114)$$

so that

$$a = \frac{L\lambda}{s}. \quad (16.115)$$

This procedure has several advantages. Random reading errors in individual fringe positions are averaged over many data points, an offset in the chosen origin can be identified through the fitted intercept, and curvature in the plot can reveal that the small-angle approximation is becoming inadequate.

Example 16.21: Slit width from a linear fit

A 532 nm laser illuminates an unknown slit. The screen is $L = 1.50$ m from the slit. For an illustrative data set, a least-squares fit of minimum position to order gives

$$y_m = (5.31 \text{ mm})m + (0.08 \text{ mm}).$$

Estimate the slit width and interpret the nonzero intercept.

Solution. The fitted slope is $s = 5.31 \times 10^{-3}$ m per order. Equation (16.115) gives

$$a = \frac{(1.50)(532 \times 10^{-9})}{5.31 \times 10^{-3}} \approx 1.50 \times 10^{-4} \text{ m} = 0.150 \text{ mm}. \quad (16.116)$$

The 0.08 mm intercept is not evidence for a new diffraction effect. It most naturally indicates that the coordinate origin used to read the screen was displaced slightly from the true optical axis. Fitting an intercept rather than forcing the line through zero prevents that alignment offset from biasing the inferred slit width.

Method

When a physical law predicts a linear relation, use the slope whenever possible rather than solving the same formula independently for every data point. A fitted slope uses the entire data set, gives a natural uncertainty estimate, and exposes systematic departures from the assumed model.

16.17.4 Uncertainty propagation in slit metrology

Suppose the slit width is inferred from

$$a = \frac{L\lambda}{s}. \quad (16.117)$$

For independent small uncertainties, the fractional standard uncertainty is approximately

$$\left(\frac{\sigma_a}{a}\right)^2 \approx \left(\frac{\sigma_L}{L}\right)^2 + \left(\frac{\sigma_\lambda}{\lambda}\right)^2 + \left(\frac{\sigma_s}{s}\right)^2. \quad (16.118)$$

This expression immediately shows where experimental effort is worthwhile. If the laser wavelength is known to much better than 0.1% but the fringe-position slope is uncertain by 2%, improving the wavelength calibration has almost no effect on the final uncertainty. The measurement should instead improve the screen-position determination, increase the propagation distance, or include more well-resolved minima.

The same logic applies throughout diffraction science. An instrument is not characterized only by a nominal resolution formula; it also has calibration uncertainties, alignment errors, finite detector sampling and environmental perturbations.

Example 16.22: A compact uncertainty budget

In the previous experiment let the relative uncertainties be $\sigma_L/L = 0.3\%$, $\sigma_\lambda/\lambda = 0.05\%$, and $\sigma_s/s = 1.2\%$. Estimate the fractional uncertainty in a .

Solution.

$$\frac{\sigma_a}{a} \approx \sqrt{(0.003)^2 + (0.0005)^2 + (0.012)^2} \quad (16.119)$$

$$\approx 0.0124. \quad (16.120)$$

Thus the relative uncertainty is about 1.24%. For $a = 0.150$ mm,

$$\sigma_a \approx 0.0019 \text{ mm}. \quad (16.121)$$

A reasonable report is therefore $a = (0.150 \pm 0.002)$ mm for this simplified uncertainty model. The dominant contribution is the fitted fringe-position slope.

16.17.5 Computational fitting of the full intensity profile

Instead of locating minima, one may fit the measured intensity profile directly to

$$I(y) = I_{\text{bg}} + A \left[\frac{\sin \beta(y)}{\beta(y)} \right]^2, \quad \beta(y) = \frac{\pi a}{\lambda} \sin \left(\arctan \frac{y - y_0}{L} \right). \quad (16.122)$$

Here A is an amplitude scale, I_{bg} is a background level, and y_0 is the pattern center. The slit width a is the physically interesting parameter. For the angular mapping, the exact geometric relation

$$\sin \theta = \frac{y - y_0}{\sqrt{L^2 + (y - y_0)^2}} \quad (16.123)$$

can be used without approximating the angle. The fitted intensity model still assumes Fraunhofer propagation, uniform illumination, linear detector response and nearly constant geometrical prefactors over the measured field. At large angles, a planar detector also needs distance, projection and obliquity factors; exact coordinate conversion alone does not supply them.

Full-profile fitting uses information from all sampled points rather than only the minima, but it also demands a more realistic detector model. Saturated central pixels, finite pixel width,

nonuniform illumination, camera nonlinearities and background light can bias the result if ignored. The computational lesson is general: using more data can improve precision only when the model for those data is adequate.

Example 16.23: Why saturation biases a profile fit

Suppose a camera saturates near the center of a diffraction pattern so that the true central maximum is clipped to a flat plateau. If these saturated pixels are included in an unmodified sinc-squared least-squares fit, what qualitative bias might occur?

Discussion. The model predicts a sharp central maximum whose height and curvature are linked to the width parameter. Clipped data make the central region artificially broad and flat. The optimizer can partly mimic that shape by altering amplitude and width, potentially biasing the inferred slit width. The proper response is to mask saturated pixels, reduce exposure, or explicitly model detector saturation rather than treating clipped values as accurate intensity measurements.

16.18 Computational exploration of diffraction patterns

Diffraction is particularly suitable for computational visualization because the formulas are simple but the pattern shapes are difficult to appreciate from equations alone. A short program can plot the normalized single-slit intensity

$$\frac{I}{I_0} = \left(\frac{\sin \beta}{\beta} \right)^2 \quad (16.124)$$

for a range of angles, and can then multiply it by an interference or grating factor.

A numerically stable implementation should treat the $\beta = 0$ point using its limiting value 1 rather than evaluating $0/0$. Most scientific libraries provide a normalized sinc function, but conventions differ. Some define $\text{sinc}(x) = \sin x/x$ and others define $\text{sinc}(x) = \sin(\pi x)/(\pi x)$. Code should document the convention explicitly.

For the finite grating, use Eq. (16.77); removable singularities at principal orders must be evaluated by continuity.

A useful computational study varies one parameter at a time:

- vary a to see the envelope narrow or widen;
- vary d to change principal-maximum spacing;
- vary N to sharpen principal maxima without changing their central angles;
- vary λ to compare spectral components;
- map angle to detector position and estimate whether a chosen pixel pitch samples the peaks adequately.

Method

For computational plots, separate *physics parameters* from *display parameters*. Compute the physical intensity on a sufficiently fine angular grid first. Only then choose plot limits, logarithmic or linear vertical scale, and detector mapping. A coarse numerical grid can falsely broaden sharp grating peaks and lead to an incorrect conclusion about resolution.

16.18.1 A reproducible finite-array calculation

The physical question is whether increasing the illuminated groove count shifts the principal orders or only narrows them. The following Python 3 code uses only the standard library. It evaluates the finite sum directly, so no 0/0 appears at a principal maximum. The coordinate $v = d \sin \theta / \lambda$ measures angle in grating-order units.

```
from math import pi, sin
from cmath import exp

def sinc_rad(x):
    return 1.0 if abs(x) < 1e-12 else sin(x)/x

def array_intensity(v, N, a_over_d=0.10):
    amplitude = sum(exp(-2j*pi*j*v) for j in range(N))/N
    envelope = sinc_rad(pi*a_over_d*v)**2
    return envelope*abs(amplitude)**2

for N in (2, 5, 20):
    print(N, array_intensity(1.0, N),
          array_intensity(1.0 + 1.0/N, N))
```

The normalized intensity at $v = 1$ is $[\sin(0.1\pi)/(0.1\pi)]^2 \approx 0.96753$ for every N , while the neighboring array zero is at $v = 1 + 1/N$. The last printed values should be zero to floating-point accuracy. A plot requires a step in v much smaller than $1/N$; otherwise a narrow peak can be missed. This test compares a physical prediction with a directly evaluated sum, rather than trusting visual smoothness. Figure 16.9 uses the same model for its array factors.

16.19 Main chapter summary

Summary

- Diffraction is coherent interference produced by a wavefront of finite transverse extent or by a periodic scattering structure.
- Fraunhofer diffraction describes the far-field angular pattern. Fresnel diffraction describes near-field situations in which propagation geometry and wavefront curvature remain important.
- A uniformly illuminated slit of width a has minima at $a \sin \theta = m\lambda$ for nonzero integer m .
- Its intensity is $I = I_0(\sin \beta / \beta)^2$ with $\beta = \pi a \sin \theta / \lambda$.
- The central single-slit maximum has small-angle width approximately $2\lambda/a$ in angle and $2L\lambda/a$ on a distant screen.
- A circular aperture produces an Airy pattern. The first dark ring lies at $\theta \approx 1.22\lambda/D$ for small angles.
- Rayleigh's criterion uses this diffraction scale as a conventional estimate of just-resolved point sources.
- Microscope resolution scales approximately as λ_0/NA , so large numerical aperture and short wavelength improve conventional diffraction-limited detail.
- Two finite slits produce interference fringes multiplied by a single-slit diffraction envelope.
- A grating has principal maxima at $d \sin \theta = m\lambda$. Its angular dispersion is $m/(d \cos \theta)$.
- The ideal resolving power of an N -groove grating in order m is $R = |m|N$.
- Bragg's law, $2d_{hkl} \sin \theta_B = n\lambda$, connects diffraction peak angles to crystal-plane spacing.

- Diffraction limits imaging resolution but enables spectroscopy, metrology and structural characterization.

Formula summary

single-slit minima:	$a \sin \theta_m = m\lambda, \quad m = \pm 1, \pm 2, \dots$
single-slit intensity:	$I(\theta) = I_0 \left(\frac{\sin \beta}{\beta} \right)^2, \quad \beta = \frac{\pi a \sin \theta}{\lambda}$
small-angle central width:	$\Delta\theta_{\text{central}} \approx \frac{2\lambda}{a}, \quad w_{\text{central}} \approx \frac{2L\lambda}{a}$
circular first minimum:	$\theta_1 \approx 1.22 \frac{\lambda}{D}$
Airy-disk diameter at focus:	$d_{\text{Airy}} \approx 2.44\lambda N_f$
Rayleigh angle:	$\theta_R \approx 1.22 \frac{\lambda}{D}$
microscope Rayleigh scale:	$\delta_R \approx 0.61 \frac{\lambda_0}{\text{NA}}$
double slit with envelope:	$I = I_0 \left(\frac{\sin \beta}{\beta} \right)^2 \cos^2 \gamma, \quad \gamma = \frac{\pi d \sin \theta}{\lambda}$
grating maxima:	$d \sin \theta_m = m\lambda$
angular dispersion:	$\mathcal{D} = \frac{d\theta}{d\lambda} = \frac{m}{d \cos \theta}$
grating resolving power:	$R = \frac{\lambda}{\Delta\lambda} = m N$
Bragg law:	$2d_{hkl} \sin \theta_B = n\lambda.$

Take-home message

Finite apertures create finite angular widths. That single statement links slit diffraction, Airy disks, telescope and microscope resolution, finite grating linewidths, and the practical information capacity of optical instruments. Periodic structures then turn the same wave-superposition physics into a precision measurement tool for wavelength and crystal spacing.

16.20 Advanced Topics for Technical and Engineering Students

The following modules are optional extensions. They organize diffraction into four connected problems: propagation of a complex field, image formation through a finite pupil, numerical sampling, and scattering from periodic matter. The core chapter remains sufficient for the standard slit, grating, Rayleigh and Bragg calculations.

16.20.1 From the scalar wave equation to the diffraction integral

The Fourier-transform description of Fraunhofer diffraction can be placed in a broader wave-theory context. For monochromatic scalar waves one writes the physical field as the real part of a complex amplitude $U(\mathbf{r})e^{-i\omega t}$. In a homogeneous source-free region the spatial amplitude obeys the Helmholtz equation

$$\nabla^2 U + k^2 U = 0, \quad k = \frac{2\pi}{\lambda}. \quad (16.125)$$

A rigorous electromagnetic treatment must also respect vector boundary conditions and polarization, but scalar diffraction is remarkably successful when aperture dimensions are large compared with wavelength and polarization-dependent edge effects are not the central issue.

The Huygens–Fresnel idea can be expressed schematically as an integral over an aperture plane:

$$U(P) \propto \iint_{\mathcal{A}} U(x', y') \frac{e^{ikr}}{r} K(\theta) dx' dy', \quad (16.126)$$

where r is the distance from aperture point (x', y') to observation point P , and $K(\theta)$ represents an obliquity factor whose detailed form depends on the chosen scalar diffraction formulation. The central physics is contained in the rapidly varying phase kr .

For an observation plane a distance z away, with transverse coordinates (x, y) , the distance is

$$r = \sqrt{z^2 + (x - x')^2 + (y - y')^2}. \quad (16.127)$$

When transverse distances are small compared with z , expand the square root:

$$r = z \sqrt{1 + \frac{(x - x')^2 + (y - y')^2}{z^2}} \quad (16.128)$$

$$\approx z + \frac{(x - x')^2 + (y - y')^2}{2z}. \quad (16.129)$$

Keeping the quadratic terms leads to Fresnel diffraction. Expanding those squares gives

$$(x - x')^2 + (y - y')^2 = x^2 + y^2 + x'^2 + y'^2 - 2xx' - 2yy'. \quad (16.130)$$

If z is sufficiently large that the aperture-dependent quadratic phase

$$\exp \left[\frac{ik}{2z} (x'^2 + y'^2) \right] \quad (16.131)$$

varies negligibly across the transmitting region, it may be approximated as unity. What remains in the aperture integral is

$$U(x, y) \propto \iint U(x', y') \exp \left[-i \frac{2\pi}{\lambda z} (xx' + yy') \right] dx' dy', \quad (16.132)$$

apart from factors depending only on the observation point. Equation (16.132) is a Fourier transform with spatial frequencies

$$f_x = \frac{x}{\lambda z}, \quad f_y = \frac{y}{\lambda z}. \quad (16.133)$$

Thus the far-field Fourier relation is not a separate postulate. It is the leading far-zone form of the same propagation integral that describes near-field diffraction.

Pedagogical boundary

The scalar diffraction integral is introduced here as an advanced model, not as a complete derivation of electromagnetic boundary diffraction from Maxwell's equations. Near subwavelength structures, strongly vectorial focusing, anisotropic media and polarization-sensitive nanophotonics require a vector treatment beyond this course.

16.20.2 Complex amplitude and the Fourier-transform viewpoint

The single-slit derivation already contained the essential Fourier-optics structure. Let the complex field immediately after an aperture in a transverse coordinate x be

$$U(x, 0) = A(x), \quad (16.134)$$

where $A(x)$ includes both the illumination and the aperture transmission. In the Fraunhofer regime, the field as a function of observation angle is, apart from overall phase and scale factors, the Fourier transform of the aperture field:

$$U(\theta) \propto \int_{-\infty}^{\infty} A(x) \exp[-ikx \sin \theta] dx. \quad (16.135)$$

The spatial-frequency variable is therefore proportional to $\sin \theta/\lambda$.

For a uniform slit,

$$A(x) = \begin{cases} A_0, & |x| \leq a/2, \\ 0, & |x| > a/2, \end{cases} \quad (16.136)$$

which is a rectangular function. The Fourier transform of a rectangle is a sinc function, explaining why the slit intensity is sinc-squared.

This perspective provides a powerful general rule:

Narrow structure in one domain corresponds to broad structure in the Fourier-conjugate domain, while broad smooth structure corresponds to narrow angular or spatial-frequency content.

A narrow slit therefore has a broad diffraction pattern. A wide grating containing many periods has narrow spectral peaks. Sharp aperture edges produce extended side lobes because abrupt spatial changes require a wide range of spatial frequencies.

Theoretical point 16.1: Diffraction as spatial-frequency analysis

In scalar Fraunhofer diffraction, the aperture field and the far-field complex amplitude form an approximate Fourier-transform pair. Intensity is the squared magnitude of that transform. This is the optical analogue of frequency analysis in signals: the aperture is a spatial signal, and diffraction reveals its spatial-frequency content.

Example 16.24: Why doubling slit width halves the diffraction scale

Suppose an aperture function $A(x)$ is replaced by $A(x/s)$, so its physical width is stretched by a factor $s > 1$. Show qualitatively why the angular diffraction pattern narrows by the same factor in the small-angle limit.

Solution. Insert the stretched function into the Fourier integral:

$$U_s(q) \propto \int A(x/s) e^{-iqx} dx, \quad (16.137)$$

where $q \approx k\theta$. Let $u = x/s$, so $x = su$ and $dx = sdu$. Then

$$U_s(q) \propto s \int A(u) e^{-i(sq)u} du = sU(sq). \quad (16.138)$$

The amplitude scale changes by s , but the angular dependence is compressed: a feature that occurred at q_0 now occurs at q_0/s . Since $q \propto \theta$ for small angles, the diffraction angle scales as $1/s$. Doubling the aperture width therefore halves the characteristic diffraction angle.

16.20.3 Two-dimensional rectangular apertures

For an aperture with independent rectangular widths a in x and b in y , the aperture function factorizes:

$$A(x, y) = A_x(x)A_y(y). \quad (16.139)$$

Its two-dimensional Fourier transform also factorizes. In small-angle coordinates θ_x and θ_y ,

$$I(\theta_x, \theta_y) = I_0 \left[\frac{\sin \beta_x}{\beta_x} \right]^2 \left[\frac{\sin \beta_y}{\beta_y} \right]^2, \quad (16.140)$$

with

$$\beta_x = \frac{\pi a \sin \theta_x}{\lambda}, \quad \beta_y = \frac{\pi b \sin \theta_y}{\lambda}. \quad (16.141)$$

A rectangular aperture therefore produces a cross-like two-dimensional pattern with different widths in the two transverse directions when $a \neq b$.

This is an instructive engineering example. If a beam is clipped by a narrow rectangular dimension, the far field spreads most strongly perpendicular to that narrow dimension. Aperture anisotropy becomes angular anisotropy.

Guided checks

1. A rectangular aperture is very wide horizontally and narrow vertically. In which angular direction is diffraction broadest?
2. Why does separability of the aperture function produce a product of two sinc-squared factors?
3. If both aperture dimensions are doubled, what happens to the angular area occupied by the central lobe?

16.20.4 Circular pupils and the Bessel-function Airy pattern

The circular-aperture result can be developed more quantitatively. Let the aperture radius be $R = D/2$. Because the pupil is rotationally symmetric, the two-dimensional Fourier integral is naturally written in polar coordinates. Writing $q = k \sin \theta$ and dividing by the on-axis aperture area gives

$$\frac{U(q)}{U(0)} = \frac{1}{\pi R^2} \int_0^R \rho d\rho \int_0^{2\pi} e^{-iq\rho \cos \varphi} d\varphi \quad (16.142)$$

$$= \frac{2}{R^2} \int_0^R \rho J_0(q\rho) d\rho = \frac{2J_1(qR)}{qR}. \quad (16.143)$$

Here J_0 is defined by the angular integral, and $d[uJ_1(u)]/du = uJ_0(u)$ evaluates the radial integral. Since $J_1(u) \sim u/2$ near zero, the normalized amplitude tends to one. Squaring its modulus gives

$$\boxed{\frac{I(\theta)}{I_0} = \left[\frac{2J_1(u)}{u} \right]^2, \quad u = \frac{\pi D \sin \theta}{\lambda}.} \quad (16.144)$$

The first positive zero of $J_1(u)$ occurs at

$$u_1 \approx 3.8317. \quad (16.145)$$

Therefore

$$\sin \theta_1 = \frac{u_1}{\pi} \frac{\lambda}{D} \approx 1.2197 \frac{\lambda}{D}, \quad (16.146)$$

which explains the familiar factor 1.22.

The Bessel form also makes clear that the Airy pattern is not a Gaussian spot. Its rings decay more slowly than a Gaussian tail, and those side lobes can matter when a faint object lies near a bright one. In astronomy and high-contrast imaging, the detailed point-spread function is therefore as important as the nominal central-disk diameter.

Note

The label “Airy disk” usually refers to the bright central region out to the first dark ring. The *Airy pattern* includes all surrounding rings. A point-spread function describes the full intensity response to a point object, not only the central disk.

16.20.5 Point-spread functions and convolution

An imaging system maps a point object to a finite intensity distribution called the point-spread function (PSF). For a diffraction-limited circular pupil under standard incoherent imaging assumptions, the PSF is proportional to the Airy intensity pattern.

Express object and image coordinates in the same plane by absorbing geometrical magnification into the coordinate scale, and normalize $\iint h \, dx dy = 1$. If the intensity system is linear and shift-invariant over the field region being considered, an extended object intensity $O(x, y)$ produces an image that can be modelled as a convolution:

$$I_{\text{image}}(x, y) = \iint O(x', y') h(x - x', y - y') \, dx' dy', \quad (16.147)$$

where h is the intensity PSF.

This relation gives a deeper meaning to diffraction blur. The image is not simply “less sharp” in a vague sense. It is the object filtered by the spatial response of the optical system. Fine object features correspond to high spatial frequencies. A finite pupil restricts which of those frequencies are transmitted with significant contrast.

The Fourier transform of the PSF is the optical transfer function (OTF), and its magnitude is the modulation transfer function (MTF). The MTF quantifies how contrast at different spatial frequencies is transmitted. Resolution criteria such as Rayleigh provide one scalar benchmark, whereas the MTF provides a richer system-level description.

Key message

Rayleigh resolution is a convenient single number. The PSF and MTF contain more information: they describe how the system distributes light from a point and how it transfers contrast across spatial frequency.

16.20.6 Abbe’s spatial-frequency picture of microscopy

Consider a thin periodic object of period p in an object medium of index n . A plane wave incident at angle α_i has transverse wave vector $k_{x,i} = (2\pi n/\lambda_0) \sin \alpha_i$. The object adds integer multiples of $2\pi/p$, so its diffracted orders obey

$$k_{x,m} = k_{x,i} + \frac{2\pi m}{p}. \quad (16.148)$$

The objective accepts only $|k_x| \leq (2\pi/\lambda_0)\text{NA}_{\text{obj}}$. Interference between an accepted zero order and at least one accepted first order is needed to reproduce the basic object periodicity.

For *on-axis coherent illumination*, $k_{x,i} = 0$. Acceptance of the first order then requires $2\pi/p \leq 2\pi\text{NA}_{\text{obj}}/\lambda_0$, or

$$p \gtrsim \frac{\lambda_0}{\text{NA}_{\text{obj}}}.$$

There is no factor of two in this on-axis result. If illumination supplies oblique incident directions up to NA_{illum} , the incident spectrum is shifted, and an appropriate first order can enter the opposite edge of the objective. The largest accessible transverse-wave-vector difference is then

$$\frac{2\pi}{\lambda_0}(\text{NA}_{\text{obj}} + \text{NA}_{\text{illum}}).$$

For matched illumination with $\text{NA}_{\text{illum}} \leq \text{NA}_{\text{obj}}$, this gives the ideal periodic-object scale

$$p_{\text{min}} \sim \frac{\lambda_0}{\text{NA}_{\text{obj}} + \text{NA}_{\text{illum}}}. \quad (16.149)$$

Matched condenser and objective apertures yield $\lambda_0/(2\text{NA})$. This is the illumination assumption behind the common Abbe expression. Contrast close to a cutoff can be very small, and a partially coherent transmitted-light microscope needs a fuller transfer model for quantitative predictions.

In incoherent fluorescence imaging, there is a separate route to the factor two: the intensity transfer function is an autocorrelation of the pupil and extends to twice the coherent *amplitude* bandwidth. The equality of scales does not make the two imaging models identical. Figure 16.12 shows the pupil acceptance argument.

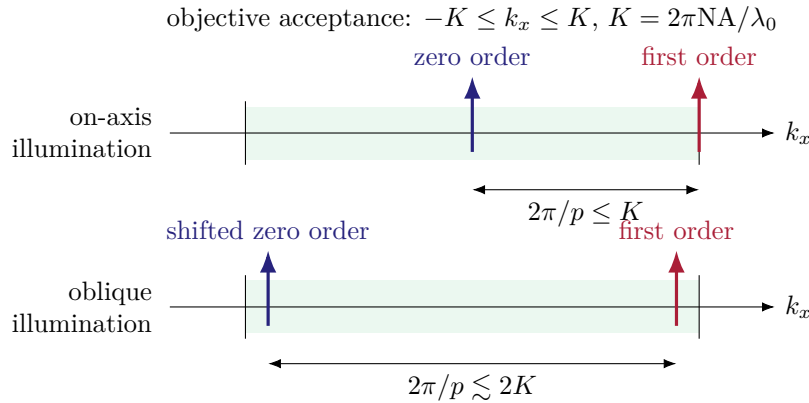


Figure 16.12: Origin of the illumination-dependent Abbe factor. With on-axis illumination the zero order is at the pupil center; the first order can be at most K away. Oblique illumination shifts the pair so that their separation can approach the full pupil width $2K$, provided both relevant orders are accepted.

Example 16.25: Comparing wavelength and NA improvements

For matched illumination and objective NA, a microscope at $\lambda_0 = 600 \text{ nm}$ and $\text{NA} = 0.75$ has an Abbe period scale $600/[2(0.75)] = 400 \text{ nm}$. Changing to 450 nm at fixed NA gives 300 nm ; increasing NA to 1.00 at 600 nm also gives 300 nm . Both changes improve the scale by 25% . These are bandwidth estimates, not guarantees of useful contrast at the limiting period, and the comparison presumes the illumination remains matched.

16.20.7 Coherent and incoherent imaging

The mathematical imaging model depends on coherence. In coherent imaging, complex field amplitudes are propagated through the system and only then squared to obtain intensity. In

incoherent imaging, intensities from mutually incoherent object points add, and the intensity PSF is the relevant convolution kernel.

This distinction changes the transfer function and spatial-frequency cutoff. It also changes the appearance of edges and periodic structures. The central undergraduate formulas for Airy patterns usually describe the diffraction response of a coherent point source or the intensity PSF under conditions where the circular pupil governs the same characteristic angular scale. A full treatment of imaging coherence requires a separate optics course.

For engineering work, the lesson is to state the illumination model. Laser illumination can create coherent speckle and interference artifacts that are absent under broadband or spatially incoherent illumination. A nominal diffraction-limited resolution number alone does not specify image texture or contrast behavior.

Pedagogical boundary

This chapter does not develop the mutual-coherence function, van Cittert–Zernike theorem or partially coherent transfer functions. Those tools are important in advanced imaging and lithography, but the present goal is to recognize that coherence changes how diffraction propagates into an image.

16.20.8 Optical transfer function and modulation transfer function

The point-spread function tells how an imaging system responds to a point. A complementary description asks how well the system transfers sinusoidal object patterns of different spatial frequency. For incoherent imaging, the Fourier transform of the intensity PSF is the *optical transfer function* (OTF). Its magnitude is the *modulation transfer function* (MTF).

Consider an object with intensity varying as

$$I_{\text{obj}}(x) = I_0 [1 + C \cos(2\pi f_x x)], \quad (16.150)$$

where C is contrast and f_x is spatial frequency. An imaging system generally reduces the contrast of fine patterns. In a linear shift-invariant incoherent model,

$$C_{\text{image}} = \text{MTF}(f_x) C_{\text{object}}. \quad (16.151)$$

The MTF therefore gives a frequency-by-frequency description of image sharpness rather than a single threshold angle.

For an ideal incoherent diffraction-limited circular pupil, the normalized MTF can be written

$$\text{MTF}(\nu) = \frac{2}{\pi} \left[\cos^{-1} \nu - \nu \sqrt{1 - \nu^2} \right], \quad 0 \leq \nu \leq 1, \quad (16.152)$$

with $\text{MTF} = 0$ for $\nu > 1$. Here $\nu = f/f_c$ is normalized spatial frequency. In object space for ideal incoherent microscopy,

$$f_c = \frac{2 \text{NA}}{\lambda_0} \quad (16.153)$$

is the cutoff in cycles per unit object-space length; λ_0 is the vacuum wavelength because NA already contains the refractive index. The coherent *field-amplitude* cutoff for on-axis illumination is NA/λ_0 . Squaring a coherent field is nonlinear in that field, so it is incorrect to apply the incoherent intensity MTF to every coherently illuminated object.

For the clear pupil, the incoherent OTF is the normalized overlap area of two identical pupil disks shifted relative to one another. It is one at zero shift and vanishes when the disks cease to overlap. The overlap area gives Eq. (16.152), plotted in Figure 16.13. This geometrical construction explains both the continuous loss of contrast and the finite support.

Equation (16.152) makes two important ideas visible. First, contrast does not remain perfect up to a sudden Rayleigh boundary; it decreases continuously with spatial frequency. Second, spatial frequencies above the diffraction cutoff are absent in this ideal linear model rather than merely weakened.

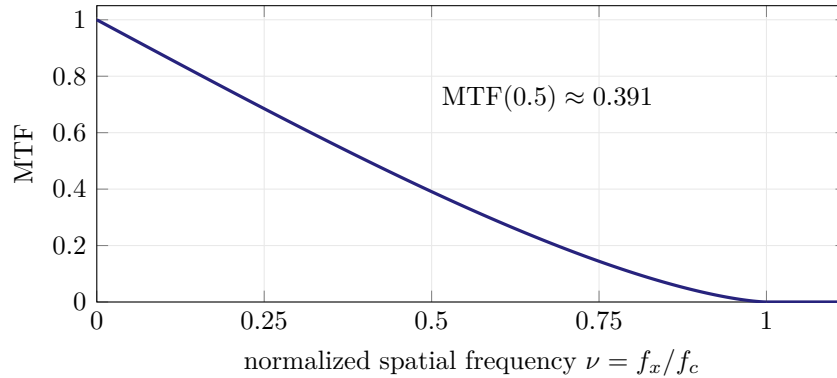


Figure 16.13: Calculated incoherent intensity MTF of an ideal clear circular pupil. Contrast falls continuously and reaches zero at the cutoff; a period exactly at cutoff is not transferred with finite contrast. Object-space $f_c = 2NA/\lambda_0$.

Example 16.26: Object period at the incoherent cutoff

For $\lambda_0 = 550$ nm and $NA = 0.80$, estimate the cutoff period $1/f_c$.

Solution.

$$f_c = \frac{2(0.80)}{550 \times 10^{-9}} \approx 2.91 \times 10^6 \text{ m}^{-1}. \quad (16.154)$$

Hence

$$\frac{1}{f_c} \approx 3.44 \times 10^{-7} \text{ m} = 344 \text{ nm}. \quad (16.155)$$

This period is not numerically identical to the Rayleigh separation because the two criteria describe different test objects and conventions. At exactly f_c the ideal MTF is zero, so this limiting period has zero transmitted contrast. A visible or statistically measurable pattern must lie below cutoff with sufficient contrast and signal-to-noise. Both scales nevertheless arise from the same finite pupil.

16.20.9 Detector sampling, Nyquist reasoning and aliasing

Even a perfect optical image can be recorded poorly if the detector sampling is too coarse. Suppose the optical system creates a point-spread function whose central width is comparable to or smaller than one detector pixel. The detector then records only a few samples across the image structure, and fine detail can be misrepresented.

For a one-dimensional band-limited signal with maximum spatial frequency $f_{\max}$, the Nyquist sampling criterion requires sample spacing

$$\Delta x \leq \frac{1}{2f_{\max}}. \quad (16.156)$$

Optical images are not always perfectly band-limited in a simplistic sense, but a diffraction-limited pupil does impose a finite transfer-function support under standard imaging models. The practical design rule is that a resolution element should span multiple pixels. Two samples per smallest period is a theoretical minimum; real imaging commonly uses more to support centroiding, interpolation, calibration and imperfect alignment.

Undersampling can create aliasing, where high spatial frequencies appear falsely as lower ones. Oversampling does not increase the intrinsic optical resolution, but it can make better use of the information that is present.

Example 16.27: Pixel pitch compared with an Airy disk

A sensor has $4.0\ \mu\text{m}$ pixels and is used behind an $f/5.6$ lens at $\lambda = 550\ \text{nm}$. Estimate the Airy-disk diameter in pixels.

Solution.

$$d_{\text{Airy}} \approx 2.44(550 \times 10^{-9})(5.6) \approx 7.52\ \mu\text{m}. \quad (16.157)$$

In pixel units,

$$\frac{d_{\text{Airy}}}{p} \approx \frac{7.52}{4.0} \approx 1.88\ \text{pixels}. \quad (16.158)$$

Thus the first-zero diameter spans about two pixels, but that is *not* sufficient to sample the full incoherent optical bandwidth. Equation (16.36) requires $p \leq (0.550)(5.6)/2 = 1.54\ \mu\text{m}$; $4.0\ \mu\text{m}$ pixels undersample this ideal image. Aberrations or pixel integration can reduce high-frequency contrast, but their effect must be modeled rather than inferred from the Airy diameter alone.

16.20.10 Advanced synthesis example: imaging trade-offs for a space telescope

Example 16.28: Wavelength, aperture and detector trade-off

A diffraction-limited space telescope has diameter $D = 1.20\ \text{m}$ and focal length $f = 18\ \text{m}$. It observes at $500\ \text{nm}$ with a detector pixel pitch of $6.0\ \mu\text{m}$. Estimate the Rayleigh angular scale, the Airy-disk diameter in the focal plane, and the number of pixels across that diameter.

Solution. The angular Rayleigh scale is

$$\theta_R \approx 1.22 \frac{500 \times 10^{-9}}{1.20} \approx 5.08 \times 10^{-7}\ \text{rad}. \quad (16.159)$$

The focal-plane radius to the first dark ring is

$$r_1 \approx f\theta_R = (18)(5.08 \times 10^{-7}) \approx 9.15\ \mu\text{m}. \quad (16.160)$$

Thus the Airy diameter is about

$$d_{\text{Airy}} \approx 18.3\ \mu\text{m}. \quad (16.161)$$

In detector pixels,

$$\frac{d_{\text{Airy}}}{p} \approx \frac{18.3}{6.0} \approx 3.05. \quad (16.162)$$

The central disk spans roughly three pixels. With $N_f = f/D = 15$, the full-band incoherent Nyquist bound is $p \leq \lambda N_f/2 = 3.75\ \mu\text{m}$. The $6.0\ \mu\text{m}$ detector therefore undersamples the full ideal bandwidth despite showing several pixels across the central disk. A shorter wavelength would shrink the Airy pattern and could eventually require finer pixels or a longer effective focal length to maintain comparable sampling.

16.20.11 Apodization, windowing and side-lobe control

Uniform illumination with a hard edge produces relatively strong side lobes because the aperture amplitude changes abruptly to zero. If the aperture is weighted smoothly toward its edges, side lobes can be reduced. This deliberate weighting is called apodization.

There is an unavoidable trade-off. Suppressing side lobes generally broadens the central lobe. This

is directly analogous to window functions in digital signal processing. A rectangular time window has a narrow main spectral lobe but high side lobes; smoother windows reduce side lobes at the cost of wider main lobes. Optical apodization is the spatial version of the same Fourier principle.

Applications include spectroscopy, antenna arrays, high-contrast imaging and beam shaping. The design question is not “Which aperture gives the narrowest possible main lobe?” but rather “What balance of main-lobe width, side-lobe level and throughput best supports the measurement?”

Guided checks

1. Why does smoothing an aperture edge tend to reduce far-field side lobes?
2. What is the expected cost of strong side-lobe suppression?
3. Why is apodization mathematically similar to windowing in a finite digital data record?

For example, take $A(x) = \cos^2(\pi x/a)$ for $|x| \leq a/2$ and zero outside. With $v = af_x$, its normalized amplitude transform is

$$H(v) = \text{sinc}(v) + \frac{1}{2} \text{sinc}(v-1) + \frac{1}{2} \text{sinc}(v+1), \quad \text{sinc}(v) = \frac{\sin \pi v}{\pi v}.$$

Its first zero is at $v = 2$, rather than $v = 1$ for the uniform slit. Figure 16.14 compares the calculated intensities, each normalized to its own central peak. At fixed incident field and peak transmission, this amplitude taper transmits $\int_{-a/2}^{a/2} \cos^4(\pi x/a) dx = 3a/8$, only 37.5% of the uniform slit power per unit height. Normalized plots conceal that throughput cost.

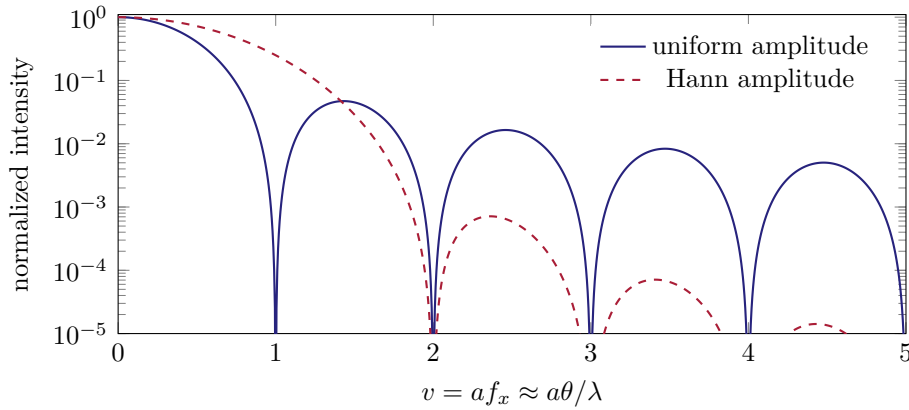


Figure 16.14: Uniform and Hann-tapered slit intensities calculated from their amplitude transforms. The logarithmic scale reveals side-lobe suppression and the broader main lobe. Exact zeros extend below the displayed range; curves are clipped only for logarithmic plotting.

16.20.12 Gaussian beams and diffraction-limited divergence

A laser beam is often better described by a Gaussian transverse profile than by a uniformly illuminated hard aperture. The fundamental Gaussian beam has a waist radius w_0 , defined conventionally as the radius where the intensity falls to $1/e^2$ of its on-axis value. Its far-field half-angle divergence in the paraxial approximation is

$$\theta_{\text{div}} \approx \frac{\lambda}{\pi w_0}. \quad (16.163)$$

The beam radius evolves as

$$w(z) = w_0 \sqrt{1 + \left(\frac{z}{z_R}\right)^2}, \quad (16.164)$$

where

$$z_R = \frac{\pi w_0^2}{\lambda} \quad (16.165)$$

is the Rayleigh range.

The coefficient in the divergence can be derived rather than guessed. At the waist, $A(x) \propto \exp(-x^2/w_0^2)$, so its Fourier amplitude is proportional to $\exp(-q^2 w_0^2/4)$ and its angular intensity to $\exp[-(k w_0 \theta)^2/2]$ in the paraxial limit. The intensity reaches e^{-2} at $k w_0 \theta = 2$, giving $\theta_{\text{div}} = \lambda/(\pi w_0)$. This derivation also fixes the radius convention. These equations express the same localization-divergence trade-off seen in slit diffraction. A smaller waist produces greater angular spread. The absence of hard edges gives a Gaussian angular profile rather than sinc side lobes.

Example 16.29: Laser-beam divergence from a small waist

A diffraction-limited Gaussian laser beam has wavelength 1064 nm and waist radius $w_0 = 0.50$ mm. Estimate the far-field half-angle divergence and Rayleigh range.

Solution.

$$\theta_{\text{div}} \approx \frac{1.064 \times 10^{-6}}{\pi(5.0 \times 10^{-4})} \approx 6.77 \times 10^{-4} \text{ rad} = 0.677 \text{ mrad.} \quad (16.166)$$

The Rayleigh range is

$$z_R = \frac{\pi(5.0 \times 10^{-4})^2}{1.064 \times 10^{-6}} \approx 0.738 \text{ m.} \quad (16.167)$$

A beam focused to an even smaller waist would have a shorter Rayleigh range and larger divergence.

Caution

The Gaussian-beam formulas assume the fundamental paraxial mode and a consistent definition of beam radius. Real beams are often characterized by a beam-quality factor $M^2 \geq 1$, which increases the divergence above the ideal value. The simple equation above is the diffraction-limited case $M^2 = 1$.

16.20.13 Wavefront aberrations and the Strehl ratio

A diffraction-limited pupil is not merely an aperture of the correct size. The phase across that pupil must also be sufficiently close to the ideal wavefront. Represent a pupil as

$$P(x, y) = A(x, y)e^{i\phi(x, y)}, \quad (16.168)$$

where A describes amplitude transmission and ϕ describes phase error relative to the ideal reference wavefront. Because the focal amplitude is the Fourier transform of this *complex* pupil, phase errors redistribute energy in the point-spread function even if $|P|^2$ is unchanged across the aperture.

One common quality metric is the Strehl ratio

$$S = \frac{I_{\text{peak,aberrated}}}{I_{\text{peak,ideal}}}, \quad (16.169)$$

with equal total transmitted power and otherwise matched conditions. Measure wavefront error as optical path difference relative to the ideal wavefront, remove irrelevant piston and compare centered images after tilt correction. For small pupil phase variance $\sigma_\phi^2 = (2\pi\sigma_W/\lambda_0)^2$, expansion of $|\langle e^{i\phi} \rangle|^2$ gives $S \approx 1 - \sigma_\phi^2$. Its commonly used exponential form, the extended Maréchal approximation, is

$$S \approx \exp \left[- \left(\frac{2\pi\sigma_W}{\lambda_0} \right)^2 \right]. \quad (16.170)$$

Here λ_0 is the vacuum wavelength, while σ_W is an optical path difference. This high-Strehl approximation is not a general predictor for large deterministic aberrations. It shows why tolerances

become more demanding at shorter wavelength: a fixed physical surface or wavefront error represents a larger phase error when divided by a smaller λ .

Example 16.30: Strehl estimate for an RMS wavefront error

An optical system has RMS wavefront error $\sigma_W = 40$ nm at $\lambda_0 = 550$ nm. Estimate its Strehl ratio using Eq. (16.170).

Solution.

$$\frac{2\pi\sigma_W}{\lambda_0} = 2\pi \frac{40}{550} \approx 0.457. \quad (16.171)$$

Thus

$$S \approx e^{-0.457^2} \approx 0.812. \quad (16.172)$$

The peak intensity is about 81% of the ideal diffraction-limited value under the assumptions of the approximation. Energy is not destroyed; it is redistributed from the central peak into surrounding structure.

16.20.14 Fresnel diffraction, Fresnel zones and zone plates

Fraunhofer diffraction is often experimentally convenient because a lens can place the far-field angular distribution in its focal plane. Yet many systems operate in the Fresnel regime, where the aperture-dependent quadratic phase in Eq. (16.129) cannot be neglected.

A useful geometric construction divides a wavefront into Fresnel half-period zones relative to an observation point. Boundaries are chosen so that the path length increases by approximately $\lambda/2$ from one zone boundary to the next. Contributions from neighboring zones therefore tend to arrive nearly opposite in phase. Partial cancellation between zones explains why opening or blocking selected radial regions can strongly modify the on-axis field.

A *Fresnel zone plate* deliberately alternates transparent and opaque annular zones. It can focus waves even though it has no conventional refractive curvature. The focusing mechanism is diffraction: the radii are chosen so that transmitted zones contribute approximately in phase at a selected focal point. In the paraxial limit the zone radii approximately satisfy

$$r_n^2 \approx n\lambda f, \quad (16.173)$$

for zone number n and design focal length f under plane-wave illumination. To see the approximation, set $\sqrt{f^2 + r_n^2} - f = n\lambda/2$. Squaring gives $r_n^2 = n\lambda f + (n\lambda/2)^2$; the second term is negligible when $n\lambda \ll f$. Alternating open and opaque zones then remove contributions that would tend to cancel the transmitted ones.

This device is conceptually important for technical physics because it demonstrates that focusing is fundamentally about controlling phase, not necessarily about glass lenses. Zone plates are useful particularly at wavelengths for which ordinary refractive optics are difficult, including X-ray regimes where absorption and refractive-index constraints make conventional lenses challenging.

Example 16.31: First Fresnel-zone radius

Estimate the first-zone radius for a zone plate designed for $\lambda = 1.0$ nm and $f = 0.25$ m.

Solution.

$$r_1 \approx \sqrt{\lambda f} = \sqrt{(1.0 \times 10^{-9})(0.25)} \approx 1.58 \times 10^{-5} \text{ m} = 15.8 \mu\text{m}. \quad (16.174)$$

The small feature scale illustrates why short-wavelength diffractive optics demand microfabrication or nanofabrication techniques.

16.20.15 Fourier-transform rules as optical design rules

Once the far field is recognized as a Fourier transform, standard transform properties become physical design rules.

Scaling. If an aperture is widened by a factor s , its transform contracts by approximately the same factor. This is the mathematical statement behind “narrow slit, broad diffraction” and extends immediately to two dimensions.

Translation. If

$$A(x) \rightarrow A(x - x_0), \quad (16.175)$$

then its Fourier transform acquires a phase factor

$$\tilde{A}(f_x) \rightarrow e^{-i2\pi f_x x_0} \tilde{A}(f_x). \quad (16.176)$$

The intensity $|\tilde{A}|^2$ is unchanged. Therefore translating an isolated aperture sideways shifts its real-space position but does not alter the ideal far-field intensity distribution; it changes only the phase reference.

Linear phase. Multiplying the aperture by $e^{i2\pi f_0 x}$ shifts the Fourier transform by f_0 . A phase ramp across a pupil therefore steers a beam. This principle is used in phased arrays and programmable spatial light modulators.

Convolution and multiplication. The transform of a product is a convolution of transforms, while the transform of a convolution is a product. A finite array of slit centers is $\sum_{j=0}^{N-1} \delta(x - jd)$. Convoluting it with a single-slit field $t(x)$ creates the finite grating aperture:

$$A(x) = t(x) * \sum_{j=0}^{N-1} \delta(x - jd).$$

Thus the transform is the single-slit transform *multiplied* by the finite array factor. Equivalently, windowing an infinite array of centers broadens its discrete Fourier orders; it is not correct to generate finite-width slits by multiplying a delta comb by a slit profile. In Fourier space these ingredients become a set of discrete orders broadened by finite illumination and shaped by the single-element envelope.

This viewpoint unifies several formulas that otherwise look unrelated. The double-slit envelope, finite- N grating line width and apodization trade-offs all follow from simple transform composition.

Key message

In Fourier optics, geometry and signal processing use the same mathematics. Aperture width controls bandwidth, periodicity controls discrete spectral locations, finite extent controls line width, and smooth weighting controls side lobes. These are not analogies imposed afterward; they are consequences of the same Fourier transform.

16.20.16 The 4f optical system and spatial filtering

A particularly elegant engineering use of diffraction is the 4f optical system. Place an object one focal length in front of a lens. In the rear focal plane, different propagation angles are mapped to different positions, and under paraxial scalar conditions the field is proportional to the Fourier transform of the object field. For equal focal lengths f , a second lens placed f beyond that Fourier plane forms the inverted image a further f beyond the second lens; the object-to-image distance is $4f$.

The plane between the lenses is therefore a *spatial-frequency plane*. A small centered transmitting aperture in an otherwise opaque mask passes low spatial frequencies and produces a smoother image. Blocking the central region while transmitting larger angles can emphasize edges and rapid spatial variations. A phase mask can modify selected spatial frequencies without simply absorbing them.

This is optical signal processing performed at the speed of propagation: the lens geometry physically organizes the angular spectrum so that a mask can act directly on spatial-frequency components. Modern digital image processing often performs analogous operations numerically after detection, but optical filtering can act before the detector and can be useful when bandwidth or light-level constraints favor analog preprocessing.

Guided checks

1. In a Fourier plane, where are the lowest spatial frequencies located relative to the optical axis?
2. Why does a small centered aperture in the Fourier plane blur fine image detail?
3. What qualitative image feature is emphasized if low spatial frequencies are strongly attenuated while higher frequencies remain?
4. Why must a phase mask be treated as part of the complex field rather than as an intensity-only transmission filter?

16.20.17 Numerical Fraunhofer diffraction with the FFT

The fast Fourier transform (FFT) provides a direct computational route from an aperture field to its Fraunhofer pattern. A one-dimensional workflow is:

1. define a sampled aperture coordinate x_j with spacing Δx ;
2. define a complex aperture field A_j , including amplitude and phase;
3. compute the discrete Fourier transform $\tilde{A}_k$ using an FFT;
4. shift the zero spatial frequency to the center of the array for plotting;
5. map discrete spatial frequency f_x to angle through

$$\sin \theta = \lambda f_x; \quad (16.177)$$

6. plot $|\tilde{A}|^2$, with proper normalization if absolute intensity is required.

The discrete calculation introduces sampling constraints of its own. The aperture-plane sampling interval determines the maximum spatial frequency that can be represented, while the total computational window determines frequency resolution. Zero-padding can produce a more finely sampled plotted Fourier pattern but does not create new physical resolution; it interpolates the discrete transform.

A simple verification test is to simulate a rectangular slit and compare the numerical result with the analytic sinc-squared pattern. Agreement of zero positions is a strong check that coordinate mapping and normalization are consistent.

Method

A trustworthy numerical diffraction calculation should pass at least three checks: (1) symmetry when the aperture is symmetric, (2) correct scaling of pattern width with λ/a , and (3) convergence when the aperture-plane grid and computational window are refined. A visually smooth plot is not by itself evidence of numerical correctness.

16.20.18 FFT sampling: connecting numerical pixels to physical angles

A numerical FFT does not automatically know the physical wavelength, aperture size or propagation distance. Those quantities enter through the coordinate mapping. Suppose an aperture field is sampled at N equally spaced points with spacing Δx . The total computational aperture window is

$$L_x = N\Delta x. \quad (16.178)$$

The FFT samples spatial frequency in increments

$$\Delta f_x = \frac{1}{L_x}, \quad (16.179)$$

with a Nyquist spatial frequency of magnitude approximately

$$f_{x,\max} \approx \frac{1}{2\Delta x}. \quad (16.180)$$

In Fraunhofer diffraction,

$$\sin \theta_x = \lambda f_x, \quad (16.181)$$

so a one-dimensional FFT bin maps to a propagating angle only when $|\lambda f_x| \leq 1$. In two dimensions the condition is $\lambda^2(f_x^2 + f_y^2) \leq 1$, with these products interpreted as direction cosines. Larger transverse frequencies are evanescent in a homogeneous medium and must not be sent blindly to an inverse sine or interpreted as far-field rays.

Two distinct numerical design requirements follow.

1. Δx must be fine enough to represent the aperture transmission or phase without aliasing.
2. The total window L_x must be large enough to give sufficiently fine angular sampling $\Delta f_x = 1/L_x$ in the far-field pattern.

Zero padding increases the computational window without adding nonzero aperture data and samples the same transform more finely. If an enlarged window contains only additional known zeros, it has the same effect. Refining Δx improves the representation of edges and phase; extending a window matters physically only when previously truncated nonzero field is added. These are distinct convergence checks.

Example 16.32: Angular spacing of FFT samples

An aperture is represented on a one-dimensional grid with $N = 4096$ samples and $\Delta x = 2.0 \mu\text{m}$. For $\lambda = 633 \text{ nm}$, estimate the small-angle separation between neighboring Fraunhofer FFT samples.

Solution. The computational width is

$$L_x = N\Delta x = (4096)(2.0 \mu\text{m}) = 8.192 \text{ mm}. \quad (16.182)$$

Thus

$$\Delta f_x = \frac{1}{L_x} \approx 122.1 \text{ m}^{-1}. \quad (16.183)$$

For small angles, Eq. (16.181) gives

$$\Delta \theta \approx \lambda \Delta f_x = (633 \times 10^{-9})(122.1) \approx 7.73 \times 10^{-5} \text{ rad}. \quad (16.184)$$

This number is a sampling interval of the computed far field, not a diffraction-limited resolution by itself.

16.20.19 Advanced mini-project: numerical aperture-to-image pipeline

A useful computational project is to build a complete scalar imaging pipeline:

1. create a two-dimensional pupil array representing a circular aperture;
2. optionally add a central obstruction, support vanes or a phase aberration;
3. compute the FFT of the complex pupil field to obtain a focal-plane amplitude;
4. square the magnitude to obtain a PSF;
5. normalize the PSF so its discrete sum is unity;
6. create a synthetic object containing point pairs and bar patterns of different separation;
7. convolve the object with the PSF;
8. sample the result onto detector pixels;
9. add Poisson or Gaussian noise at controlled level;
10. test how apparent resolution changes with aperture, wavelength, aberration and signal-to-noise ratio.

The project teaches several transferable skills: FFT usage, normalization, convolution, sampling, noise modelling, parameter scans and the distinction between an optical limit and a detection limit. It also creates a direct bridge between optics and computer-science topics such as image filtering and inverse problems.

Method

When comparing simulated systems, hold all but one physical parameter fixed. If aperture diameter, wavelength, detector pitch and noise level are changed simultaneously, it becomes difficult to identify which mechanism caused the observed change. Controlled parameter scans are the computational analogue of controlled laboratory experiments.

16.20.20 Spectrometer design: dispersion, slit width and detector sampling

A real grating spectrometer is a chain of subsystems. An entrance slit defines a spatial source width. A collimator sends approximately parallel light to the grating. The grating disperses wavelength into angle. A camera optic maps angle to detector position. The detector samples the resulting line-spread function.

Suppose the reciprocal linear dispersion at the detector is defined as

$$\mathcal{R}_L \equiv \frac{d\lambda}{dx}. \quad (16.185)$$

A detector pixel of width p then corresponds locally to a wavelength interval approximately

$$\Delta\lambda_{\text{pix}} \approx \mathcal{R}_L p. \quad (16.186)$$

If the optical line-spread function has full width at half maximum spanning much less than one pixel, the detector undersamples it. If it spans many pixels, sampling is excellent but detector area may be used inefficiently. A common engineering convention allocates at least two or three detector samples to a stated line-width measure. This is a useful design heuristic; the Nyquist theorem itself concerns the highest frequency of the sampled signal, not an arbitrary first-zero or FWHM width. Precise sampling requirements depend on the line-spread function and the chosen recovery task.

The entrance slit creates a second trade-off. A narrow slit improves spectral line definition but reduces light throughput. In photon-limited measurements, too narrow a slit can decrease signal-to-noise enough that theoretical resolution becomes practically useless.

Example 16.33: Detector sampling of a spectral resolution element

A spectrograph near 600 nm has resolving power $R = 30,000$. Its detector has $15 \mu\text{m}$ pixels. If the design aims for three pixels per resolution element, what reciprocal linear dispersion $d\lambda/dx$ is required near 600 nm?

Solution. The wavelength width of one resolution element is

$$\Delta\lambda = \frac{\lambda}{R} = \frac{600 \text{ nm}}{30,000} = 0.020 \text{ nm}. \quad (16.187)$$

Three pixels span

$$\Delta x = 3(15 \mu\text{m}) = 45 \mu\text{m}. \quad (16.188)$$

Therefore

$$\frac{d\lambda}{dx} \approx \frac{0.020 \text{ nm}}{45 \mu\text{m}} \approx 4.44 \times 10^{-4} \text{ nm } \mu\text{m}^{-1}. \quad (16.189)$$

Equivalently, the detector should provide about 2.25 mm per nanometer in this local linearization.

16.20.21 Advanced synthesis example: designing a compact grating spectrometer

Example 16.34: First-pass design of a compact spectrometer

In an illustrative design, a compact spectrometer is intended to operate near $\lambda = 650 \text{ nm}$ with resolving power at least $R = 20,000$. A reflective grating with 1200 lines/mm is available. The design uses first order and illuminates a width $W = 20 \text{ mm}$. The camera axis is aligned with the central first-order ray under normal grating incidence. The camera focal length is $f = 150 \text{ mm}$ and the detector pixel pitch is $10 \mu\text{m}$. Assess the ideal groove-count resolving power, the diffraction angle, the local wavelength interval per pixel in the paraxial detector mapping, and whether two pixels per ideal resolution element are sufficient.

Solution. The groove spacing is

$$d = \frac{1 \text{ mm}}{1200} = 8.333 \times 10^{-7} \text{ m}. \quad (16.190)$$

The illuminated groove count is

$$N = 1200 \text{ mm}^{-1} \times 20 \text{ mm} = 24,000. \quad (16.191)$$

In first order, the ideal resolving power is

$$R_{\text{ideal}} = |m|N = 24,000, \quad (16.192)$$

which exceeds the target 20,000 but leaves only modest margin for slit and aberration broadening.

The diffraction angle satisfies

$$\sin \theta = \frac{650 \times 10^{-9}}{8.333 \times 10^{-7}} = 0.780, \quad (16.193)$$

so

$$\theta \approx 51.3^\circ, \quad \cos \theta \approx 0.626. \quad (16.194)$$

The angular dispersion is

$$\frac{d\theta}{d\lambda} = \frac{1}{d \cos \theta} \approx 1.92 \times 10^6 \text{ rad m}^{-1}. \quad (16.195)$$

Using $x = f \tan(\theta - \theta_0)$ about this camera axis, the local derivative is $dx = f d\theta$, even though $\theta_0 \approx 51.3^\circ$:

$$\frac{dx}{d\lambda} \approx (0.150)(1.92 \times 10^6) \approx 2.88 \times 10^5 \text{ m m}^{-1}. \quad (16.196)$$

Therefore one $10 \mu\text{m}$ pixel corresponds to

$$\Delta\lambda_{\text{pix}} \approx \frac{10 \times 10^{-6}}{2.88 \times 10^5} \approx 3.48 \times 10^{-11} \text{ m} = 0.0348 \text{ nm}. \quad (16.197)$$

The ideal grating resolution element is

$$\Delta\lambda_{\text{ideal}} = \frac{650 \text{ nm}}{24,000} \approx 0.0271 \text{ nm}. \quad (16.198)$$

Thus the ideal line width is less than one pixel in this simplified mapping. Two pixels span about 0.0695 nm , which corresponds to an illustrative two-sample resolving-power scale $R \sim 9300$. This is a sampling design estimate, not a theorem that forbids more precise model-based line centroiding. The detector/focal-length combination therefore undersamples the ideal grating resolution. A longer camera focal length, smaller pixels or a different optical magnification would be required if the full groove-count resolving power is to be recorded.

This example demonstrates why $|m|N$ alone is not an instrument design. Grating diffraction, camera optics and detector sampling must be matched as one system.

16.20.22 Reciprocal lattices and the deeper meaning of Bragg diffraction

Bragg's law is geometrically intuitive, but a more general description of crystal diffraction uses reciprocal space. A periodic crystal potential can be expanded in spatial Fourier components indexed by reciprocal-lattice vectors $\mathbf{G}$. Elastic scattering changes the wave vector from $\mathbf{k}_i$ to $\mathbf{k}_f$ while preserving its magnitude:

$$|\mathbf{k}_i| = |\mathbf{k}_f| = \frac{2\pi}{\lambda}. \quad (16.199)$$

Constructive crystal diffraction occurs when the scattering vector

$$\mathbf{q} = \mathbf{k}_f - \mathbf{k}_i \quad (16.200)$$

coincides with a reciprocal-lattice vector:

$$\boxed{\mathbf{k}_f - \mathbf{k}_i = \mathbf{G}.} \quad (16.201)$$

This is a vector form of the Laue diffraction condition.

For a family of planes with spacing d_{hkl} , the corresponding reciprocal-lattice vector magnitude is

$$G_{hkl} = \frac{2\pi}{d_{hkl}}. \quad (16.202)$$

The wave-vector triangle gives $q = 2k \sin \theta_B$. Setting $q = G_{hkl}$ yields $2d_{hkl} \sin \theta_B = \lambda$. A higher order n of the same plane spacing uses nG_{hkl} and is conventionally indexed as the corresponding higher-index reflection. For a cubic conventional cell of side a_c ,

$$d_{hkl} = \frac{a_c}{\sqrt{h^2 + k^2 + l^2}},$$

before applying the structure selection rules. This explicitly reproduces Bragg's law. Thus Bragg and Laue descriptions are not competing theories; they are equivalent viewpoints adapted to different geometrical and mathematical preferences.

The reciprocal-space picture is powerful because diffraction is fundamentally a measurement of spatial frequencies in a periodic structure. It connects naturally to the Fourier-transform viewpoint developed earlier in this advanced section.

Theoretical point 16.2: Unification through Fourier structure

A slit, a grating and a crystal differ enormously in physical scale, but each converts spatial structure into angular structure through wave interference. The slit transforms a finite aperture, the grating transforms a one-dimensional periodic array, and the crystal transforms a three-dimensional periodic density. Fourier analysis provides a common mathematical language for all three.

16.20.23 Crystal structure factors and systematic absences

Bragg's law determines geometrically allowed plane spacings, but it does not determine the intensity of every reflection. The waves scattered from atoms within one unit cell can interfere with each other. For a crystal basis with atoms at fractional coordinates (x_j, y_j, z_j) , a kinematic structure factor is

$$F_{hkl} = \sum_j f_j \exp [2\pi i(hx_j + ky_j + lz_j)], \quad (16.203)$$

where f_j is the scattering factor of atom j . The ideal integrated intensity contains a factor proportional to $|F_{hkl}|^2$, together with geometric, polarization, multiplicity, thermal and instrumental contributions.

Using conventional cubic-cell indices for a body-centered cubic crystal with identical scatterers at $(0, 0, 0)$ and $(1/2, 1/2, 1/2)$,

$$F_{hkl} = f [1 + e^{i\pi(h+k+l)}]. \quad (16.204)$$

If $h + k + l$ is odd, the two terms cancel and the reflection is systematically absent; if $h + k + l$ is even, they add.

For a face-centered cubic crystal, the conventional cell has sites $(0, 0, 0)$, $(0, 1/2, 1/2)$, $(1/2, 0, 1/2)$ and $(1/2, 1/2, 0)$. Summing their phases gives $F = f[1 + (-1)^{k+l} + (-1)^{h+l} + (-1)^{h+k}]$, nonzero only when h, k, l are all even or all odd. These conventional-cell restrictions are consistent with the reciprocal lattice of the corresponding Bravais lattice; they do not remove allowed reciprocal-lattice vectors arbitrarily. These systematic absences contain structural information that is invisible to Bragg's scalar spacing relation alone.

Example 16.35: Why a Bragg-allowed reflection can vanish

Consider a BCC crystal and the (100) plane family. Geometrically, a spacing d_{100} exists and Bragg's law can be satisfied at some angle. Yet $h + k + l = 1$ is odd, so

$$F_{100} = f(1 - 1) = 0. \quad (16.205)$$

In the ideal kinematic model the reflection is absent. This is not a failure of Bragg's law; Bragg's law identifies geometric phase matching between lattice planes, while the structure factor accounts for interference among the basis atoms inside each unit cell.

16.20.24 Powder diffraction and line broadening

In a powder, many small crystallites are oriented approximately randomly. For a given plane spacing, some crystallites will satisfy the Bragg condition at a particular angle, producing a cone of diffracted intensity. A detector scan in 2θ records peaks characteristic of the set of lattice spacings present in the material.

Peak positions are determined mainly by lattice geometry. Peak widths contain additional information. Finite coherent domain size can broaden diffraction peaks. A commonly used estimate is the Scherrer relation

$$L_c \approx \frac{K\lambda}{\beta_{2\theta} \cos \theta_B}, \quad (16.206)$$

where L_c is a characteristic coherent-domain size, K is a shape factor of order unity, and $\beta_{2\theta}$ is the line broadening in radians after appropriate correction for instrumental broadening.

The width convention (for example FWHM in 2θ) and peak profile determine the appropriate shape factor K . For Gaussian instrument and sample profiles, their variances add, so $\beta_{\text{sample}} \approx \sqrt{\beta_{\text{obs}}^2 - \beta_{\text{inst}}^2}$; Lorentzian widths subtract linearly instead. Mixed profiles require a suitable convolution fit. The Scherrer equation is useful but easy to overinterpret. Strain, defects, instrument response and peak-shape conventions also contribute to line width. Therefore it should be presented as a model-dependent estimate, not a universal nanoparticle-size formula.

Caution

Never insert a raw peak width in degrees directly into the Scherrer equation. The width must be expressed in radians and should represent sample broadening after instrumental effects are considered. The inferred L_c is a coherent diffraction length, which need not equal the geometric particle diameter.

16.20.25 Electron diffraction and matter waves

Diffraction is not unique to electromagnetic waves. Quantum particles have de Broglie wavelength

$$\lambda = \frac{h}{p}. \quad (16.207)$$

Electrons accelerated through a potential difference can have wavelengths comparable to atomic spacings and therefore diffract from crystals.

For a nonrelativistic electron accelerated from rest through voltage V ,

$$eV = \frac{p^2}{2m_e}, \quad (16.208)$$

so

$$\lambda = \frac{h}{\sqrt{2m_e eV}}. \quad (16.209)$$

At sufficiently high voltage, relativistic corrections are required.

Example 16.36: Electron wavelength for diffraction

Estimate the nonrelativistic de Broglie wavelength of electrons accelerated through 150 V.

Solution. Using Eq. (16.209),

$$\lambda = \frac{6.626 \times 10^{-34}}{\sqrt{2(9.11 \times 10^{-31})(1.602 \times 10^{-19})(150)}} \approx 1.00 \times 10^{-10} \text{ m} = 0.100 \text{ nm}. \quad (16.210)$$

This wavelength is comparable to interatomic spacings, so crystal diffraction is expected. The result anticipates the quantum-physics chapters in which matter-wave behavior is developed more fully.

16.20.26 Diffraction and information: what a finite pupil removes

The Fourier viewpoint suggests an information-theoretic interpretation. A finite pupil acts as a spatial-frequency filter. Object components beyond the transmitted bandwidth are strongly suppressed or absent from the measured field. Noise and finite photon number further limit how reliably the remaining components can be estimated.

This perspective helps clarify what computational imaging can and cannot do. Deconvolution can partially compensate a known blur when the relevant spatial frequencies are attenuated but still measurable. If the transfer function is exactly zero over a range of frequencies, reconstructing arbitrary unknown content in that range is impossible without additional prior information or additional measurements. Modern super-resolution techniques circumvent classical limits by changing the measurement model—for example through nonlinear fluorescence response, structured illumination, near-field coupling or strong prior constraints—not by simply magnifying a diffraction-limited image.

Pedagogical boundary

The phrase “diffraction limit” depends on the imaging modality, resolution criterion, noise level and prior information. This chapter uses the conventional scalar far-field limits for ordinary linear imaging. Advanced super-resolution methods deliberately modify one or more of those assumptions and therefore require separate analysis rather than a claim that diffraction has been “violated.”

16.20.27 Advanced section summary**Summary**

The advanced viewpoint unifies diffraction through Fourier structure. The far field is determined by the Fourier transform of the complex pupil or scattering distribution. A rectangular aperture gives a sinc pattern; a circular aperture gives a Bessel/Airy pattern; a finite periodic grating gives narrow array-factor peaks; a crystal gives reciprocal-lattice diffraction. Imaging systems are characterized by point-spread and transfer functions, while numerical aperture limits the spatial frequencies accepted by a microscope. Gaussian beams show that even smoothly confined waves diffract. Apodization trades main-lobe width against side-lobe suppression. Real instruments must match optical resolution to slit width, detector sampling, noise and calibration. These ideas connect classical wave optics directly to signal processing, spectroscopy, imaging and materials characterization.

16.21 Original Practice Problems

The problems in this section are written to support both independent study and auditorium-class work. Unless a problem states otherwise, assume monochromatic light in air or vacuum, treat the stated apertures as ideal, and use the Fraunhofer approximation whenever an angular diffraction formula is invoked. Numerical answers should include units and a physically sensible number of significant figures. When a small-angle approximation is used, state it explicitly and check that it is reasonable.

Method

A reliable workflow for nearly every diffraction problem is:

1. identify the physical element—single slit, circular aperture, finite double slit, grating or crystal planes;
2. write the corresponding angular condition before inserting numbers;
3. decide whether the exact trigonometric relation or a small-angle approximation is appropriate;
4. distinguish angular quantities from positions on a screen or focal plane;
5. perform a dimensional and limiting check;
6. interpret the result in terms of spreading, resolution or spectral separation.

Conceptual reasoning

1. A laser beam passes through two adjustable slits, one twice as wide as the other. Without calculation, compare the angular widths of their central diffraction maxima. Explain the physical reason for the comparison rather than merely quoting a formula.
2. A student says, “The central diffraction maximum is bright because all points in the slit are exactly in phase for every direction close to the axis.” Identify the error and replace the statement with a correct one.
3. Explain why the condition $a \sin \theta = m\lambda$ locates single-slit *minima*, whereas $d \sin \theta = m\lambda$ locates principal maxima of an ideal diffraction grating.
4. A camera lens is stopped down from a large aperture to a very small aperture. Geometrical aberrations may decrease, yet the final image can become less sharp. Explain why these two tendencies can coexist.
5. Rayleigh’s criterion is often called a resolution criterion rather than a fundamental binary law. What does this distinction mean experimentally?
6. A double slit has finite slit width. Why are some interference fringes absent even though the two-slit interference condition would otherwise predict a maximum there?
7. Why do the principal maxima of an N -slit grating become narrower as N increases while their allowed central angles remain governed by the same grating equation?
8. Explain the difference between *angular dispersion* and *resolving power*. Can an instrument have large dispersion but inadequate resolving power? Give a qualitative example.
9. Two wavelengths produce overlapping spectral orders in a grating. Explain what order overlap means and why filters or restricted wavelength ranges may be required in practical spectrometers.
10. A crystalline sample produces sharp diffraction peaks, while an amorphous sample produces broad features. Explain qualitatively what this reveals about spatial order in the two materials.

11. Explain why a phase-only change across an aperture can alter a far-field diffraction pattern even if the transmitted intensity immediately after the aperture is unchanged.
12. A detector has extremely small pixels but is placed behind a poor optical system. Explain why reducing pixel pitch further cannot by itself restore spatial frequencies that the optical system does not transmit.

Analytical problems

13. Monochromatic light of wavelength 632.8 nm illuminates a slit of width 0.120 mm. Determine the angular positions of the first three minima using the exact sine relation. Then compare them with the small-angle estimates.
14. Derive the full width of the central maximum on a screen a distance L from a single slit without initially using the small-angle approximation. Show explicitly how the usual result $w \approx 2L\lambda/a$ follows.
15. For a single-slit pattern, show that the first secondary maximum is determined by the nonzero solution of $\tan \beta = \beta$. Explain why the points $\beta = m\pi$ are not maxima. Estimate the location of the first secondary maximum using a numerical root or a known approximate value.
16. A slit produces a central maximum of width 18.0 mm on a screen 2.40 m away when illuminated with 532 nm light. Determine the slit width using the small-angle approximation and then check the corresponding first-minimum angle with the exact formula.
17. A thin wire of unknown diameter is placed in a laser beam. The distance between the first minima on opposite sides of the central maximum is 24.0 mm on a screen 1.80 m away. For $\lambda = 650$ nm, estimate the wire diameter using Babinet's principle.
18. Two equal, mutually incoherent point sources are observed through a circular aperture of diameter 25.0 cm at $\lambda = 600$ nm. Find their Rayleigh angular separation. If the sources are 12.0 km away, what transverse separation corresponds to that angle?
19. A microscope operates at vacuum wavelength $\lambda_0 = 500$ nm with an oil-immersion objective of $n = 1.515$ and half-angle $\alpha = 60^\circ$. Compute the numerical aperture and a Rayleigh-type lateral resolution estimate.
20. Two slits each have width $a = 0.080$ mm and center-to-center spacing $d = 0.40$ mm. Determine how many nominal constructive orders of the interference factor lie within the central single-slit diffraction envelope. Identify any missing orders.
21. A grating has 600 lines per millimetre. For normally incident 500 nm light, calculate the angles of all physically allowed principal maxima and state the highest possible diffraction order.
22. Differentiate the grating equation to derive the angular dispersion

$$\frac{d\theta}{d\lambda} = \frac{m}{d \cos \theta}.$$

Then derive the focal-plane linear dispersion for a spectrometer using a lens of focal length f and coordinate $x = f \tan \theta$.

23. A grating of 1200 lines/mm is used in first order near 600 nm. Calculate the angular dispersion in radians per nanometre. With a 0.50 m focal-length camera lens, estimate the local linear dispersion in mm/nm using the exact derivative of $x = f \tan \theta$.
24. A grating has $N = 8.0 \times 10^3$ illuminated grooves and is used in second order. Estimate its ideal resolving power and the smallest wavelength difference resolvable near 500 nm.

25. A grating spectrometer is required to resolve wavelengths 589.00 nm and 589.59 nm in first order. What minimum number of illuminated grooves is required by the ideal Rayleigh result? If the groove density is 600 lines/mm, what illuminated width does this correspond to?
26. Derive the free spectral range estimate $\Delta\lambda_{\text{FSR}} \approx \lambda/m$ for neighboring orders by equating approximately the angle of order m at wavelength $\lambda + \Delta\lambda$ to order $m + 1$ at wavelength λ .
27. X rays of wavelength 0.154 nm produce a first-order Bragg peak at $2\theta = 44.0^\circ$ in a conventional diffractometer. Determine the plane spacing d . Be explicit about the relation between the reported scan angle 2θ and the Bragg angle θ .
28. Starting from the scattering-vector magnitude $q = (4\pi/\lambda)\sin\theta$, show that the Bragg condition implies $q = 2\pi m/d$. Interpret this relation in reciprocal-space language.
29. An X-ray diffraction peak at $2\theta = 40.0^\circ$ has corrected FWHM 0.20° . Estimate a characteristic coherent domain size using $K = 0.90$ and $\lambda = 0.154$ nm. Clearly convert the angular width to radians and state the limitations of the estimate.

Computation and data

30. Numerically compare the exact screen coordinate $y = L \tan \theta$ with the approximation $y \approx L\theta$ for the first three minima of a slit with $a = 20\lambda$. At what order does the relative error first exceed 1%?
31. Implement the N -slit array factor

$$\left[\frac{\sin(N\gamma)}{\sin \gamma} \right]^2, \quad \gamma = \frac{\pi d \sin \theta}{\lambda},$$
 for $N = 2, 5, 20, 100$. Normalize each curve to its principal maximum and measure numerically how the main-peak width scales with N .
32. Create a two-dimensional binary circular pupil on a square grid and compute its FFT. Radially average the resulting intensity and compare the first dark-ring radius with the Airy prediction. Discuss at least three numerical effects that can shift or blur the apparent zero.
33. Construct a simple imaging simulation in which two equal, mutually incoherent point sources are convolved with a diffraction-limited circular-aperture PSF. Vary their separation through 0.5, 1.0, and 1.5 times the Rayleigh angle. Plot the one-dimensional intensity cuts and discuss how the visual distinction changes continuously rather than discontinuously.
34. Generate a synthetic powder-diffraction pattern containing several Gaussian peaks. Add random noise, fit the peak centers and widths, and propagate the fitted angular uncertainty into an uncertainty in lattice spacing through Bragg's law.
35. Simulate a one-dimensional rectangular aperture with both uniform and Hann-like amplitude transmission. Compare the full width of the central lobe and the highest side-lobe level in the two cases. Explain the result using the language of apodization.

Engineering applications

36. An idealized small-format camera has a physical focal length of 6.0 mm and operates at f/1.8 and f/11. For 550 nm light, compare the Airy-disk diameters at the two f-numbers. If the pixel pitch is $1.4 \mu\text{m}$, discuss the relation between diffraction spot size and detector sampling.
37. A ground-based telescope of diameter 0.80 m observes at 550 nm. Calculate its diffraction-limited angular resolution. Typical atmospheric seeing is 1.2 arcsec during an observation. Which limit dominates, and by approximately what factor?

38. A microscope is required to resolve 250 nm features at vacuum wavelength $\lambda_0 = 550$ nm. What numerical aperture would be required according to $\delta = 0.61\lambda_0/\text{NA}$? Is this plausible in air? How does immersion change the design space?
39. A compact spectrometer uses a 1200 lines/mm grating in first order, a 100 mm focal-length camera lens, and a detector with $10\text{ }\mu\text{m}$ pixels. For the explicitly assumed projection $x = f \tan \theta$, near 500 nm determine the wavelength interval per pixel. Also state how the result changes if the camera axis is instead centered on that diffracted ray. Compare this sampling interval with the grating's ideal $\Delta\lambda$ if 15 mm of grating width is illuminated.
40. A CD has approximately 625 tracks/mm. Treat it as a reflection grating and estimate the first-order diffraction angle for 650 nm light at normal incidence. Explain qualitatively why a white-light reflection from a CD separates into colors.
41. A materials laboratory uses Cu K α radiation with $\lambda \approx 0.154$ nm. A new sample shows a strong peak at $2\theta = 36.0^\circ$. Estimate the associated plane spacing. If the peak shifts to larger 2θ after processing, does the inferred spacing increase or decrease? Explain using Bragg's law.

Advanced challenges

42. For a grating with finite illuminated width W , use $R = |m|N$ and $N = W/d$ together with $d \sin \theta = m\lambda$ to show how the ideal resolving power can be expressed in terms of W , λ and θ . Discuss what the result says about grazing diffraction angles.
43. A grating is illuminated at oblique incidence angle α rather than normal incidence. Using a consistent sign convention, derive a generalized grating equation of the form $d(\sin \theta_m - \sin \alpha) = m\lambda$ or its equivalent. Check that the normal-incidence result is recovered.
44. An X-ray peak position is $2\theta = (60.00 \pm 0.04)^\circ$ for $\lambda = 0.15406$ nm. Assuming wavelength uncertainty is negligible, propagate the angular uncertainty into the spacing d for first-order diffraction. Report both d and its uncertainty.
45. Starting from the Fraunhofer integral for a one-dimensional aperture $A(x)$, show that shifting the aperture by x_0 changes the far-field amplitude only by a phase factor. Explain why the intensity pattern is therefore unchanged by a rigid transverse translation of an isolated aperture.
46. Starting from the same Fourier viewpoint, show that multiplying a pupil field by a linear phase ramp shifts the far-field pattern. Relate this to beam steering by a phased array or spatial light modulator.
47. A two-dimensional aperture is rectangular with dimensions $a \times b$. Derive the separable far-field intensity and determine the angular widths between first minima along both transverse directions. Explain how this pattern could be used experimentally to infer both aperture dimensions.
48. An imaging system has a circular pupil and detector pixels that subtend 0.40 times the Rayleigh angle. Discuss whether the image is adequately sampled. Then repeat the discussion for pixels subtending 1.5 times the Rayleigh angle. Your answer should distinguish optical resolution, sampling, aliasing and signal-to-noise ratio.

16.22 Selected Solutions to Representative Problems

The solutions below are deliberately more explicit than a typical answer key. They illustrate setup, assumptions, units and interpretation. Problems not solved here are intended for independent work or tutorial discussion.

Problem 13: first three single-slit minima

For $a = 0.120 \text{ mm} = 1.20 \times 10^{-4} \text{ m}$ and $\lambda = 632.8 \text{ nm} = 6.328 \times 10^{-7} \text{ m}$,

$$\sin \theta_m = \frac{m\lambda}{a}. \quad (16.211)$$

The ratio λ/a is

$$\frac{\lambda}{a} = 5.273 \times 10^{-3}. \quad (16.212)$$

Thus

$$\theta_1 = \arcsin(5.273333 \times 10^{-3}) \approx 0.30214^\circ, \quad (16.213)$$

$$\theta_2 = \arcsin(1.054667 \times 10^{-2}) \approx 0.60429^\circ, \quad (16.214)$$

$$\theta_3 = \arcsin(1.582000 \times 10^{-2}) \approx 0.90646^\circ. \quad (16.215)$$

Using $\theta_m \approx m\lambda/a$ in radians gives 0.30214° , 0.60428° and 0.90642° , respectively (extra digits are shown to compare the approximations). The difference is tiny because every angle is below one degree.

Problem 15: condition for secondary maxima

Let

$$f(\beta) = \left(\frac{\sin \beta}{\beta} \right)^2. \quad (16.216)$$

Away from $\beta = 0$, extrema satisfy $df/d\beta = 0$. It is convenient first to differentiate the amplitude factor $g(\beta) = \sin \beta/\beta$:

$$g'(\beta) = \frac{\beta \cos \beta - \sin \beta}{\beta^2}. \quad (16.217)$$

Because $f = g^2$,

$$f' = 2gg'. \quad (16.218)$$

Thus extrema occur either when $g = 0$, which gives the diffraction minima $\beta = m\pi$, or when

$$\beta \cos \beta - \sin \beta = 0. \quad (16.219)$$

For nonzero $\cos \beta$ this becomes

$$\tan \beta = \beta. \quad (16.220)$$

The nonzero solutions between successive zeros locate the secondary maxima. The first solution beyond π is $\beta \approx 4.493$. The important conceptual point is that the minima $m\pi$ are also stationary points of the intensity, but they are zeros, not maxima. One must therefore classify the stationary points rather than treating every derivative zero identically.

Problem 16: determining slit width from central width

The central maximum extends between the first minima at $\pm y_1$, so $w = 2y_1$. Under the small-angle approximation,

$$w \approx \frac{2L\lambda}{a}, \quad (16.221)$$

which gives

$$a \approx \frac{2L\lambda}{w} = \frac{2(2.40)(532 \times 10^{-9})}{18.0 \times 10^{-3}} \approx 1.42 \times 10^{-4} \text{ m} = 0.142 \text{ mm}. \quad (16.222)$$

The exact first-minimum condition gives

$$\sin \theta_1 = \frac{\lambda}{a} \approx 3.75 \times 10^{-3}, \quad (16.223)$$

so $\theta_1 \approx 0.215^\circ$. Such a small angle fully justifies the approximation.

Problem 17: wire diameter from Babinet diffraction

The measured distance between the first minima on opposite sides is 24.0 mm, so

$$y_1 = 12.0 \text{ mm} = 1.20 \times 10^{-2} \text{ m}. \quad (16.224)$$

For a thin wire, Babinet's principle permits the same angular-minimum condition as for a slit of equal width. Under the small-angle approximation,

$$a \approx \frac{L\lambda}{y_1}. \quad (16.225)$$

Using $L = 1.80 \text{ m}$ and $\lambda = 650 \text{ nm}$,

$$a \approx \frac{(1.80)(650 \times 10^{-9})}{1.20 \times 10^{-2}} = 9.75 \times 10^{-5} \text{ m} = 97.5 \mu\text{m}. \quad (16.226)$$

The first-minimum angle is about $y_1/L = 6.67 \times 10^{-3} \text{ rad}$, so the small-angle approximation is excellent. In a real measurement, nonuniform wire diameter and uncertainty in the exact pattern center may dominate over trigonometric corrections.

Problem 18: angular and transverse telescope resolution

For $D = 0.250 \text{ m}$ and $\lambda = 600 \text{ nm}$,

$$\theta_R = 1.22 \frac{600 \times 10^{-9}}{0.250} = 2.93 \times 10^{-6} \text{ rad}. \quad (16.227)$$

At distance $L = 12.0 \text{ km} = 1.20 \times 10^4 \text{ m}$, the corresponding transverse separation is approximately

$$s = L\theta_R = (1.20 \times 10^4)(2.93 \times 10^{-6}) \approx 3.51 \times 10^{-2} \text{ m} = 3.51 \text{ cm}. \quad (16.228)$$

This is a diffraction-only scale for two ideal point-like sources with the Rayleigh convention. Contrast, turbulence, aberrations and detector sampling can make the practical distinguishable separation larger.

Problem 19: microscope numerical aperture

The numerical aperture is

$$\text{NA} = n \sin \alpha = (1.515) \sin 60^\circ \approx 1.31. \quad (16.229)$$

The Rayleigh-type lateral separation is

$$\delta \approx 0.61 \frac{\lambda_0}{\text{NA}} = 0.61 \frac{500 \text{ nm}}{1.312} \approx 232 \text{ nm}. \quad (16.230)$$

The value is smaller than the vacuum wavelength λ_0 because the microscope accepts a large cone of spatial frequencies in a medium whose refractive index exceeds unity. This does not violate wave optics; the relevant dimensionless acceptance is $n \sin \alpha$.

Problem 20: fringes inside a diffraction envelope

Interference maxima satisfy

$$d \sin \theta = n\lambda, \quad (16.231)$$

where n is an integer. The first single-slit minima satisfy

$$a \sin \theta = \pm \lambda. \quad (16.232)$$

Therefore interference maxima lie inside the central envelope when

$$\left| \frac{n\lambda}{d} \right| < \frac{\lambda}{a} \implies |n| < \frac{d}{a} = 5. \quad (16.233)$$

Hence the maxima $n = 0, \pm 1, \pm 2, \pm 3, \pm 4$ lie strictly inside the central envelope: nine nominal constructive orders in total. The orders $n = \pm 5$ coincide with the first diffraction minima and are missing. More generally, because $d/a = 5$, every fifth two-slit order coincides with a single-slit zero.

Problem 23: dispersion of a 1200 lines/mm grating

The spacing is

$$d = \frac{1}{1.20 \times 10^6 \text{ m}^{-1}} = 8.333 \times 10^{-7} \text{ m}. \quad (16.234)$$

At $\lambda = 600 \text{ nm}$ in first order,

$$\sin \theta = \frac{600 \times 10^{-9}}{8.333 \times 10^{-7}} = 0.720, \quad (16.235)$$

so $\theta \approx 46.05^\circ$ and $\cos \theta \approx 0.694$. The angular dispersion is

$$\frac{d\theta}{d\lambda} = \frac{1}{d \cos \theta} \approx 1.73 \times 10^6 \text{ rad/m} = 1.73 \times 10^{-3} \text{ rad/nm}. \quad (16.236)$$

For $x = f \tan \theta$,

$$\frac{dx}{d\lambda} = f \sec^2 \theta \frac{d\theta}{d\lambda} = \frac{fm}{d \cos^3 \theta}. \quad (16.237)$$

With $f = 0.50 \text{ m}$,

$$\frac{dx}{d\lambda} \approx 1.80 \text{ mm/nm}. \quad (16.238)$$

The sizeable difference between $f d\theta/d\lambda$ and the derivative of the specified projection model arises because θ is no longer very small.

Problem 24: ideal grating resolution

For $N = 8.0 \times 10^3$ and $m = 2$,

$$R = |m|N = 1.6 \times 10^4. \quad (16.239)$$

Near 500 nm ,

$$\Delta\lambda_{\min} \approx \frac{\lambda}{R} = \frac{500 \text{ nm}}{1.6 \times 10^4} \approx 3.13 \times 10^{-2} \text{ nm}. \quad (16.240)$$

This is an ideal grating-limited value. A real spectrometer can have poorer resolution because of entrance-slit width, aberrations, detector sampling and imperfect illumination.

Problem 25: resolving the sodium doublet

Using the mean wavelength $\lambda \approx 589.295 \text{ nm}$ and separation $\Delta\lambda = 0.59 \text{ nm}$,

$$R_{\text{req}} = \frac{589.3}{0.59} \approx 999. \quad (16.241)$$

In first order, $R = N$, so the leading Rayleigh estimate requires $N \geq \lceil 589.295/0.59 \rceil = 999$ illuminated grooves. About one thousand is an appropriate practical scale. With 600 lines/mm,

$$W = \frac{N}{600 \text{ mm}^{-1}} \approx 1.67 \text{ mm}. \quad (16.242)$$

The small illuminated width illustrates that this particular doublet is not an extreme resolving-power challenge for a laboratory grating; practical line broadening can nevertheless dominate if the entrance slit is too wide.

Problem 27: crystal spacing from a 2θ peak

The instrument reports $2\theta = 44.0^\circ$, so the Bragg angle is

$$\theta = 22.0^\circ. \quad (16.243)$$

For first order,

$$d = \frac{\lambda}{2 \sin \theta} = \frac{0.154 \text{ nm}}{2 \sin 22.0^\circ} \approx 0.206 \text{ nm}. \quad (16.244)$$

Using 44° directly in Bragg's law would give a substantially different and incorrect spacing for this conventional θ - 2θ geometry.

Problem 29: coherent domain size from peak width

The peak is at $2\theta = 40.0^\circ$, hence $\theta = 20.0^\circ$. The corrected FWHM is

$$\beta_{2\theta} = 0.20^\circ \frac{\pi}{180} = 3.49 \times 10^{-3} \text{ rad.} \quad (16.245)$$

Using the Scherrer estimate,

$$L_c \approx \frac{K\lambda}{\beta_{2\theta} \cos \theta}, \quad (16.246)$$

with $K = 0.90$ and $\lambda = 0.154 \text{ nm}$ gives

$$L_c \approx \frac{0.90(0.154 \text{ nm})}{(3.49 \times 10^{-3}) \cos 20^\circ} \quad (16.247)$$

$$\approx 42.3 \text{ nm.} \quad (16.248)$$

The result should be described as a characteristic coherent domain size within the assumptions of the Scherrer model. It is not automatically the physical particle diameter. Residual instrumental broadening, strain, defects and peak-shape conventions can all alter the inferred value.

Problem 39: compact spectrometer sampling

For 1200 lines/mm,

$$d = 833.3 \text{ nm.} \quad (16.249)$$

At $\lambda = 500 \text{ nm}$ and $m = 1$,

$$\sin \theta = \frac{500}{833.3} = 0.600, \quad \theta \approx 36.87^\circ, \quad \cos \theta = 0.800. \quad (16.250)$$

With $x = f \tan \theta$,

$$\frac{dx}{d\lambda} = \frac{fm}{d \cos^3 \theta}. \quad (16.251)$$

For $f = 0.100 \text{ m}$,

$$\frac{dx}{d\lambda} \approx \frac{0.100}{(8.333 \times 10^{-7})(0.800)^3} \approx 2.34 \times 10^5 \frac{\text{m}}{\text{m}} = 0.234 \text{ mm/nm.} \quad (16.252)$$

A $10 \mu\text{m} = 0.010 \text{ mm}$ pixel therefore spans approximately

$$\Delta\lambda_{\text{pix}} \approx \frac{0.010}{0.234} \text{ nm} \approx 0.0427 \text{ nm.} \quad (16.253)$$

The illuminated groove number is

$$N = (1200 \text{ mm}^{-1})(15 \text{ mm}) = 1.8 \times 10^4, \quad (16.254)$$

so the ideal grating-limited interval is

$$\Delta\lambda_{\text{grating}} \approx \frac{500 \text{ nm}}{1.8 \times 10^4} \approx 0.0278 \text{ nm.} \quad (16.255)$$

With a camera centered on the central diffracted ray, the local dispersion would instead be $f/(d \cos \theta) = 0.150 \text{ mm/nm}$, giving 0.0667 nm per pixel. Thus the detector mapping changes the number, and in either stated geometry one pixel spans a larger interval than the grating's Rayleigh separation. Multiple pixels per resolution element would normally be desired, so detector sampling and/or focal length must be considered in the design rather than assuming the grating alone determines system resolution.

Problem 44: uncertainty in Bragg spacing

For first order,

$$d = \frac{\lambda}{2 \sin \theta}, \quad \theta = \frac{2\theta_{\text{meas}}}{2} = 30.00^\circ. \quad (16.256)$$

Thus

$$d = \frac{0.15406 \text{ nm}}{2 \sin 30.00^\circ} = 0.15406 \text{ nm}. \quad (16.257)$$

The uncertainty in 2θ is 0.04° , so the uncertainty in θ is 0.02° , or

$$\sigma_\theta = 0.02 \frac{\pi}{180} \approx 3.49 \times 10^{-4} \text{ rad}. \quad (16.258)$$

Differentiating,

$$\frac{1}{d} \frac{dd}{d\theta} = -\cot \theta. \quad (16.259)$$

Hence the magnitude of the fractional uncertainty is

$$\frac{\sigma_d}{d} \approx |\cot \theta| \sigma_\theta = (\sqrt{3})(3.49 \times 10^{-4}) \approx 6.05 \times 10^{-4}. \quad (16.260)$$

Therefore

$$\sigma_d \approx (0.15406 \text{ nm})(6.05 \times 10^{-4}) \approx 9.3 \times 10^{-5} \text{ nm}. \quad (16.261)$$

A suitable report is

$$\boxed{d = (0.15406 \pm 0.00009) \text{ nm}}, \quad (16.262)$$

under the stated assumption that wavelength and systematic angular calibration uncertainties are negligible.

Problem 45: the shift theorem and diffraction intensity

Let the original aperture be $A(x)$ with Fourier amplitude

$$\tilde{A}(f) = \int_{-\infty}^{\infty} A(x) e^{-i2\pi f x} dx. \quad (16.263)$$

Translate the aperture by x_0 :

$$A_{x_0}(x) = A(x - x_0). \quad (16.264)$$

Its transform is

$$\tilde{A}_{x_0}(f) = \int A(x - x_0) e^{-i2\pi f x} dx. \quad (16.265)$$

Let $u = x - x_0$, so $x = u + x_0$ and $dx = du$. Then

$$\tilde{A}_{x_0}(f) = \int A(u) e^{-i2\pi f(u+x_0)} du \quad (16.266)$$

$$= e^{-i2\pi f x_0} \tilde{A}(f). \quad (16.267)$$

The multiplicative factor has unit magnitude, so

$$|\tilde{A}_{x_0}(f)|^2 = |\tilde{A}(f)|^2. \quad (16.268)$$

Hence the ideal far-field intensity pattern of an isolated aperture is unchanged by rigid transverse translation, even though the complex phase of the far field changes. This is a precise example of why intensity alone does not contain all information about the optical field.

Bibliography

- [1] D. Halliday, R. Resnick and J. Walker, *Fundamentals of Physics*, 12th ed., Vol. 2, Wiley, Chapter 36, especially single-slit diffraction, circular apertures, gratings and X-ray diffraction.
- [2] R. A. Serway and J. W. Jewett, Jr., *Physics for Scientists and Engineers with Modern Physics*, 9th ed., Brooks/Cole, Cengage Learning, Chapter 38, Sections 38.1–38.5.
- [3] S. J. Ling, J. Sanny and W. Moebs, *University Physics, Volume 2*, OpenStax, Chapter 16. Electromagnetic-wave prerequisites for the diffraction theory in this chapter.
- [4] S. J. Ling, J. Sanny and W. Moebs, *University Physics, Volume 3*, OpenStax, Chapter 4; [Single-Slit Diffraction](#); [Circular Apertures and Resolution](#).
- [5] Nikon MicroscopyU, *Modulation Transfer Function and Resolution*; [MTF article](#); [resolution article](#).
- [6] Edmund Optics, *Sensors and Lenses*; [online application note](#).