Waves and optics

Follow a wave from its source through interfaces and apertures, then use interference and diffraction to explain what an image or a spectrum can resolve.

Read the subject in depth

Each volume develops its arguments in sequence, with worked problems and links to experiments.

- ### Optical fields, instruments and observations

  Trace an optical field through a physical instrument. Calculate the signal a detector measures and test the approximations on which the interpretation rests.

Advanced study across the subject

Each sequence connects the course to detailed chapters in neighbouring subjects. The complete course PDF includes their derivations, examples and exercises.

1. ### A wave in space, matter and a guide

   Rebuild a propagating mode from the field equations. Track vector polarisation and flux through an interface, then separate material from geometric dispersion.

2. ### What a detector samples

   Compare radiation, scattering and the measured diffraction pattern. Examine how aperture sampling, transfer calibration and finite coherence alter the observable.

3. ### Mechanical and thermal effects of light

   Calculate how a beam can translate a bead, couple resonances or exchange energy with a moving boundary. Test each proposal against a measured signal.

How the course fits together

1. **Chapters 1–4**

   A local force balance on a string gives a wave speed. Superpose travelling waves to satisfy interfaces and make packets. In the short-wavelength limit their phase fronts lead to rays, focal lengths and images.

2. **Chapters 5–8**

   Restore the information that ray tracing discards. Polarisation describes a vector field, coherence determines which fields can interfere, and an aperture maps a wavefront to a diffraction pattern that limits image detail.

3. **Chapters 9–12**

   Repeated optical paths select frequencies in an interferometer. Many illuminated grooves resolve nearby wavelengths. Beams and widely separated receivers offer two different ways to infer spatial structure from wave phase.

Contents

1. Chapter 1 · Travelling waves and energy on a string
2. Chapter 2 · Impedance, reflection and standing waves
3. Problem sheet 1 · Pulses, junctions and resonances
4. Chapter 3 · Wave packets and dispersion
5. Chapter 4 · Optical paths, lenses and images
6. Problem sheet 2 · Packets and paraxial imaging
7. Chapter 5 · Polarisation and phase retardation
8. Chapter 6 · Coherence and interferometry
9. Problem sheet 3 · Polarisation and coherence
10. Chapter 7 · Diffraction as a Fourier transform
11. Chapter 8 · Images through a finite pupil
12. Problem sheet 4 · Fourier pupils and image sampling
13. Chapter 9 · Multiple-beam interference in a cavity
14. Chapter 10 · How a grating separates wavelengths
15. Problem sheet 5 · Orders, linewidths and spectral identity
16. Chapter 11 · Gaussian beams and focusing
17. Chapter 12 · Measuring a star with interference
18. Problem sheet 6 · Beam transport and spatial coherence
19. Chapter 13 · Reflection, polarisation and the field beyond a reflecting surface
20. Chapter 14 · How a dielectric slab chooses its guided modes
21. Chapter 15 · Why opening a circular aperture can darken its centre
22. Chapter 16 · When do two curved mirrors support a confined beam?
23. Chapter 17 · Threshold, gain saturation and the decay of stored light
24. Chapter 18 · Turning invisible phase into visible intensity in a Fourier plane
25. Chapter 19 · Building a dielectric mirror from weakly reflecting interfaces
26. Chapter 20 · Why a grating makes images of itself in empty space
27. Problem sheet 7 · Choosing an output mirror for a small optical resonator
28. Problem sheet 8 · Measure a double source with a two-aperture interferometer
29. Optical fields, instruments and observations
30. A wave in space, matter and a guide
31. What a detector samples
32. Mechanical and thermal effects of light
33. Applied study
34. Notation and assumptions
35. Further reading

Chapter 1 · Travelling waves and energy on a string

Derive wave speed from the local force balance and determine how much average power a transverse travelling wave carries.

FROM FIRST PRINCIPLES

Derive the wave speed from transverse force balance

Select a short string segment of unperturbed length dx and mass $\mu\,dx$, where μ is mass per unit length. Assume a flexible string with uniform tension T, small slopes and negligible damping. Its neighbours pull tangentially at either end. For small angles the transverse component is approximately T times the slope. Oppositely directed pulls cancel when the two ends have equal slopes, while a change in slope produces a net transverse force.

$$dF_y = T\,[\partial_x y(x+dx,t) - \partial_x y(x,t)] \simeq T\,\partial_x^2 y\,dx.$$

Newton's second law equates this force to the mass times transverse acceleration. Divide by segment length and compare with the derivatives of a profile translated at speed c.

$$\partial_t^2 y = \frac{T}{\mu}\,\partial_x^2 y,$$

$$y = f(x-ct) \implies \partial_t^2 y = c^2 f'',$$

$$\partial_x^2 y = f''.$$

A freely travelling shape therefore moves at $c = \sqrt{T/\mu}$. Extra tension supplies a larger force for the same curvature. More mass per unit length reduces the acceleration produced by that force.

Background · Wavenumber

$$k = 2\pi/\lambda$$

The wavenumber gives phase change per unit distance in a homogeneous medium. Wavelength must be specified in the medium in which propagation is calculated.

DERIVATION AND APPLICATIONS

Calculate the energy and power of a travelling pulse

A single pulse has no unique frequency, but a measurement of its slope can still determine the total energy it carries. Model a long string with constant tension T, line density μ, negligible bending stiffness and small transverse slope. Choose a right-moving displacement $y(x,t) = f(x-ct)$, with $c = \sqrt{T/\mu}$. Stretching a segment increases its length by approximately $(\partial_x y)^2 dx/2$, while its transverse velocity is $\partial_t y = -c\partial_x y$. The kinetic and elastic energies per unit unperturbed length are consequently equal.

$$u = \tfrac{1}{2}\mu(\partial_t y)^2 + \tfrac{1}{2}T(\partial_x y)^2 = T[f'(x-ct)]^2.$$

The left portion of the string does work on the right portion through its transverse force at a cut. With positive power defined towards increasing x, multiply that force by the velocity of the point at the cut. The result is nonnegative even as the two slopes of the passing pulse have opposite signs.

$$P = -T(\partial_x y)(\partial_t y) = cT[f'(x-ct)]^2 = cu.$$

Consider an isolated Gaussian displacement with peak height A and spatial scale s,

$$f(\xi) = A\exp[-\xi^2/(2s^2)].$$

Its slope is $-\xi f/s^2$. Integrating over the entire string using the second moment of a Gaussian gives

$$E = T\frac{A^2}{s^4}\int_{-\infty}^{\infty}\xi^2 e^{-\xi^2/s^2}d\xi = \frac{\sqrt{\pi}TA^2}{2s}.$$

A broader pulse of the same height has gentler slopes and stores less energy. The displacement maximum itself has zero instantaneous power because its slope vanishes. Two power peaks pass a fixed sensor on either side of that maximum.

WORKED EXAMPLE

Budget the energy of a laboratory pulse

Change the inputs and investigate this derivation.

A string under 40.0 N tension carries a Gaussian pulse of height 4.00 mm with $s = 50.0$ mm. Its energy is $(\sqrt{\pi}/2)(40.0)(0.00400)^2/0.0500 = 11.3$ mJ. The largest slope is $(A/s)e^{-1/2} = 0.0485$, supporting the small-slope approximation. If the string absorbs the pulse completely at its far end, the absorber receives the full 11.3 mJ.

SECOND APPLICATION

Reconstruct pulse width from two power maxima

Change the inputs and investigate this derivation.

On a second string the pulse speed is 25.0 m/s. Since P is proportional to $\xi^2 e^{-\xi^2/s^2}$, differentiating gives maxima at $\xi = \pm s$. Two measured power peaks 6.00 ms apart therefore imply $s = c\Delta t/2 = 75.0$ mm. For $A = 3.00$ mm and $T = 18.0$ N the predicted peak power is $cT(A/s)^2/e = 0.265$ W, an independent amplitude check.

Try it · String pulse

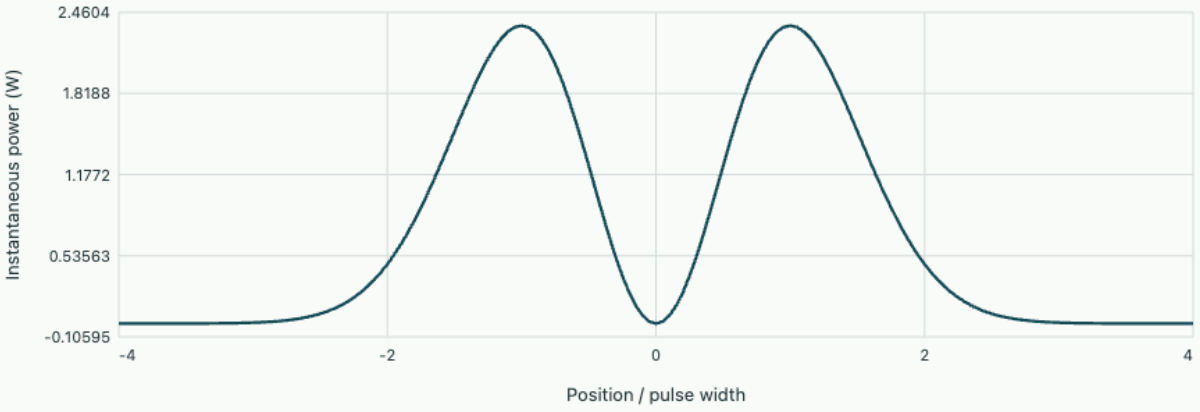
Vary the physical parameters in the derived relation. Use the units and assumptions in the readout to connect the curves to the two worked examples. A parameter study can test the forward prediction; the same relation can then be inverted to infer a parameter from a measurement.

Open this live calculation

Try the calculation

Pulse height (mm) **4**

Gaussian width (mm) **50**



Travelling Gaussian pulse

Tension 40 N, wave speed 25 m/s. Total pulse energy 11.344 mJ. Maximum slope 0.048522; the linear string approximation requires small slope.

Choose the displacement and the approximation

Strike one end of a stretched rope. A crest moves down its length, yet a marked point on the rope moves mainly up and down. Distinguishing the motion of a disturbance from the motion of the material is the starting point for predicting wave speed. Consider a uniform flexible string under tension F_T with mass per unit length μ. Its transverse displacement from a straight equilibrium line is $y(x,t)$. Assume slopes $|\partial_x y| \ll 1$, negligible bending stiffness and negligible damping.

These assumptions have measurable consequences. To first order a segment of horizontal length dx retains that length and has mass $\mu\,dx$. The tension magnitude can be treated as its equilibrium value, while the directions at the two ends vary with the local slope. For larger slopes the tension can change as the string stretches, and longitudinal motion need no longer decouple from transverse displacement.

Balance forces on one small segment

At the two ends of a short segment the vertical tension components have opposite signs. Expand the difference of the slopes to first order in the segment length. Newton's law then reads

$$\begin{aligned}\mu\, dx\, \partial_t^2 y &= F_T\left[\partial_x y(x+dx,t) - \partial_x y(x,t)\right] \\ &= F_T\, dx\, \partial_x^2 y + O(dx^2).\end{aligned}$$

Dividing by the segment mass and taking the limit gives the wave equation with speed c,

$$\partial_t^2 y = c^2 \partial_x^2 y,$$

$$c = \sqrt{\frac{F_T}{\mu}}.$$

The equation predicts that a local curvature accelerates material towards straightening the string. Any twice-differentiable shape $g(x - ct)$ travels right at speed c because its second time derivative is $c^2 g''$. A second independent shape $h(x + ct)$ travels left. Their superposition solves the same linear equation, and initial displacement alone generally does not specify how the disturbance splits. An initial velocity field supplies the second piece of information.

For example, lifting a narrow region and releasing the whole string from rest produces opposite-going copies of half the initial displacement on a long uniform string. Launching the same shape exclusively to the right requires $y_t = -c y_x$ at release. A sequence of still images can reveal the direction of motion, but a single displacement profile usually cannot.

Read a frequency from moving wavefronts

A sinusoidal displacement at a fixed point repeats every source period T_s. A microphone in motion does not stay at that point between observations. To relate its measured frequency to the source clock, choose the positive axis along the sound ray, from the source towards the receiver. Let c be the sound speed in still air and v_s, v_r the signed source and receiver velocities in that same air frame. Assume steady subsonic motion along the ray and negligible reflection.

At emission of the next crest, the first crest has travelled cT_s and the emitter has moved $v_s T_s$. The separation of the crests ahead of the emitter is consequently $\lambda = (c - v_s)T_s$. The receiver intercepts them at relative speed $c - v_r$. Dividing the distance swept per second by the spacing gives

$$\frac{f_r}{f_s} = \frac{c - v_r}{c - v_s}.$$

A speaker on a test trolley emits 680 Hz while travelling at 15 m/s towards a microphone moving towards it at 5 m/s. Take $c = 340$ m/s. The crest spacing is $325/680 = 0.478$ m, and the microphone sweeps through the crests at 345 m/s. It registers 721.85 Hz. Holding the microphone still instead gives 711.38 Hz. The two instruments approach one another in either experiment, but moving the emitter changes wavelength while moving only the microphone leaves it fixed.

Two checks prevent sign errors. If both instruments travel at the same speed the ratio is one. For velocities small compared with c, expansion to first order gives

$$\Delta f/f_s \simeq (v_s - v_r)/c.$$

Separate dependence on the two air-frame velocities reappears at second order. The distinction matters in a wind tunnel, where ground speed and air-frame speed differ. The derivation uses a material medium and cannot be used unchanged for light in vacuum.

Practise on sheet 1 · Recover a source speed from two microphones. Further reading · University of Illinois Urbana-Champaign · Physics 406, Acoustical Physics of Music · Lecture III, part 2 · Doppler effect for a moving source and observer.

Locate the energy being transmitted

Each segment stores kinetic energy $\mu y_t^2 dx/2$. Stretching it adds length $[\sqrt{1+y_x^2}-1]dx \simeq y_x^2 dx/2$, costing potential energy $F_T y_x^2 dx/2$ at the assumed constant tension. Defining their sum per length as u, direct differentiation and substitution of the wave equation yield

$$\partial_t u = \partial_x (F_T y_t y_x).$$

Thus the positive-direction power is $P = -F_T y_t y_x$ and $\partial_t u + \partial_x P = 0$.

For a right-moving wave $y_t = -c y_x$, so the power is positive and equals cu. A left-moving wave carries power in the opposite direction. The signs come from the mechanical work done across a cut. In a standing wave the two powers need not add constructively even where the displacement amplitude is large. An average amplitude alone therefore measures power only once the propagation conditions have been established.

Compare two launches

Let $F_T = 100$ N, $\mu = 0.010$ kg/m and frequency $f = 20.0$ Hz. The wave speed is 100 m/s and wavelength is 5.0 m. A sinusoidal right-moving displacement of peak amplitude $A = 3.0$ mm has peak transverse material speed $2\pi f A = 0.377$ m/s. The largest slope is $2\pi A/\lambda = 0.00377$, supporting the small-slope approximation.

The mean power is

$$\mu c (2\pi f A)^2/2 = 0.0711\,\mathrm{W}.$$

Doubling the amplitude at the same frequency quadruples mean power while keeping the propagation speed fixed. If the tension is doubled at fixed amplitude and frequency, the speed and wavelength both increase by $\sqrt{2}$ and the mean power increases by $\sqrt{2}$. These predictions separate the energy required to maintain a sinusoidal drive from the arrival time of a signal along the string.

The two field measurements needed for instantaneous power are the transverse velocity and the local slope. Measuring displacement alone at one point gives velocity after differentiation, but not the spatial slope. A

second nearby displacement probe can approximate that slope if its spacing is short compared with the wavelength. Standing waves illustrate why this matters: a point may move rapidly while the time-averaged transport across it is zero because counterpropagating components carry equal powers.

Hear the initial shape in the harmonics of a string

On a finite uniform string with fixed ends the boundary values must vanish at $x = 0, L$ at every time. Separating the wave equation gives spatial eigenfunctions $\sin(n\pi x/L)$ and angular frequencies $n\pi c/L$. If the initial transverse velocity is zero, each excited mode starts in its cosine time phase. To determine its amplitude multiply the initial displacement $f(x)$ by the corresponding sine and integrate. Orthogonality gives

$$y(x,t) = \sum_{n=1}^{\infty} b_n \sin(n\pi x/L)\cos(n\pi c t/L),$$

$$b_n = \frac{2}{L}\int_0^L f(x)\sin(n\pi x/L)\,dx.$$

Pluck the midpoint upward by height $h \ll L$ and release the resulting triangle from rest. Its slopes are $2h/L$ and $-2h/L$. Integrating the coefficient by parts makes the endpoint term vanish because f vanishes at both ends. The two remaining constant-slope integrals give

$$b_n = 8h\sin(n\pi/2)/(n^2\pi^2).$$

Reflection symmetry suppresses every even mode. The negative coefficient for the third mode fixes its phase relative to the fundamental. Its energy remains positive.

Take $L = 0.600$ m, $c = 120$ m/s and $h = 3.00$ mm. The first and third frequencies are 100 and 300 Hz and the amplitudes are $+2.43$ and -0.270 mm. Energy in a mode is proportional to $n^2 b_n^2$ because its maximum spatial slope scales with n. The third mode therefore has one ninth of the first mode's energy, even though its displacement amplitude magnitude is one ninth and its amplitude squared is only one eighty-first. A pickup at an individual mode node may suppress the measured line even when the mode is present in the string.

Practise on sheet 1 · Predict which note disappears for an off-centre pluck. Further reading · Massachusetts Institute of Technology · 8.03SC, The Physics of Waves · §6.2.1 The String with Fixed Ends; §6.2.4 Plucking a String.

Reconstruct two outgoing pulses from initial data

A long uniform string can be started by specifying both its displacement and its velocity. Assume small slopes, constant tension and observations made before reflections return from the ends. The one-dimensional wave equation factors into left- and right-travelling pieces, so write $y(x,t) = F(x - ct) + G(x + ct)$. At the release time the measured shape is $f(x) = F(x) + G(x)$ and the transverse velocity is $g(x) = -cF'(x) + cG'(x)$. Differentiating the shape equation and solving this pair gives

$$F' = (f' - g/c)/2,$$
$$G' = (f' + g/c)/2.$$

Integrate those derivatives and evaluate the two pieces at their respective moving arguments. Constants cancel when the prescribed initial shape is restored, yielding the solution

$$y(x,t) = \frac{f(x-ct) + f(x+ct)}{2} + \frac{1}{2c}\int_{x-ct}^{x+ct} g(s)\,ds.$$

The interval in this formula contains every initial point whose signal can have reached the observation point by time t. A localized shape released from rest splits into two half-height copies. A localized velocity impulse produces a different response because the integral accumulates the initial velocity over a growing interval. Measuring only a release photograph cannot distinguish these preparations. A right-travelling shape alone requires the specific initial velocity $g = -cf'$, consistent with the local slope-to-velocity relation of a travelling wave.

Strike a short region of the string

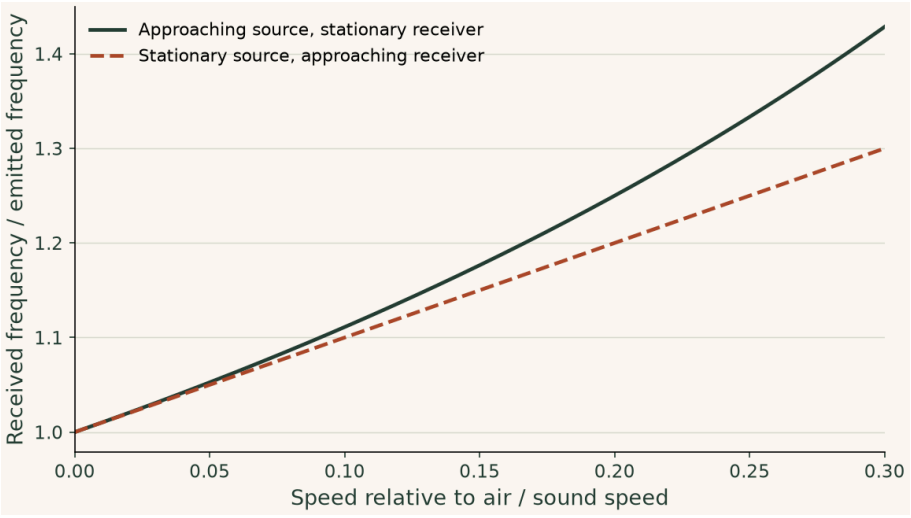
A string has $c = 50.0$ m/s and line density $\mu = 0.0200$ kg/m. Initially it is straight. Impart transverse velocity $g_0 = 0.400$ m/s over the interval $-a < x < a$ with $a = 0.100$ m, and zero elsewhere. This rectangular velocity profile idealizes a short, distributed tap. At the centre, while $ct < a$, the integration interval lies inside the struck region and $y(0,t) = g_0 t$. At 1.00 ms the centre has moved 0.400 mm. After $a/c = 2.00$ ms the entire initial velocity region lies inside the interval, and the centre takes the constant displacement $g_0 a/c = 0.800$ mm.

At that later stage the centre is at rest and locally straight while two wavefront regions move away in opposite directions. The energy that initially resided in the moving patch was $\mu(2a)g_0^2/2 = 0.320$ mJ, with half propagating in each direction by symmetry. A permanent shift at a fixed point of an infinite ideal string costs no local curvature energy after the fronts pass. Finite fixed ends eventually send return waves, invalidating the infinite-string prediction at late times. An array of displacement sensors would see a linear rise followed by a local plateau, while a velocity probe would record a finite-time pulse and return to zero.

The example also predicts what a position probe placed outside the struck region records before any boundary reflection. Its signal begins when the nearer edge of the initial interval enters the backwards propagation cone. It reaches the plateau when the farther edge enters. At a point 0.500 m from centre, those events occur at 8.00 and 12.0 ms. The local rise lasts 4.00 ms, twice the centre rise time, but only one travelling front contributes there. A second sensor equally far on the opposite side should give the same record, providing an immediate symmetry check on the applied tap.

Does it matter which instrument is moving?

Compare two acoustic measurements at the same closing speed in still air. Move the source in the first experiment and move the microphone in the second. Their rates of approach agree but their crest spacings differ.



At one fifth of the speed of sound the factors are 1.25 and 1.20. Both curves have slope one at rest. Their separation away from zero shows where an approximation depending only on the closing speed loses information.

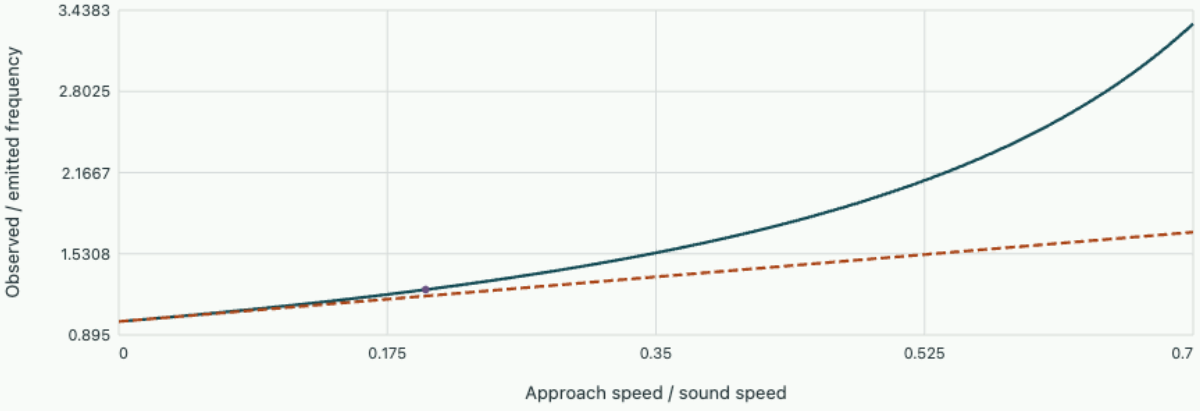
Try it · Doppler

Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Approach speed / sound speed **0.2**



Moving source, fixed receiver | Moving receiver, fixed source | Selected source speed

At selected speed: moving-source factor 1.25; moving-receiver factor 1.2. The medium defines the rest frame.

▶ **Read values from the plot**

Chapter 2 · Impedance, reflection and standing waves

Derive amplitude and energy reflection at a string junction and recover the resonance conditions of a finite resonator.

FROM FIRST PRINCIPLES

Match displacement and force at a string joint

Join two long strings under the same tension T, with linear densities μ_1, μ_2. Assume small transverse displacements and a massless joint with no external force. A sinusoidal wave arrives from string one. Write A, B, C for complex incident, reflected and transmitted displacement amplitudes at the join. Because the strings remain attached, their endpoint positions agree at each instant.

$$A + B = C.$$

The transverse components of the two tensions must balance. A right-going wave has slope amplitude ikA, whereas reversing its direction changes the sign of that slope. With $k_j = \omega/c_j$, the force equality becomes

$$Tk_1(A - B) = Tk_2C,$$

$$Z_j = T/c_j = \sqrt{T\mu_j}.$$

Eliminate C using continuity and solve for the two wave amplitudes. The sign of the reflection records a possible reversal of displacement.

$$\frac{B}{A} = \frac{Z_1 - Z_2}{Z_1 + Z_2},$$

$$\frac{C}{A} = \frac{2Z_1}{Z_1 + Z_2}.$$

A matched joint gives zero reflection. Equal-frequency incident and reflected waves overlap to produce fixed interference minima and maxima. Two reflecting endpoints select a standing pattern when the round trip reproduces the original phase.

DERIVATION AND APPLICATIONS

Scatter a string wave from a small bead

Two strings with identical impedance can reflect a wave at their joint if the joint has inertia. Attach a bead of mass M between two long uniform portions of a string under tension T. Assume a pointlike bead, small transverse motion and negligible damping. Write displacements proportional to $e^{-i\omega t}$, with incident, reflected and transmitted amplitudes A, B and C. Continuity of the string displacement gives $C = A + B$. Newton's law for the bead requires the difference in transverse tension components to accelerate this common displacement.

$$-M\omega^2 C = T\left[\partial_x y(0^+) - \partial_x y(0^-)\right]_{\text{amplitude}} = ikT[C - (A - B)] = 2ikTB.$$

The reflected wave has the opposite spatial phase gradient, which produces the second copy of B in the tension jump. Introduce the travelling-wave impedance $Z = T/c$ and dimensionless mass parameter $s = \omega M/(2Z)$. Using $k = \omega/c$, the junction equation gives $B = isC$. Eliminate C through displacement continuity to obtain

$$r = \frac{B}{A} = \frac{is}{1 - is},$$
$$t = \frac{C}{A} = \frac{1}{1 - is}.$$

For identical strings the mean incident and outgoing powers carry the same impedance and frequency factors. Squaring the amplitude ratios gives the energy fractions and explicitly accounts for the entire incident power.

$$\mathcal{R} = \frac{s^2}{1 + s^2},$$
$$\mathcal{T} = \frac{1}{1 + s^2},$$
$$\mathcal{R} + \mathcal{T} = 1.$$

For slowly oscillating signals the bead follows the string and becomes almost transparent. At high frequency the transmitted amplitude tends to zero and r approaches -1, the displacement reversal of a fixed endpoint.

WORKED EXAMPLE

Predict the power passing a marker bead

Change the inputs and investigate this derivation.

With $Z = 0.500$ kg/s, $M = 5.00$ g and $\omega = 200$ rad/s, the junction has $s = 1.00$. Each outgoing string carries half of the incident power. The transmitted displacement magnitude is $1/\sqrt{2} = 0.707$ of the incident value, and $t = (1 + i)/2$ in the selected phase convention. An incident 12.0 mW wave therefore delivers 6.00 mW beyond the bead.

SECOND APPLICATION

Weigh an attached marker by reflection

Change the inputs and investigate this derivation.

Another string has $Z = 0.300$ kg/s. At 30.0 Hz its reflected power fraction is 0.200. The inversion $s = \sqrt{\mathcal{R}/(1 - \mathcal{R})}$ gives 0.500. Its inferred bead mass is

$$2Zs/(2\pi f) = 1.59\,\mathrm{g}.$$

Doubling the frequency should bring s to unity and the reflected power to 50.0 percent, a check independent of the calibrated incident displacement.

Try it · Bead scattering

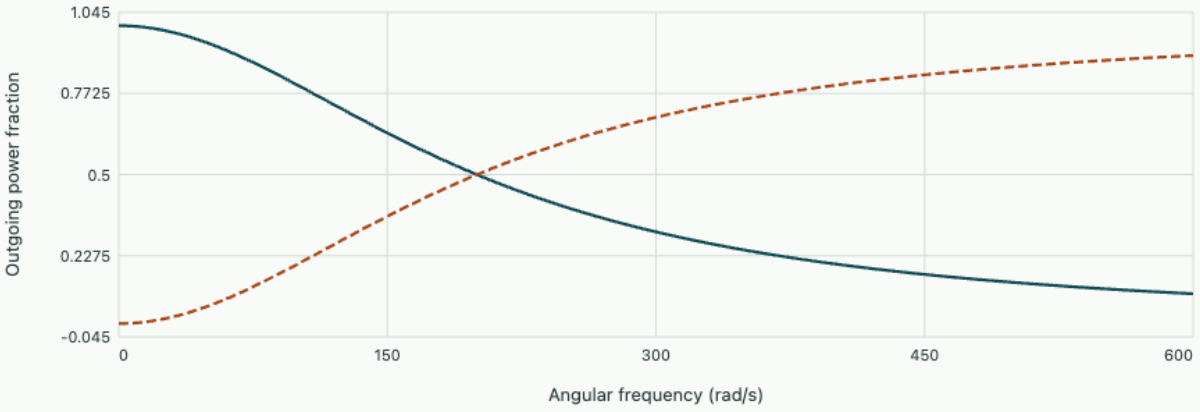
Vary the physical parameters in the derived relation. Use the units and assumptions in the readout to connect the curves to the two worked examples. A parameter study can test the forward prediction; the same relation can then be inverted to infer a parameter from a measurement.

Open this live calculation

Try the calculation

Attached mass (g) **5**

String impedance (kg/s) **0.5**



Transmitted power Reflected power

Junction mass on two identical semi-infinite strings: s=ωM/(2Z), T=1/(1+s²), R=s²/(1+s²). Their sum is one.

Make the joint part of the model

A travelling disturbance reaches the junction between a light string and a heavy one held at the same tension. The animation of the junction might show an inverted reflected crest. To tell which sign and amplitude are physically possible, isolate the joint. We assume it has negligible mass, that the strings are tied without slip and that it experiences no externally applied transverse force. The displacement must be continuous and the transverse components of tension must balance.

Place the junction at $x = 0$ and let positive x point into the transmitted string. Use a harmonic time convention $e^{-i\omega t}$. In the first string write the complex displacement as $A_i e^{ik_1 x} + A_r e^{-ik_1 x}$, and in the second as $A_t e^{ik_2 x}$. Since the joint experiences one oscillation frequency, wavelength changes with the respective speed. A change of wavelength alone says nothing about the reflected power.

Match motion and transverse force

At the joint the displacement condition gives

$$A_i + A_r = A_t.$$

The force condition gives

$$F_{T1} k_1 (A_i - A_r) = F_{T2} k_2 A_t.$$

Defining the string impedance as $Z_j = F_{Tj}/c_j = \sqrt{F_{Tj}\mu_j}$ converts the second equation into $Z_1(A_i - A_r) = Z_2 A_t$. Under the equal static tensions assumed here, different impedances arise from different linear mass densities. Eliminating the transmitted amplitude gives

$$r = \frac{A_r}{A_i} = \frac{Z_1 - Z_2}{Z_1 + Z_2},$$

$$t = \frac{A_t}{A_i} = \frac{2Z_1}{Z_1 + Z_2}.$$

The sign of r describes displacement relative to the incident displacement at the junction. For $Z_2 \gg Z_1$ the junction approaches a fixed end, $r \to -1$. For $Z_2 \ll Z_1$ it approaches a free end, $r \to +1$. These limiting experiments check the force signs independently.

Background · Which sign changes at a free end?

A free end of an ideal stretched string carries zero transverse force, hence zero slope. For harmonic displacement $y = (Ae^{ikx} + Be^{-ikx})e^{-i\omega t}$, differentiating reverses the sign of the reflected term. If the end lies at $x = 0$, zero slope requires $B = A$. Holding the same end at zero displacement instead gives $B = -A$.

$$y(0) = 2Ae^{-i\omega t} \quad \text{(free end)},$$

$$\partial_x y(0) = 2ikAe^{-i\omega t} \quad \text{(fixed end)}.$$

Either end returns all incident power. A sensor measuring string displacement sees an antinode at the free end and a node at the fixed end. A force sensor responds to slope and gives the opposite pattern. State the measured quantity when assigning a phase reversal to a reflected wave.

Compare amplitude and power

The mean power of a sinusoidal travelling wave is $Z\omega^2|A|^2/2$. Thus the reflected power fraction is $|r|^2$ and the transmitted power fraction is $(Z_2/Z_1)|t|^2$. Adding the two fractions gives one for the ideal lossless junction. A transmitted displacement larger than the incoming displacement is consistent with energy conservation when the second medium has sufficiently small impedance.

Suppose both strings have tension 80 N, while $\mu_1 = 0.0050$ and $\mu_2 = 0.020$ kg/m. The ratio of impedances is two, so $r = -1/3$ and $t = 2/3$. Of 0.90 W entering, 0.10 W is reflected and 0.80 W travels onwards.

Reversing the incident direction makes $r = +1/3$ and $t = 4/3$ while leaving the reflected fraction at one ninth. A sensor that records only displacement peaks could wrongly interpret the reverse experiment as amplification.

Build a resonator using two reflections

Put reflecting ends a distance L apart. After a round trip a monochromatic wave acquires propagation phase $2kL$ and the phases of both reflection coefficients. A repeated wave pattern requires their sum to be an integer multiple of 2π. With two fixed ends, the two sign reversals add a complete cycle and $k_n = n\pi/L$, $n = 1, 2, \ldots$. With one fixed and one free end there is only one reversal, giving $k_n = (2n-1)\pi/(2L)$. The boundary conditions, rather than the length alone, determine whether an antinode or a node sits at an end.

For a 0.60 m string with speed 120 m/s the fixed–fixed fundamental is 100 Hz. The fixed–free fundamental is 50 Hz. Replacing one ideal end by a partially transmitting attachment also leaks energy each round trip. One then observes a finite resonance width and a decaying signal after switching off the drive. A resonance frequency tests phase closure, whereas the decay time tests energy escape and internal loss.

A pulse experiment can also distinguish a fixed end from a free end without measuring force. At a fixed end the incoming and reflected displacements must cancel, so an upward pulse returns downward. At a free end the transverse force is zero, requiring slopes to cancel; the displacement returns with the original sign. In both cases ideal reflection conserves energy. A negative displacement reflection coefficient records a phase reversal, even though the reflected energy remains positive.

Terminate a string with a spring and a damper

A fixed or free termination gives only two extreme reflection phases. Connect the end of a uniform string to a linear spring of stiffness k_s and a viscous damper b acting in parallel against ground. The endpoint is massless, and the incoming transverse displacement is harmonic. Use temporal phasors $e^{i\omega t}$ and let the string extend from negative x to its terminal at zero. Define the positive load force as the force the string supplies to the termination. The load force divided by endpoint velocity is $Z_L = b + k_s/(i\omega)$. A travelling string wave has characteristic force-to-velocity ratio $Z_0 = \sqrt{T\mu}$.

The incident and reflected endpoint velocities add, but their forces enter with opposite signs because their directions of transport differ. If r is the reflected-to-incident velocity or displacement amplitude, the boundary equation is

$$Z_0(1 - r) = Z_L(1 + r).$$

Thus

$$r = \frac{Z_0 - Z_L}{Z_0 + Z_L},$$

$$1 - |r|^2 = \frac{4Z_0 \operatorname{Re} Z_L}{|Z_0 + Z_L|^2}.$$

For a pure spring the reflected power fraction is one, although its phase varies continuously with frequency. The spring temporarily stores energy and returns it. A matched pure damper $b = Z_0$ with no spring absorbs

an incident wave completely. In the low-frequency limit a nonzero spring has a large impedance magnitude and its end approximates a fixed boundary. As frequency increases the spring contributes less force for the same velocity, and the damper determines the response.

Infer the absorbed power from an endpoint motion

Choose $T = 100$ N and $\mu = 0.0100$ kg/m, giving $Z_0 = 1.00$ N s/m. Fit $b = 1.00$ N s/m and $k_s = 20.0$ N/m, and drive at $\omega = 20.0$ rad/s. The load is $Z_L = 1.00 - i1.00$ N s/m, hence

$$r = i/(2 - i) = (-1 + 2i)/5.$$

Its magnitude squared is 0.200. Of 10.0 mW incident power, 2.00 mW returns and 8.00 mW is dissipated.

The incident velocity amplitude is

$$\sqrt{2P_{\mathrm{in}}/Z_0} = 0.141\,\mathrm{m/s}.$$

Total endpoint velocity amplitude is $|1 + r|$ times this, or 0.126 m/s. Independent damper heating is $b|v_{\mathrm{end}}|^2/2 = 8.00$ mW as required. The endpoint displacement amplitude is

$$0.126/20.0 = 6.32\,\mathrm{mm}.$$

At a much higher frequency the same termination approaches the matched damper limit. A complex reflection measurement therefore separates reactive force from dissipative force even when an amplitude-only measurement happens to give the same reflected fraction for two different fixtures.

One can invert the experiment to estimate the fixture properties from reflected phase. Algebraically the boundary equation gives

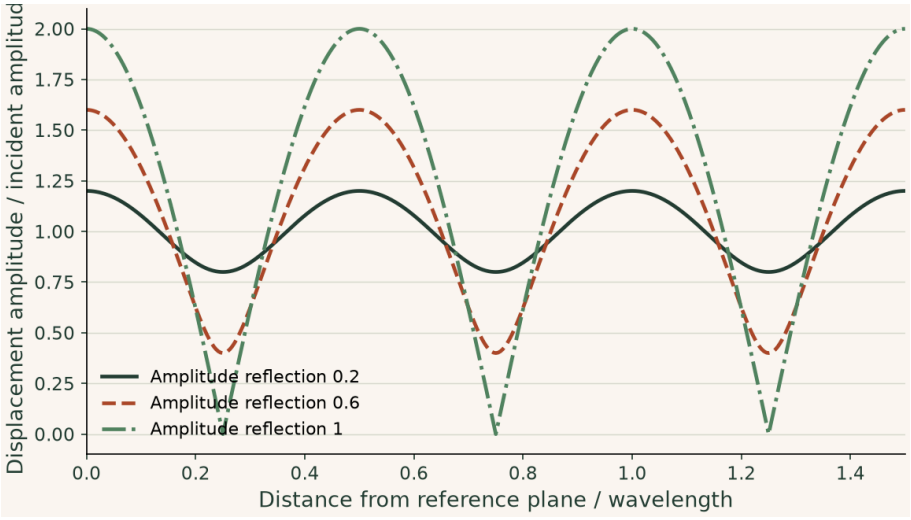
$$Z_L = Z_0(1 - r)/(1 + r).$$

The real part estimates the damper coefficient, and multiplying the negative imaginary part by angular frequency estimates spring stiffness. Near a fixed boundary the measured reflection is close to minus one. The small denominator then makes the stiffness estimate sensitive to phase errors even when reflected power is measured accurately. Repeating the experiment near the fixture's spring–damper crossover gives a better-conditioned comparison between the two independently inferred components.

Repeating the inversion at two frequencies also tests whether the assumed spring and damper coefficients remain constant.

Infer a reflection from a standing-wave scan

Superpose an incident harmonic wave of unit amplitude and a returned wave of real amplitude r. A probe that measures the oscillation amplitude at each point reports the modulus of their complex sum. Squaring first and adding intensities would erase the interference.



For r = 0.6 the extrema are 1.6 and 0.4, so their ratio is four. Successive minima lie half a wavelength apart. A nonzero minimum can represent partial reflection even with perfect experimental resolution.

Try it · Standing wave

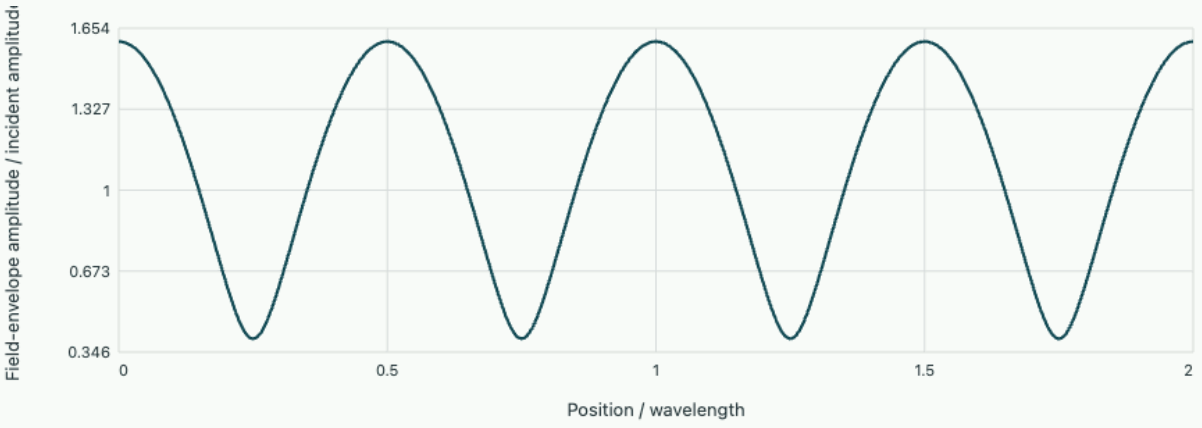
Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Reflected field amplitude ratio **0.6**

Reflection phase (degrees) **0**



Standing-wave envelope

Maximum/minimum 1.6 / 0.4. Standing-wave ratio 4. Minima are spaced by half a wavelength.

▶ **Read values from the plot**

Problem sheet 1 · Pulses, junctions and resonances

Use the force and energy balances of Lectures 1 and 2 to interpret motion at a string junction.

Select the direction of a pulse

A string has tension 72 N and linear density 0.020 kg/m. At one instant its displacement is $y = Ae^{-(x/w)^2}$ with $A = 2.0$ mm and $w = 0.10$ m. Specify the initial velocity field that sends the pulse only to the right. Find the material velocity at $x = w$ and the time for its peak to travel 3.0 m.

▼ **Relate the initial velocity to the slope**

The speed is

$$c = \sqrt{72/0.020} = 60\,\mathrm{m/s}.$$

A right-moving profile satisfies

$$y_t = -c y_x = 2cAx\,e^{-(x/w)^2}/w^2.$$

At $x = w$, the positive material velocity is

$$2(60)(0.0020)e^{-1}/0.10 = 0.883\,\mathrm{m/s}.$$

The crest travels 3.0 m in 0.050 s. Setting all velocities to zero instead produces two pulses in opposite directions, each with half the original displacement profile.

Match a measured reflection

A small junction connects strings under equal tension. A displacement pulse reflects with amplitude -0.20 times the incident amplitude. Find the ratio of impedances, the ratio of linear mass densities and the transmitted power fraction for a long sinusoidal train. Does the reflected displacement sign change when the source is moved to the other side?

▼ **Invert the matching equations**

Solving $r = (1 - z)/(1 + z)$ with $z = Z_2/Z_1$ gives

$$z = (1 - r)/(1 + r) = 1.50.$$

Equal tension implies

$$\mu_2/\mu_1 = z^2 = 2.25.$$

Four percent of the incident power reflects, so 96% transmits. Reversing incidence makes $r = +0.20$ and leaves the reflected power fraction at 0.04. A sign reversal records phase; a power measurement alone cannot distinguish positive and negative r.

Identify the boundary by spectroscopy

A string of length 0.75 m has speed 90 m/s. A vibration record contains adjacent peaks at 90 and 150 Hz. Compare the fixed–fixed and fixed–free spectra and predict one lower-frequency peak. Find how the peak spacing changes if the tension is increased by 44% without changing density.

▼ **Combine the mode conditions with the speed law**

For fixed–fixed ends the frequencies are multiples of 60 Hz, inconsistent with both readings. A fixed–free string has odd multiples of $c/(4L) = 30$ Hz, producing 30, 90 and 150 Hz. The predicted lower peak is 30 Hz if the drive and probe couple to it. Increasing tension by a factor of 1.44 multiplies speed and all frequencies by 1.20. The spacing between adjacent allowed modes rises from 60 to 72 Hz.

Predict which note disappears for an off-centre pluck

A string is released from a triangular displacement of height h at $x = L/3$. Obtain its coefficients by integrating the two constant slopes after one integration by parts. Compare the second and first modal energies and identify the first missing harmonic.

▼ **Project the two sides onto the sine modes**

For pluck position αL, the slopes are $h/(\alpha L)$ and $-h/[(1-\alpha)L]$. The coefficient is

$$b_n = 2h \sin(n\pi\alpha)/[\pi^2 n^2 \alpha(1-\alpha)].$$

At $\alpha = 1/3$ the first two sines are equal, so

$$b_2/b_1 = 1/4,$$

$$E_2/E_1 = 4(1/4)^2 = 1/4.$$

The third harmonic is absent because its sine vanishes at the pluck point. Moving the pluck slightly off that point restores it.

Revisit chapter 1 · Hear the initial shape in the harmonics of a string.

Recover a source speed from two microphones

A trolley carries a tone generator down a straight track in still air at $c = 340$ m/s. A stationary microphone ahead of it measures 714 Hz, and a stationary microphone behind it measures 646 Hz during the same steady run. Recover the trolley speed and emitted frequency without assuming a small Doppler shift. Would averaging the two readings recover the source frequency exactly?

▼ **Worked solution**

With speed v towards the front microphone, the measured frequencies are

$$f_+ = f_s c/(c - v),$$
$$f_- = f_s c/(c + v).$$

Taking their difference divided by their sum eliminates the unknown tone frequency. Thus

$$v = c(f_+ - f_-)/(f_+ + f_-) = 17.0\,\mathrm{m/s}.$$

Adding the reciprocal frequencies gives

$$1/f_+ + 1/f_- = 2/f_s.$$

The emitted frequency is their harmonic mean, $2f_+f_-/(f_+ + f_-) = 678.3$ Hz. The arithmetic mean is 680 Hz, high by 1.7 Hz. Substitution with $v/c = 0.05$ gives 678.3/0.95 = 714 Hz and 678.3/1.05 = 646 Hz.

Revisit chapter 1 · Read a frequency from moving wavefronts.

Explore these relations

- For the pulse direction, compare displacement and initial velocity and follow the separated travelling components. Open the live calculation.
- For the boundary inference, compare the standing-wave mode sequence. Open the live calculation.
- For the microphone observations, vary source speed and compare the approach and recession frequencies. Open the live calculation.

Chapter 3 · Wave packets and dispersion

Derive phase and group velocity from a pair of waves, then use a dispersion relation to predict how a finite packet moves and spreads.

FROM FIRST PRINCIPLES

Derive the speed of an envelope from two nearby waves

Consider two small waves in a linear medium, with equal amplitudes and nearby wavenumbers k_1, k_2. The medium supplies their frequencies through its dispersion relation $\omega(k)$. Linearity allows the displacements to add. Let overbars denote arithmetic means and let $\Delta k = k_2 - k_1$, $\Delta\omega = \omega_2 - \omega_1$. The sum of the two cosines factors into a rapidly oscillating carrier and a slowly varying multiplier.

$$y = 2A\cos[(\Delta k\, x - \Delta\omega\, t)/2]\cos(\bar{k}x - \bar{\omega}t).$$

To follow a particular carrier crest, keep its phase constant. To follow an envelope maximum, hold the phase of the first cosine fixed instead. Differentiating these two constant-phase conditions gives two speeds.

$$v_{\text{crest}} = \bar{\omega}/\bar{k},$$
$$v_{\text{envelope}} = \Delta\omega/\Delta k.$$

As the two waves approach each other in wavenumber, the ratio of differences becomes a derivative.

$$v_p = \omega/k,$$
$$v_g = d\omega/dk.$$

A narrow packet moves approximately at this group velocity when nearby components sample nearly the same derivative. Curvature of $\omega(k)$ makes different parts of its spectrum translate at slightly different speeds.

Background · Group velocity

$$v_g = d\omega/dk$$

This is the local slope of the dispersion curve. A sufficiently narrow-band packet with weak distortion can travel with its envelope at this velocity.

DERIVATION AND APPLICATIONS

Predict the exact width of a Gaussian envelope

The range of group velocities estimates when a packet starts to spread. For a Gaussian input one can calculate its width at every time and fix the numerical factor in that estimate. Assume a linear, lossless medium, a narrow spectrum around k_0 and negligible cubic and higher dispersion across that spectrum. At

$t = 0$ take the complex envelope $A(x, 0) = A_0 e^{-x^2/(2s^2)}$ with no initial spatial phase variation. Its wave-number spectrum is proportional to $e^{-s^2 q^2/2}$ for $q = k - k_0$.

$$\omega(k_0 + q) = \omega_0 + v_g q + \tfrac{1}{2} D q^2,$$

$$D = \left.\frac{d^2\omega}{dk^2}\right|_{k_0}.$$

After removing the carrier, the linear phase translates the packet to $\xi = x - v_g t$. The quadratic phase combines with its spectral amplitude to give a Gaussian with complex coefficient $s^2 + iDt$. Integrating each spatial Fourier component and fixing the prefactor by its initial value gives

$$A(\xi, t) = \frac{A_0 s}{\sqrt{s^2 + iDt}} \exp\left[-\frac{\xi^2}{2(s^2 + iDt)}\right].$$

A square-law detector measures the modulus squared. The real part of the reciprocal complex width determines the new spatial scale w, defined here as the distance from the centre to the $1/e$ intensity point.

$$|A|^2 = |A_0|^2 \frac{s}{w} e^{-\xi^2/w^2},$$

$$w = s\sqrt{1 + \left(\frac{Dt}{s^2}\right)^2}.$$

The product of peak intensity and width stays constant, as expected for lossless propagation. The sign of curvature affects the phase across the packet but drops out of this width measurement. A pre-existing spatial phase gradient varying across the envelope would change that conclusion about increasing width.

WORKED EXAMPLE

Separate attenuation from geometric spreading

Change the inputs and investigate this derivation.

A packet starts with $s = 0.500$ m in a medium with $D = -0.100$ m²/s. At 5.00 s, $Dt/s^2 = -2.00$, so its width is 1.12 m. The peak intensity then equals $1/\sqrt{5} = 0.447$ of its initial value. A receiver following the envelope could see this reduction even with no dissipation. Integrating intensity across the packet tests that interpretation.

SECOND APPLICATION

Infer local dispersion from a profile

Change the inputs and investigate this derivation.

For a second packet $s = 0.200$ m increases to 0.250 m after 0.600 s. Solving for curvature magnitude gives $|D| = (s^2/t)\sqrt{(w/s)^2 - 1} = 0.0500$ m²/s. An independently measured $v_g = 1.80$ m/s puts the centre 1.08 m from launch at this time. The centre displacement fixes the first derivative of the dispersion relation, while the width change fixes the magnitude of its second derivative.

Track a crest and an envelope separately

A burst contains many neighbouring frequencies. Individual crests can travel across its envelope as the burst moves past a detector. To see this with the simplest possible signal, superpose cosines with wave numbers $k_0 \pm \Delta k/2$ and angular frequencies $\omega_0 \pm \Delta\omega/2$. Applying the sum-to-product identity gives a rapid carrier $\cos(k_0 x - \omega_0 t)$ multiplied by $2\cos[(\Delta k\, x - \Delta\omega\, t)/2]$. A crest of the carrier moves at ω_0/k_0, while a maximum of the envelope moves at $\Delta\omega/\Delta k$.

As the two wave numbers approach one another, the envelope velocity tends to $v_g = d\omega/dk$. The phase velocity is $v_p = \omega/k$. These coincide for a straight dispersion relation through the origin, as on the ideal nondispersive string. They can differ when restoring physics, geometry or material response introduces another length or time scale. Their distinction comes from which observable is followed.

Move a localised packet

Describe the packet as an integral of linear modes with complex amplitudes concentrated near k_0. Write $k = k_0 + q$. The phase of each component is $k_0 x - \omega_0 t + q(x - v_g t) - \omega''(k_0) q^2 t/2$ to second order. The common first term is the carrier. Retaining only the linear term in q translates the original envelope without changing its shape, since all its Fourier components now depend on $x - v_g t$.

The quadratic term changes the relative phases within the envelope. A packet spanning a wave-number width Δk samples a range of group velocities of order $|\omega''|\Delta k$. After a time t the associated spatial separation is of order $|\omega''|\Delta k\, t$. When this approaches the initial packet width, a single arrival time ceases to describe its full shape. This estimate concerns a narrow band and small enough bandwidth for the local expansion to describe the dispersion relation. Dissipation can also distort a pulse by attenuating frequencies unequally.

Apply the argument to surface gravity waves

For small-amplitude waves on deep water, neglect surface tension and viscosity. Incompressible, initially irrotational motion admits a velocity potential Φ with $\nabla^2 \Phi = 0$. A wave that decays rapidly with depth $z < 0$ has a potential proportional to $e^{kz} e^{i(kx - \omega t)}$. At the surface the vertical velocity follows the elevation η through $\eta_t = \Phi_z$, while the linearised pressure condition from Bernoulli gives

$$\Phi_t + g\eta = 0.$$

Differentiating the latter and eliminating η gives $\Phi_{tt} + g\Phi_z = 0$ at $z = 0$. This yields

$$\omega^2 = gk.$$

Consequently $v_p = \sqrt{g/k}$ and $v_g = v_p/2$. The disturbance penetrates through a depth of order $1/k$. The assumption of deep water means bottom depth much greater than this scale. A finite-depth bottom alters the dispersion and therefore the time predicted for the arrival of a burst.

Predict an arrival and a change of width

Take a packet with carrier wavelength 4.0 m on deep water. Its $k_0 = 1.571$ m⁻¹ gives $\omega_0 = 3.93$ rad/s, $v_p = 2.50$ m/s and $v_g = 1.25$ m/s. An envelope maximum takes approximately 16.0 s to cover 20 m. A crest travels twice as fast, so it crosses the packet rather than remaining attached to the envelope maximum. Timing successive crests at one point would measure the carrier frequency, while timing envelope peaks at two points measures the group velocity.

For a spread $\Delta k = 0.20$ m⁻¹ the magnitude of the curvature is $|\omega''| = v_g/(2k_0) = 0.398$ m²/s. Over 16 s the different group velocities produce about 1.3 m of relative separation in this estimate. Comparing that with the initial envelope width is essential. Dispersion that is small in a broad pulse can be a major correction when the initial pulse is sharply localised.

An envelope measurement requires an operational timing definition. For a nearly undistorted narrowband packet, its maximum and its energy centroid travel almost together. Strong reshaping can move the maximum relative to the centroid, so arrival times defined by a threshold crossing can depend on amplitude as well as propagation. Saving the full waveform at two stations allows cross-correlation delays and centroid delays to be compared before assigning a group velocity.

Connect deep-water waves to shallow-water travel

When the water depth is not large compared with the wavelength, the bottom alters both crest speed and energy transport. For an inviscid, incompressible liquid of uniform depth h, assume a small-amplitude plane surface wave and irrotational flow. A velocity potential ϕ then satisfies Laplace's equation. Take the mean free surface at $z = 0$ and the flat impermeable bottom at $z = -h$. Horizontal dependence $e^{i(kx-\omega t)}$ makes the vertical solution proportional to $\cosh k(z+h)$, since its derivative must vanish at the bottom.

The free surface elevation η follows fluid vertically, so $\partial_t \eta = \partial_z \phi$ there. The linear dynamic condition from atmospheric pressure and gravity is

$$\partial_t \phi + g\eta = 0.$$

Differentiate the second condition in time, eliminate $\partial_t \eta$ with the first and insert the hyperbolic profile. The result and its group-velocity derivative are

$$\omega^2 = gk \tanh(kh),$$

$$\frac{v_g}{v_p} = \frac{1}{2}\left[1 + \frac{2kh}{\sinh(2kh)}\right].$$

In deep water $\tanh(kh) \to 1$ and the group speed is half the phase speed. In shallow water $\tanh(kh) \simeq kh$, giving $\omega \simeq k\sqrt{gh}$. Both speeds then approach $\sqrt{gh}$. The controlling small parameter is the depth measured against $1/k$, so even the same tank can appear deep to short ripples and shallow to longer waves.

Use an arrival time to test the depth approximation

Take wavelength 4.00 m and tank depth 0.500 m. Then $k = 1.571$ m⁻¹, $kh = 0.785$ and $\tanh(kh) = 0.656$. The angular frequency is

$$\sqrt{9.81(1.571)(0.656)} = 3.18\,\mathrm{rad/s}.$$

Phase speed is 2.02 m/s and the ratio formula gives $v_g/v_p = [1 + 1.571/\sinh(1.571)]/2 = 0.841$, hence group speed 1.70 m/s. A narrow packet takes 11.8 s to travel 20.0 m. The shallow-water estimate would give 2.21 m/s and a 9.03 s arrival, a difference readily observable with two surface gauges.

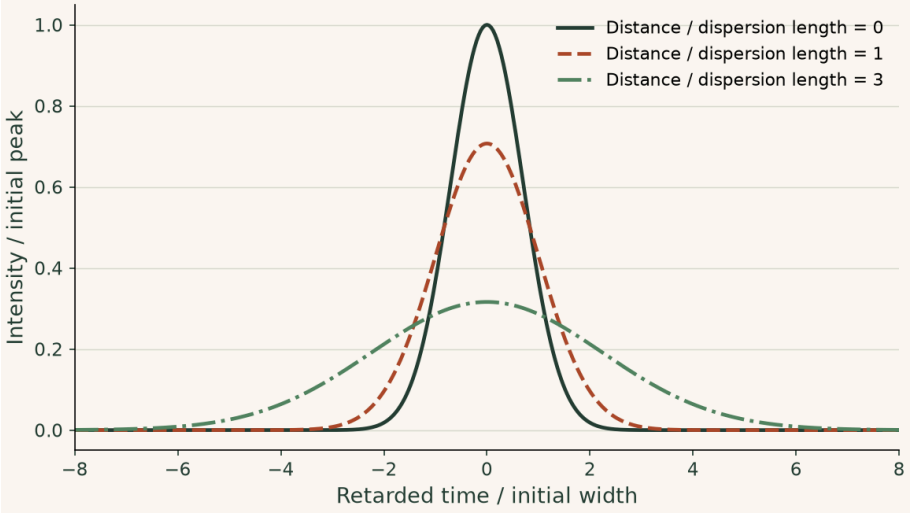
The calculation holds the wavelength fixed when comparing model predictions. A wave entering a different depth from a steady source instead retains its frequency, and its wavenumber must be found again using the local dispersion relation. Keeping both wavelength and frequency fixed across a depth change would overconstrain the problem. If depth varies gradually, timing a packet across the tank involves integrating the inverse local group speed. Sharp bottom steps introduce reflection, so a single travel-time calculation cannot then predict the transmitted amplitude.

The shape of the subsurface motion gives another depth test independent of packet timing. For the potential derived above, horizontal velocity amplitude at the bottom is reduced relative to the free surface by $1/\cosh(kh)$. At the stated depth this ratio is about 0.755: substantial horizontal motion remains even at the bottom. At much larger depths the ratio becomes exponentially small. Tracers placed at several depths therefore distinguish bottom influence directly, although viscous effects right at a real solid wall can alter the potential-flow velocity in a thin boundary layer.

The bottom must be flat on the scale of the wavelength for a single standing vertical profile to describe the whole observation region.

Can a pulse broaden without losing energy?

Propagate a transform-limited Gaussian envelope with second-order dispersion in a lossless medium. In retarded time, the intensity width increases by $b = \sqrt{1 + (z/L_D)^2}$. The peak must fall as $1/b$ if the integrated pulse energy is conserved.



At three dispersion lengths the width is √10 times its initial value and the peak is 31.6% as high. Integrating the full broadened Gaussian yields the original energy. A fixed, short detector gate could report an apparent loss.

Try it · Pulse width

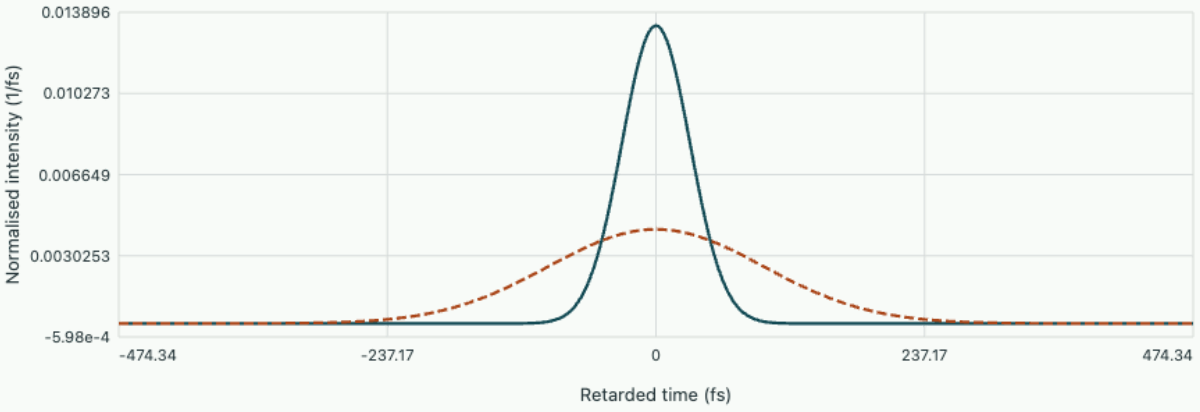
Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Distance / dispersion length 3

Initial pulse rms duration (fs) 30



Initial pulse Dispersed pulse

Rms duration 94.868 fs. Full pulse energy is conserved under quadratic phase dispersion; a finite detector gate can lose part of the broadened pulse.

▶ **Read values from the plot**

Study the full derivation

Dispersive pulse compression and a detector's view of a tilted pulse · Optical fields, instruments and observations

An ultrashort pulse can broaden because its spectral components acquire different group delays or because different parts of its beam arrive at different times. Derive the two mechanisms separately, predict when added glass can shorten an already chirped pulse, and decide which measurements distinguish them.

Chapter 4 · Optical paths, lenses and images

Use stationary optical path length and the paraxial approximation to predict refraction, focal length and the image formed on a screen.

FROM FIRST PRINCIPLES

Obtain refraction by varying the interface crossing point

In geometrical optics a ray between two specified points makes its travel time stationary against small changes of path. Take a flat interface between homogeneous transparent media with refractive indices n_1, n_2, and light speed c/n_j. Put the source a perpendicular distance h_1 from the interface and the receiver a distance h_2 beyond it. Their horizontal separation is D. If the ray crosses at coordinate x measured from the source projection, its time is

$$t(x) = \frac{n_1}{c}\sqrt{x^2 + h_1^2} + \frac{n_2}{c}\sqrt{(D-x)^2 + h_2^2}.$$

Moving the crossing point increases one segment and decreases the other. At a stationary path their first-order contributions to travel time cancel.

$$c\,\frac{dt}{dx} = \frac{n_1 x}{\sqrt{x^2 + h_1^2}} - \frac{n_2(D-x)}{\sqrt{(D-x)^2 + h_2^2}} = 0.$$

The geometric fractions are the sines of the ray angles to the surface normal. Hence $n_1 \sin\theta_1 = n_2 \sin\theta_2$. A curved optical surface applies this same local relation using its own normal at each crossing point, redirecting a family of rays towards an image.

Background · Optical path length

$$\mathcal{L} = \int n\, ds$$

Differences of this quantity determine phase delays in geometrical optics. For a pulse in a dispersive medium one must also account for group delay.

DERIVATION AND APPLICATIONS

Calculate the apparent depth of a submerged point

A submerged mark appears closer to a flat water surface when viewed from air. How accurately does a single apparent depth describe rays reaching an eye with a wide pupil? Consider a point at depth h in a clear medium of refractive index n_1. A plane boundary separates it from a second medium with index n_2. Ignore

diffraction and boundary curvature. A ray reaches the surface a horizontal distance x from the vertical through the mark. Its incidence angle θ_1 and exit angle θ_2 are measured from the normal.

$$x = h \tan\theta_1,$$
$$n_1 \sin\theta_1 = n_2 \sin\theta_2.$$

Extend the emerging straight ray backwards until it crosses the original vertical. The crossing lies at a depth h_{app} which depends on the ray chosen. Eliminating the horizontal distance and then using Snell's law shows how the cosines correct the familiar small-angle result.

$$h_{\text{app}} = \frac{x}{\tan\theta_2} = h\frac{\tan\theta_1}{\tan\theta_2} = h\frac{n_2}{n_1}\frac{\cos\theta_2}{\cos\theta_1}.$$

For a small aperture, both cosines approach unity. To see the first error term let $m = n_1/n_2$ and expand around normal incidence, measuring angles in radians.

$$h_{\text{app}} = \frac{h}{m}\left[1 - \frac{m^2 - 1}{2}\theta_1^2 + O(\theta_1^4)\right].$$

For observation from a lower-index medium, rays farther from the normal cross at smaller apparent depths. As incidence approaches its critical angle, the emerging ray grazes the surface and the inferred crossing approaches the interface. A wide bundle therefore has a range of apparent axial origins.

WORKED EXAMPLE

Focus on a mark below water

Change the inputs and investigate this derivation.

A mark lies 120 mm below water with $n_1 = 4/3$ and $n_2 = 1$. Near normal incidence the apparent depth is 90.0 mm. A ray entering the interface at 30.0° emerges with $\sin\theta_2 = 2/3$, so $\cos\theta_2 = \sqrt{5}/3$. Its crossing is at $90.0(0.7454/0.8660) = 77.5$ mm. Restricting the rays admitted by the viewing optics therefore changes the focus of this mark.

SECOND APPLICATION

Recover a refractive index with a narrow pupil

Change the inputs and investigate this derivation.

The lower surface of a 15.0 mm thick transparent slab appears 9.50 mm below its top surface in a near-normal air observation. The slab index is approximately $15.0/9.50 = 1.58$. An internal incidence angle of 5.00° then gives a relative depth correction of $-(1.58^2 - 1)(0.0873)^2/2 = -0.0057$, or about 0.054 mm. This estimates the aperture-induced error against which the depth gauge resolution should be compared.

Try it · Apparent depth

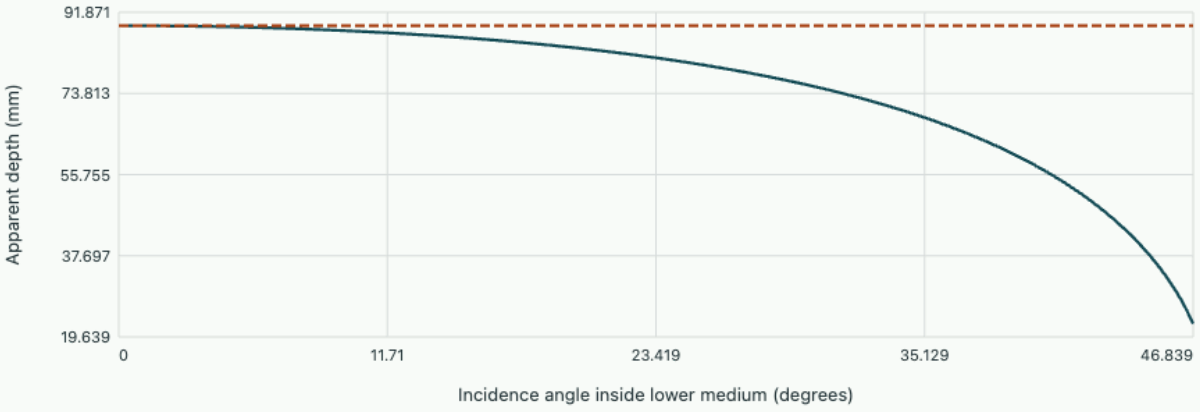
Vary the physical parameters in the derived relation. Use the units and assumptions in the readout to connect the curves to the two worked examples. A parameter study can test the forward prediction; the same relation can then be inverted to infer a parameter from a measurement.

Open this live calculation

Try the calculation

Lower-medium refractive index **1.35**

Physical depth (mm) **120**



Back-projected emergent ray Paraxial limit

Upper medium is air. Different oblique rays back-project to different depths; the paraxial apparent depth 88.889 mm applies to a narrow pupil.

Travel time chooses a ray

A pulse travels from a point in air to a point in glass through a plane interface. A straight line minimises geometrical distance, but the pulse spends different times per metre in the two media. For a narrow-band wave in transparent, effectively nondispersive media, optical path length is $\int n\,ds$. The ray connecting two fixed endpoints makes that integral stationary under small path variations. This description requires that the refractive index vary slowly on the scale of a wavelength within each material region and treats interfaces separately.

Let the crossing point slide horizontally by dx. If the angles relative to the normal are θ_1, θ_2, the length of the first segment changes by $\sin\theta_1\,dx$ and that of the second by $-\sin\theta_2\,dx$. Setting the first variation of optical path length to zero gives

$$n_1 \sin\theta_1 = n_2 \sin\theta_2.$$

The same condition follows from tangential phase matching. Stationarity gives the ray direction; the reflected and transmitted powers require the wave boundary conditions as well.

Follow field amplitude through the critical angle

Snell's law determines transmitted direction by matching the tangential phase across an interface. It does not determine how much energy is reflected. For light incident from isotropic nonmagnetic index n_1 onto index n_2, take the electric field perpendicular to the plane of incidence. Continuity of tangential electric and magnetic fields gives

$$1 + r_s = t_s,$$

$$n_1 \cos\theta_i (1 - r_s) = n_2 \cos\theta_t \, t_s.$$

All amplitudes refer to the same tangential electric direction. Eliminating t_s yields

$$r_s = \frac{n_1 \cos\theta_i - n_2 \cos\theta_t}{n_1 \cos\theta_i + n_2 \cos\theta_t},$$

$$T_s = \frac{n_2 \cos\theta_t}{n_1 \cos\theta_i} |t_s|^2$$

for propagating transmission in lossless media. The flux factor is essential: an electric amplitude larger than one need not imply transmitted power larger than the incoming power. At normal incidence from air into glass of index 1.50, $r = -0.200$, $t = 0.800$, $R = 0.0400$ and $T = 0.960$.

Now send 500 nm light from the glass toward air at 45°. Matching the tangential wavenumber gives

$$k_\parallel = k_0 n_1 \sin\theta_i.$$

The outgoing normal wavenumber obeys $k_z^2 = k_0^2 n_2^2 - k_\parallel^2$. Above the critical angle 41.8°, choose $k_z = i\kappa$ so fields decay into air. Here $\kappa / k_0 = \sqrt{1.50^2 \sin^2 45^\circ - 1} = 0.354$, giving amplitude decay length $1/\kappa = 225$ nm and intensity decay length 113 nm. The reflection coefficient has unit magnitude and a phase shift. A single evanescent field in the lossless half-space carries no mean power normal to the interface even though its local electric amplitude is nonzero.

Practise on sheet 2 · Distinguish transmission amplitude, flux and penetration. Further reading · Delft University of Technology · Interactive Optics · Basic Electromagnetic and Wave Optics, "Properties of the Fresnel Coefficients" and "Total Internal Reflection and Evanescent Waves".

Approximate a spherical refracting surface

Take light travelling from left to right. Put the surface vertex at zero and its centre of curvature at distance R to the right for a convex interface. A near-axis ray reaching height y from a real on-axis object a distance s to the left makes a small angle y/s to the axis. The normal through the centre makes angle approximately $-y/R$. For a refracted ray directed to an image s' to the right, its angle is approximately $-y/s'$. Small-angle Snell refraction applied to the signed differences between ray and normal gives

$$\frac{n_1}{s} + \frac{n_2}{s'} = \frac{n_2 - n_1}{R}.$$

This sign convention takes positive object distance on the incoming side and positive image distance on the outgoing side. A negative solution for s' means that rays diverge after refraction but their backwards extensions intersect on the incoming side. The algebra therefore distinguishes a real screen image from a virtual image.

Combine two surfaces

For a thin lens in air, treat the image from the first surface as the object for the second. Since their vertices are taken at the same axial position, these intermediate signed distances have opposite signs. Adding the two surface equations eliminates the intermediate image term, leaving

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f},$$

$$\frac{1}{f} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right).$$

A symmetric biconvex glass lens has $R_1 > 0$ and $R_2 < 0$, hence positive focal length. Rays from a distant object then converge approximately one focal length beyond the lens. The formula assumes thin geometry and paraxial rays. Wide beams can generate spherical aberration, and index dispersion makes the focal length wavelength dependent. Even a perfect paraxial lens forms a diffraction spot when its pupil is finite.

Place a detector and interpret a virtual image

A lens has $n = 1.50$, $R_1 = +0.20$ m and $R_2 = -0.20$ m, giving $f = 0.20$ m. An object at $s = 0.60$ m gives $1/s' = 5.00 - 1.67 = 3.33$ m⁻¹, hence a real image at 0.30 m. The central undeviated ray gives transverse magnification $-s'/s = -0.50$. A 12 mm tall object therefore produces a 6 mm inverted image. In practice a screen scan locates its sharpest plane, and the measured magnification independently checks the distance convention.

Move the object to 0.15 m. The same lens now gives $s' = -0.60$ m and magnification $+4.0$. The eye behind the lens sees rays diverging as if from an upright image on the object side, while a screen placed 0.60 m behind the lens finds no sharp real image. As the object passes through the focal plane, the sign changes through infinite image distance. This limiting case is readily checked using a collimated output beam.

A real image can be tested by moving a diffuse screen through its predicted location and recording the sharpness. For a virtual image, extend at least two emerging ray directions backwards to locate their apparent origin. An observer’s eye lens can still create a real retinal image from those diverging rays.

Treat two separated lenses as a single optical system

A compound objective can have an effective focal length different from its back focal distance. Work in air with paraxial rays and two centred thin lenses separated by distance d. Describe a ray by height y and small

angle u relative to the optical axis. Free propagation sends (y, u) to $(y + du, u)$ and a lens of focal length f sends it to $(y, u - y/f)$. These operations are linear, so their matrices can be multiplied in the order the ray encounters them.

For the first lens f_1, a spacing d and second lens f_2, the ray immediately after the second lens is $M(y, u)$ with

$$M = \begin{pmatrix} 1 - d/f_1 & d \\ -[1/f_1 + 1/f_2 - d/(f_1 f_2)] & 1 - d/f_2 \end{pmatrix}.$$

For an initially parallel ray, its final angle is Cy where C is the lower-left entry. Define effective focal length $f_{\text{eff}} = -1/C$. Its ray height at the second lens is Ay, where $A = 1 - d/f_1$. After further distance s the height is $(A + sC)y$, so the back focal distance measured from the second lens is $-A/C = A f_{\text{eff}}$. Conflating the effective focal length with that physically measured clearance can place a detector incorrectly.

Locate the sensor in a compact two-lens assembly

Let $f_1 = 100$ mm, $f_2 = -50.0$ mm and $d = 75.0$ mm. The total optical power is $0.0100 - 0.0200 + 0.0150 = 0.00500$ mm⁻¹, giving $f_{\text{eff}} = 200$ mm. The upper-left matrix entry is 0.250, so the back focal distance is only 50.0 mm. The sensor lies 125 mm behind the first lens even though the angular image scale corresponds to a 200 mm lens.

Trace a parallel ray entering 2.00 mm above the axis. The first lens sets its angle to -0.0200 rad. Propagation over 75.0 mm reduces height to 0.500 mm. The diverging lens adds $+0.0100$ rad, leaving -0.0100 rad, so after another 50.0 mm the ray reaches the axis. If light instead arrives at a small field angle, focal-plane image displacement is $f_{\text{eff}} u$ in this ideal system: at $u = 1.00$ mrad that is 0.200 mm. In the zero-spacing limit powers add directly and the two focal distances coincide. When the total power instead tends to zero, a parallel incident beam exits parallel and no finite back focus exists.

For a further alignment check, reverse the propagation direction so that the diverging lens is encountered first. The equivalent focal length remains 200 mm because the expression for total power is symmetric in the two lenses. The back focal distance, however, now uses $A = 1 - d/f_2 = 2.5$, giving 500 mm after the converging lens. A pair of lenses can therefore preserve its overall power when reversed while placing a collimated focus at a very different distance from the final glass surface. A measurement locating the two external focal points identifies this front–back asymmetry without having to inspect the lens curvatures.

Study the full derivation

From a slowly varying wave to a curved ray in a graded index · Optical fields, instruments and observations

A ray bends continuously in a glass whose index changes across its width. Derive where that ray equation comes from, predict the focusing length of a graded-index rod, and locate the points at which a smooth ray description ceases to determine an optical amplitude.

Problem sheet 2 · Packets and paraxial imaging

Use Lectures 3 and 4 to distinguish phase measurements, arrival-time measurements and image positions.

Extract two speeds from one dispersion curve

Near $k_0 = 10$ m⁻¹ a guide has $\omega(k) = 120 + 8(k - k_0) + 0.50(k - k_0)^2$ in consistent SI units. Find phase velocity, group velocity and the estimated spread of group velocity across a full wave-number band of 2 m⁻¹. Predict the carrier and envelope travel times across 24 m.

▼ **Use the value, slope and curvature**

At the centre $v_p = 120/10 = 12$ m/s, while differentiating gives $v_g = 8$ m/s and $\omega'' = 1.0$ m²/s. The group velocities at $k_0 \pm 1$ are 7 and 9 m/s, spanning 2 m/s. The phase travel time is 2.0 s and the envelope-centre estimate is 3.0 s. Over that time velocities at the band edges build up a separation of about 6 m; an initially very narrow packet would be noticeably distorted.

Place an object and a screen

A 150 mm thin converging lens must project a 10 mm object as an inverted 30 mm image on a screen. Find both lens distances, then determine how far the lens itself can move to a second sharp position if the object and screen stay fixed.

▼ **Solve the conjugate distances**

The magnification magnitude gives $s' = 3s$. Substitution in $1/f = 1/s + 1/(3s)$ gives $s = 4f/3 = 200$ mm and $s' = 600$ mm. Object and screen are 800 mm apart. Interchanging the two distances is a second solution because the lens equation is symmetric, giving a lens motion of 400 mm. The new image magnification is $-1/3$. A check of both sharpness and image size distinguishes which solution was reached.

Estimate a colour focus shift

A symmetric biconvex lens has $R_1 = +0.12$ m and $R_2 = -0.12$ m. Its red and blue refractive indices are 1.50 and 1.52. Locate both foci for a distant object and the axial shift, then explain why a ray calculation of two foci does not by itself give the size of a white-light spot.

▼ **Use dispersion in the lensmaker equation**

For the symmetric lens $f = R/[2(n - 1)]$. Red focuses at 0.120 m and blue at 0.1154 m, about 4.62 mm closer to the lens. The refractive index is larger for blue, increasing optical power. To predict a spot on one fixed screen one also needs the illuminated pupil, spectral weighting, aberrations and diffraction. A focal shift is a longitudinal distance; the observed blur is transverse.

Distinguish transmission amplitude, flux and penetration

At normal incidence send light from index 1.60 into air. Find the electric amplitude and transmitted power fraction. At 50.0° incidence and vacuum wavelength 640 nm, find how far one must move into air for the squared field to fall to a quarter of its surface value.

▼ **Apply the two boundary regimes**

At normal incidence $t = 2(1.60)/(1.60+1) = 1.231$ and $T = (1/1.60)t^2 = 0.947$. Reflection supplies the remaining 0.0533. At the oblique angle $\kappa = (2\pi/640\,\mathrm{nm})\sqrt{1.60^2\sin^2 50^\circ - 1} = 0.00696$ nm⁻¹. The squared-field ratio is $e^{-2\kappa z}$, hence $z = \ln 2/\kappa \simeq 99.6$ nm. A Fresnel formula for real transmitted ray angles is inapplicable in this regime.

Revisit chapter 4 · Follow field amplitude through the critical angle.

Explore these relations

- For dispersion, vary propagation distance and compare pulse broadening with the initial envelope; the plotted model uses quadratic dispersion. Open the live calculation.
- For interface transmission, compare field amplitude, energy flux and evanescent decay. Open the live calculation.

Chapter 5 · Polarisation and phase retardation

Predict the output of a polariser–retarder arrangement and distinguish changes of polarisation from loss of beam power.

FROM FIRST PRINCIPLES

Project a polarised field onto an analyser

A linear polariser transmits the electric-field component along its transmission axis. Assume an ideal lossless transmitted axis, complete rejection of the perpendicular component and a monochromatic linearly polarised incident wave. For incident field amplitude E_0 at angle θ to the analyser axis $\mathbf{a}$, the emerging field is the projection

$$\mathbf{E}_{\text{out}} = (\mathbf{E}_{\text{in}} \cdot \mathbf{a})\mathbf{a} = E_0 \cos\theta\, \mathbf{a}.$$

A detector averages energy flux over many optical cycles. For equal propagation impedances, intensity is proportional to squared field amplitude. This gives Malus's law and predicts extinction at crossed linear axes.

$$I_{\text{out}} = I_{\text{in}} \cos^2\theta.$$

A retarder takes a different action on its two orthogonal components. Both propagate through the crystal thickness d with different phase wavenumbers $2\pi n_f/\lambda_0$ and $2\pi n_s/\lambda_0$. Their relative phase on emerging is therefore

$$\delta = \frac{2\pi d}{\lambda_0}(n_s - n_f).$$

The analyser can turn this phase difference into a measurable intensity because it projects both coherent components onto one axis before the detector squares their sum.

DERIVATION AND APPLICATIONS

Infer a plate's retardance from a rotation scan

A plate between crossed polarisers looks dark whenever an optical axis coincides with the first polariser. Away from those angles, both its orientation and its retardance control the transmitted light. Assume ideal polarisers, a uniform lossless linear retarder, normal incidence and monochromatic illumination. Let the first polariser transmit x, the analyser transmit y and the fast axis of the plate make angle θ with x. Write the input field magnitude immediately after the first polariser as E_0. Projection into the fast and slow axes gives

$$E_f = E_0 \cos\theta,$$
$$E_s = -E_0 \sin\theta.$$

The slow component emerges with relative phase δ in the chosen complex convention. Its projection onto y has coefficient $\cos\theta$, while the fast component projects with $\sin\theta$. The two contributions need to be added

as complex fields before computing detector intensity.

$$E_y = E_0 \sin\theta \cos\theta (1 - e^{i\delta}).$$

Use $|1 - e^{i\delta}|^2 = 4\sin^2(\delta/2)$ and normalise to the intensity I_0 after the first polariser. Expressing the angular factor through a quadruple angle identifies an observable modulation at four times the rotation frequency.

$$\frac{I_\perp}{I_0} = \sin^2(2\theta)\sin^2(\delta/2) = \frac{1 - \cos 4\theta}{2}\sin^2(\delta/2).$$

The scan maximum provides $\sin^2(\delta/2)$ independently of where the instrument labels zero angle, provided the full rotation is sampled. Lossless plates that differ in retardance by a full wave give identical readings, and this intensity measurement alone also loses the sign of the phase delay.

WORKED EXAMPLE

Determine a birefringent film thickness

Change the inputs and investigate this derivation.

At 600 nm the measured maximum is 0.250 of I_0. Independent fabrication information places δ between zero and π, so

$$\delta = 2\arcsin(0.500) = \pi/3.$$

With index difference 0.0100 and $\delta = 2\pi\Delta n\, d/\lambda$, the thickness is $d = 10.0$ µm. At a 15.0° axis angle the predicted intensity is $0.250\sin^2(30.0^\circ) = 0.0625$ of I_0.

SECOND APPLICATION

Diagnose angular misalignment at a nominal extinction

Change the inputs and investigate this derivation.

A calibrated half-wave plate gives 0.0100 of I_0 at a setting labelled parallel to the input polariser. If angular misalignment dominates, its magnitude near the expected minimum solves $\sin^2(2\theta) = 0.0100$. Thus

$$|\theta| = \arcsin(0.100)/2 = 2.87^\circ.$$

A full rotation tests that explanation: the same angular offset should shift all four minima, and each corrected minimum should reach the independently measured background.

Resolve the electric field into two components

A narrow-band plane wave propagating along z in a transparent medium has a transverse electric field. Write its physical value as the real part of $(A_x\hat{\mathbf{x}} + A_y\hat{\mathbf{y}})e^{i(kz-\omega t)}$. A common phase attached to both complex

amplitudes shifts the time origin of the field oscillation. Their relative phase and amplitude ratio determine how the tip of the field vector moves within a cycle.

For amplitudes in phase the field traces a line. Equal magnitudes with relative phase $\pi/2$ produce a circle, while intermediate combinations usually give an ellipse. A two-component Jones vector describes one fully polarised coherent beam. For randomly changing amplitudes, predictions require averaging intensities or field correlations over the exposure. A single Jones vector cannot represent every partially polarised state.

Project onto a measuring axis

An ideal linear polariser with transmission direction $\hat{\mathbf{a}} = (\cos\alpha, \sin\alpha)$ passes the component $\hat{\mathbf{a}} \cdot \mathbf{A}$ and emits a field along $\hat{\mathbf{a}}$. Its matrix is the outer product $\hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathsf{T}}$. A linearly polarised input at angle β therefore transmits a fraction $\cos^2(\beta - \alpha)$ of the intensity. This is Malus's law. It applies to the projection of the field before the detector takes a squared magnitude.

Two ideal polarisers at right angles transmit zero for an input aligned with the first. Inserting a third between them at 45° can increase transmitted power. After the middle polariser the direction of the transmitted wave has changed. The first projection passes half the original power and the second passes half of that half, so one quarter reaches the detector. The inserted component changes the state incident on the last analyser.

Background · Can a rotating linear analyser determine handedness?

Take two normalised Jones vectors $(1, i)/\sqrt{2}$ and $(1, -i)/\sqrt{2}$. Their electric vectors rotate in opposite senses with time. An ideal linear analyser responds to a real projection direction $(\cos\alpha, \sin\alpha)$. The two squared projections are identical for every angle.

$$\frac{I_{\pm}(\alpha)}{I_0} = \frac{1}{2}|\cos\alpha \pm i\sin\alpha|^2 = \frac{1}{2}.$$

A quarter-wave plate that delays the second component by $\pi/2$ first maps the two inputs to $(1, -1)/\sqrt{2}$ and $(1, 1)/\sqrt{2}$ in the chosen phasor convention. A linear analyser at 45° can now distinguish them. An unpolarised beam also gives a flat angular scan through a linear analyser, so constant throughput on its own cannot establish circular polarisation.

Transform a phase delay into an intensity reading

A lossless birefringent plate has two orthogonal principal axes with refractive indices n_f and n_s. Through thickness d, their phase difference is $\delta = 2\pi(n_s - n_f)d/\lambda_0$, where λ_0 is vacuum wavelength. In the plate-axis basis its Jones matrix, apart from a common phase, is $\mathrm{diag}(1, e^{i\delta})$. The matrix preserves the sum of squared amplitudes. Only the relative phase changes.

Let the incident linear polarisation make an angle β to the fast axis. An analyser crossed with the original input has unit vector $(-\sin\beta, \cos\beta)$ in that basis. Projecting the output onto it gives amplitude proportional to $\sin\beta\cos\beta(e^{i\delta} - 1)$. Squaring yields

$$I_{\perp}/I_{\mathrm{in}} = \sin^2(2\beta)\sin^2(\delta/2).$$

Thus a dark crossed-analyser setting can indicate an axis-aligned input or an integer number of full-wave delays. One dark reading alone does not identify thickness or birefringence.

Distinguish a quarter-wave plate from a half-wave plate

Take $\lambda_0 = 600$ nm and $\Delta n = 0.010$. The thinnest plate with $\delta = \pi/2$ has $d = \lambda_0/(4\Delta n) = 15.0$ µm. At $\beta = 45^\circ$, equal input amplitudes acquire a relative quarter-cycle delay, producing circular polarisation. A rotating ideal analyser then transmits half the intensity at every angle. With 4.0 mW arriving at the plate, the analyser passes 2.0 mW throughout the rotation.

Double the thickness. The delay becomes π, so the two components have opposite signs and the output is linearly polarised along the original crossed analyser. It transmits the full 4.0 mW in that setting and zero in its orthogonal setting. Flat response under rotation does not by itself distinguish circularly polarised from unpolarised light. Adding a known quarter-wave plate before the analyser converts circular light to a linear state but leaves ideal unpolarised light unpolarised, giving an independent measurement.

Rotating the input polarization relative to a plate also checks the orientation of its principal axes. If the input is aligned with either axis, only one component is present and retardation has no effect on the polarization ellipse. The phase remains physically detectable only through interference with an independent reference. For comparable input components both axes contribute and an analyser becomes sensitive to their relative phase. This is why a small alignment error can suppress a retardance measurement even for a perfectly fabricated plate.

Recover polarization from six intensity readings

A rotating linear analyser measures orientation and some of the shape of a polarization ellipse, but opposite handedness gives the same result at every linear angle. Let the complex field components be E_x, E_y and average products over the detector integration time. Choose intensity units that absorb the common medium factor. Define $S_0 = \langle |E_x|^2 + |E_y|^2 \rangle$, $S_1 = \langle |E_x|^2 - |E_y|^2 \rangle$,

$$S_2 = 2\mathrm{Re}\langle E_x^* E_y \rangle,$$

$$S_3 = 2\mathrm{Im}\langle E_x^* E_y \rangle.$$

The sign convention for the last component specifies the handedness label used throughout the measurement.

An ideal linear analyser at angle θ projects onto $(\cos\theta, \sin\theta)$. Squaring the projected amplitude and using double-angle identities gives

$$I(\theta) = \tfrac{1}{2}[S_0 + S_1\cos(2\theta) + S_2\sin(2\theta)].$$

Hence the difference of horizontal and vertical readings gives S_1, while the difference between $+45^\circ$ and -45° gives S_2. Finally pass the input through a calibrated quarter-wave analyser combination selecting the two circular basis vectors $(1, \pm i)/\sqrt{2}$. The difference of their outputs in this order gives S_3. Their sum, like the sum of either linear pair, gives S_0. Multiple estimates of total intensity check for source drift and detector gain imbalance.

Distinguish a partially polarized state from a pure ellipse

Suppose the two principal linear readings are 7.00 and 3.00 mW, the diagonal readings are 6.50 and 3.50 mW, and the circular readings are 8.00 and 2.00 mW. Each pair sums to 10.0 mW. The inferred Stokes parameters are $(10.0, 4.00, 3.00, 6.00)$ mW. The degree of polarization is

$$\sqrt{S_1^2 + S_2^2 + S_3^2}/S_0 = \sqrt{61}/10 = 0.781.$$

A single fixed Jones vector has degree one, so a pure elliptical state cannot reproduce these repeated intensity readings.

The polarized fraction has major-axis azimuth $\psi = \frac{1}{2}\operatorname{atan2}(S_2, S_1) = 18.4^\circ$. Its signed ellipticity angle satisfies $\sin(2\chi) = S_3/\sqrt{S_1^2 + S_2^2 + S_3^2}$, giving $\chi = 25.1^\circ$ and minor-to-major axis magnitude $\tan|\chi| = 0.468$. At zero polarized fraction both angles cease to describe a measurable ellipse. A fitted degree above one signals an inconsistent set of readings, since the correlation matrix must have nonnegative eigenvalues. Sequential readings should therefore include a reference power monitor when source brightness can change during the six settings.

One can predict readings at analyser settings never used in reconstruction. For a linear analyser at 30° relative to horizontal, the stated parameters give

$$[10 + 4\cos 60^\circ + 3\sin 60^\circ]/2 = 7.30\,\mathrm{mW}.$$

A separate fourth linear orientation therefore validates the fitted linear components but has no sensitivity to the circular component. To test handedness, interchange the two calibrated circular analyser states. The difference should reverse while their sum remains constant. If different linearly polarized calibration standards all generate a large circular signal, an analyser retardance or alignment error is a more immediate explanation than an exotic input state.

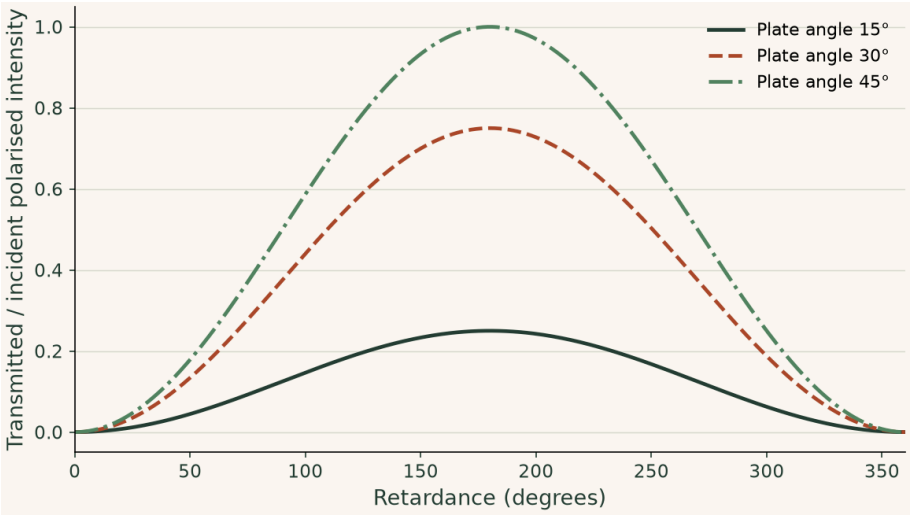
The total intensity is itself useful for another check. An ideal lossless wave plate changes the three polarized components while preserving their length and the total intensity. Measuring before and after a known plate therefore distinguishes a true change of polarization fraction from a simple rotation of its polarization coordinates.

Read retardance between crossed polarisers

A linear retarder sits between crossed ideal polarisers. At plate angle θ its two principal waves acquire relative phase δ; projection onto the final analyser gives

$$I/I_0 = \sin^2(2\theta)\sin^2(\delta/2).$$

Here I_0 is the intensity after the first polariser.



With the plate at 45°, a quarter-wave retardance transmits 1/2 and a half-wave retardance transmits all the incident polarised intensity. A single intensity does not uniquely determine retardance over multiple periods.

Try it · Retarder

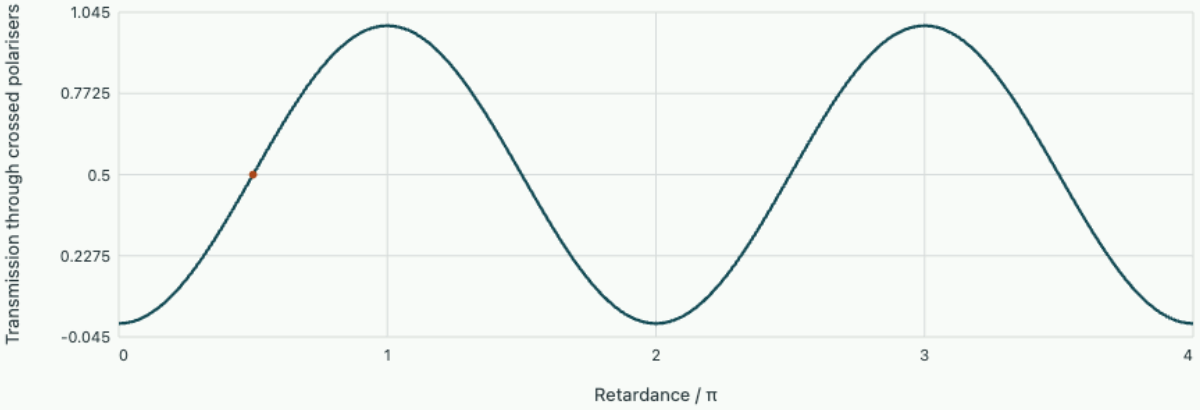
Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Plate-axis angle (degrees) **45**

Selected retardance / π **0.5**



Lossless retarder and analyser Selected retardance

Selected transmitted fraction 0.5. Intensity alone repeats every 2π of retardance and does not uniquely identify its order.

▶ **Read values from the plot**

Study the full derivation

Which polarisation measurements identify an optical element? · Optical fields, instruments and observations

A retarder rotates the state of a coherent beam, while unresolved mixtures can reduce its measured degree of polarisation. Derive the passage from Jones fields to a Mueller matrix, reconstruct a simple instrument from measured intensities, and identify claims the available input states cannot support.

Chapter 6 · Coherence and interferometry

Use a variable path delay to connect fringe visibility to a source spectrum and estimate how far two copies of a wave can remain correlated.

FROM FIRST PRINCIPLES

Derive fringe visibility from the field correlation

Split stationary light into two arms and recombine them in the same polarisation and spatial mode. Let E_1, E_2 denote complex optical fields with detector averaging over many periods, and scale units so intensity equals $\langle |E|^2 \rangle$. A controllable relative phase ϕ changes the interference term while the individual intensities I_1, I_2 stay fixed. Expanding the squared sum gives

$$I(\phi) = I_1 + I_2 + 2\,\mathrm{Re}[e^{i\phi}\langle E_1^* E_2 \rangle].$$

The largest and smallest readings occur when the phased correlation points along the positive or negative real axis. Normalise the difference of those readings by their sum to define visibility $\mathcal{V}$. Normalising the correlation by the geometric mean of the arm intensities defines the degree of coherence γ.

$$\gamma = \frac{\langle E_1^* E_2 \rangle}{\sqrt{I_1 I_2}},$$

$$\mathcal{V} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}|\gamma|.$$

Equalising the arms makes visibility directly measure $|\gamma|$. A relative delay then reveals how long a field retains correlation with a delayed copy. Particular spectra lead to different delay dependences.

Background · Fringe visibility

$$\mathcal{V} = (I_{\max} - I_{\min})/(I_{\max} + I_{\min})$$

Use maxima and minima from the same scan. Unequal intensities can reduce contrast in a fully coherent field.

DERIVATION AND APPLICATIONS

Find the interferogram of a rectangular spectral band

An optical filter with a flat passband and sharp edges predicts a distinctive pattern of coherence zeros. Consider stationary light whose spectral power per unit frequency is constant between $\nu_0 - \Delta\nu/2$ and $\nu_0 + \Delta\nu/2$. Combine two equal-intensity copies with adjustable delay τ and ignore polarisation mismatch. Different frequencies add in mean intensity, while each frequency acquires its own relative phase. Normalising by total spectral power gives the complex degree of temporal coherence

$$\gamma(\tau) = \frac{1}{\Delta\nu} \int_{\nu_0 - \Delta\nu/2}^{\nu_0 + \Delta\nu/2} e^{2\pi i \nu \tau}\, d\nu.$$

Shift the integration variable to $s = \nu - \nu_0$. Frequencies equally spaced above and below the centre contribute opposite imaginary parts to the shifted integral. Integrating the real cosine contribution yields an envelope multiplying the carrier phase.

$$\gamma(\tau) = e^{2\pi i \nu_0 \tau} \frac{2}{\Delta\nu} \int_0^{\Delta\nu/2} \cos(2\pi s \tau)\, ds = e^{2\pi i \nu_0 \tau} \frac{\sin(\pi \Delta\nu \tau)}{\pi \Delta\nu \tau}.$$

A fine local phase scan measures the fringe visibility $|\gamma|$. The signed envelope changes sign after a zero, shifting the bright optical fringes by half a cycle relative to a continuation of the carrier. In a vacuum Michelson interferometer an arm-mirror displacement x changes the optical path by $2x$, hence

$$\tau = \frac{2x}{c},$$

$$x_{\text{first zero}} = \frac{c}{2\Delta\nu}.$$

Reducing the spectral width stretches the whole envelope and moves its zeros apart. A spectrum flat per unit wavelength transforms with a Jacobian when expressed per unit frequency, so a broad wavelength passband would require its measured frequency weights in the integral.

WORKED EXAMPLE

Predict a fringe revival beyond the first zero

Change the inputs and investigate this derivation.

For $\Delta\nu = 100$ GHz and $c = 3.00 \times 10^8$ m/s, the first zero occurs at 1.50 mm mirror travel. At 2.25 mm, $\Delta\nu\tau = 1.50$. The signed envelope is $-1/(1.50\pi) = -0.212$, predicting 21.2 percent visibility and a phase reversal. Loss of visible fringes at the first zero therefore need not persist as the mirror continues moving.

SECOND APPLICATION

Infer a filter width from neighbouring nulls

Change the inputs and investigate this derivation.

Another scan shows consecutive envelope zeros separated by 0.750 mm of mirror travel. The inferred bandwidth is

$$c/(2\Delta x) = 200\,\text{GHz}.$$

For a narrow fractional band centred on 750 nm the wavelength width is approximately $\lambda_0^2 \Delta\nu / c = 0.375$ nm. This conversion uses the local slope of $\nu = c/\lambda$, and the null-spacing inference uses the known double transit of the moving arm.

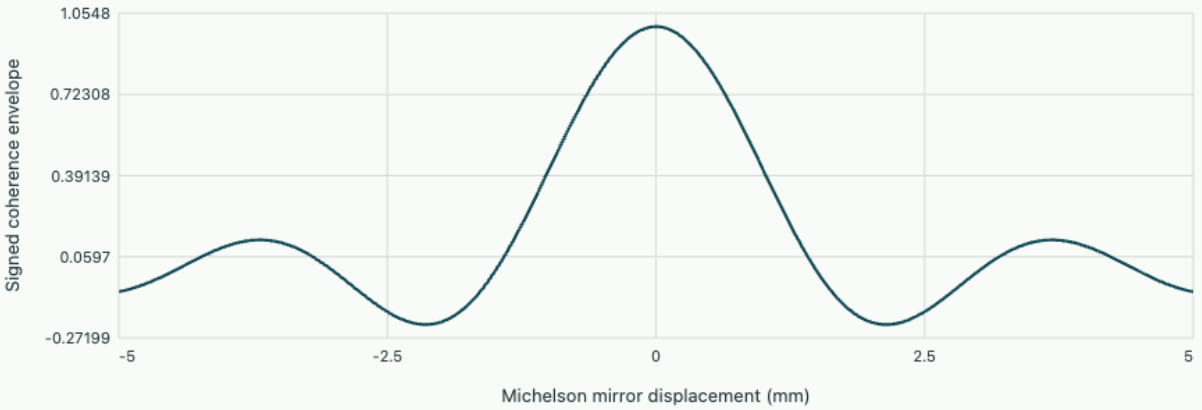
Try it · Rectangular spectrum

Vary the physical parameters in the derived relation. Use the units and assumptions in the readout to connect the curves to the two worked examples. A parameter study can test the forward prediction; the same relation can then be inverted to infer a parameter from a measurement.

Open this live calculation

Try the calculation

Spectral width (GHz) 100



Rectangular equal-frequency spectrum

Consecutive envelope nulls are separated by 1.499 mm. Negative signed envelopes indicate a π fringe-phase reversal; visibility is their magnitude.

Start from this chapter. Add two complex amplitudes of equal magnitude with a fixed relative phase. Predict where they cancel. The model drives coherent sound sources, so it isolates the path-difference contribution to fringe contrast.

OneSimulation laboratory · Find the dark fringes of a double slit

Compare a wave with its delayed copy

A Michelson interferometer divides one incoming beam into two arms, reflects the beams and recombines them. Moving one end mirror by d changes the round-trip optical path by $2d$ in air. With equal arm intensities, identical polarisations and a monochromatic source, the recorded intensity at one output varies from dark to bright and back. One complete fringe passes a detector whenever the mirror moves by half a wavelength.

A real source contains several frequencies or a continuous band. Different frequencies then accumulate different phases for the same delay. A detector averages over many optical cycles and, under the usual stationary-source assumption, adds the intensity contributions of mutually uncorrelated spectral

components. Bright fringes of one component can coincide with dark fringes of another. The visibility of the combined pattern can decay even when the beam paths remain perfectly aligned.

Write the measurement in terms of a correlation

Write $E(t)$ for the complex field amplitude in one arm and $E(t-\tau)$ for its delayed copy, with optical delay time τ. Expanding the detected intensity gives two arm intensities plus a cross term proportional to $\mathrm{Re}\langle E^*(t)E(t-\tau)\rangle$. Normalise the correlation by its zero-delay value to define $g^{(1)}(\tau)$. For equal arm intensities the contrast of fringes around a chosen delay is $|g^{(1)}|$, provided the envelope varies slowly across one fringe.

The magnitude expresses temporal coherence over that delay. A fixed imbalance of arm intensities also reduces visibility. If one arm has four times the power of the other, perfect coherence gives contrast 0.8 even at zero delay. To interpret a decrease with delay as spectral information, measure both arm powers individually and check that they remain stable during the scan. Spatial misalignment and unequal polarisation can also reduce the cross term without broadening the source spectrum.

Compute the signature of two spectral lines

Let two narrow lines of equal power have frequencies $\nu_0 - \Delta\nu/2$ and $\nu_0 + \Delta\nu/2$. For each line, the interference term varies with $\cos(2\pi\nu\tau)$. Their sum is proportional to $\cos(2\pi\nu_0\tau)\cos(\pi\Delta\nu\,\tau)$. The fast oscillation gives ordinary optical fringes and the envelope gives visibility $|\cos(\pi\Delta\nu\,\tau)|$. Its first zero occurs at $\tau = 1/(2\Delta\nu)$. Beyond that zero the fringes return with a phase reversal.

For a continuous spectral density $S(\nu)$, the same sum becomes an integral. The normalised correlation is the Fourier transform of S divided by its integrated power, up to the chosen phase convention. A uniform spectrum across a band of width B produces a sinc envelope $\sin(\pi B\tau)/(\pi B\tau)$. Its first zero occurs at $1/B$. A coherence time therefore depends on the chosen definition of width and on the line shape. A single universal numerical factor would be misleading.

Read an interferogram in the laboratory

For a pair of lines separated by 100 GHz, the first visibility minimum is at 5.0 ps. The optical path difference in air is $c\tau = 1.50$ mm; the required mirror displacement from equal paths is 0.750 mm. The first visibility revival occurs after 1.50 mm of mirror travel. A mirror scan that ends before the first dip will not resolve the two lines reliably from their contrast modulation.

Suppose the average wavelength near the centre is 600 nm. During the first 0.750 mm mirror displacement roughly $2(0.750\,\mathrm{mm})/(600\,\mathrm{nm}) = 2500$ fine fringes pass. Counting fine fringes calibrates the path scale, while locating the envelope minimum estimates spectral separation. The two observables use different parts of the same measurement record. This is an example of extracting information about a source through interference without physically separating its colours on a detector.

A drifting mirror can reduce the measured visibility by averaging different optical phases within one exposure. Shorter exposures or a separate stable reference fringe allow this instrumental loss to be distinguished from the coherence time of the source.

From a Lorentzian line to a coherence time

Two sharp lines produce a periodic visibility envelope. A single broadened line instead produces a decaying correlation. Let the power spectral density around carrier frequency ν_0 be a normalized Lorentzian with full width at half maximum $\Delta\nu$. Write frequency offset $u = \nu - \nu_0$ and $\Gamma = \Delta\nu/2$. The normalized offset density is

$$S(u) = \Gamma/[\pi(u^2 + \Gamma^2)].$$

Extending the narrow line to all real offsets is an excellent approximation when $\Delta\nu \ll \nu_0$.

For a stationary field, the complex first-order coherence at delay τ is the Fourier transform of the normalized spectral density. The carrier supplies a rapidly rotating phase factor. The remaining integral is $\int S(u)e^{2\pi i u\tau}du$. For positive delay close the integration contour in the upper half-plane where the exponential decays. Only the pole at $i\Gamma$ contributes; its residue gives $e^{-2\pi\Gamma\tau}$. Negative delays give the complex conjugate. Hence

$$g^{(1)}(\tau) = e^{2\pi i\nu_0\tau}e^{-\pi\Delta\nu|\tau|},$$

$$\tau_{1/e} = \frac{1}{\pi\Delta\nu}.$$

An equal-arm-intensity interferometer measures $|g^{(1)}|$ as fringe visibility. Several definitions of coherence time are in use, so an experimental statement should name the chosen width or threshold. A Gaussian spectral line gives a Gaussian delay envelope, allowing visibility scans to distinguish line shapes with the same half-maximum frequency width.

Infer linewidth without resolving the optical carrier

A Michelson interferometer in air gives calibrated zero-delay visibility 0.900. Moving one mirror by 30.0 mm reduces visibility to 0.450, after correcting for arm power. The optical path difference is twice the displacement, hence delay is

$$0.0600/c = 200\,\mathrm{ps}.$$

Division by the zero-delay visibility gives a normalized envelope of 0.500. For a Lorentzian this implies $\Delta\nu = \ln 2/(\pi\tau) = 1.10$ GHz. The $1/e$ coherence time is 289 ps and the corresponding path difference is

$$c\tau_{1/e} = 86.6\,\mathrm{mm}.$$

Doubling the mirror displacement to 60.0 mm doubles the delay. The normalized visibility should then fall to one quarter, giving measured visibility 0.225. A Gaussian envelope adjusted to the original half-visibility point would instead fall to $0.5^4 = 0.0625$ of its initial value. Measuring a second delay therefore tests the assumed line shape rather than merely refining a fitted linewidth. As linewidth tends to zero, the predicted envelope remains near one over every fixed path difference, although imperfect spatial overlap or phase jitter can still reduce the optical fringe itself.

A linewidth inferred from delay data is fundamentally a measurement of temporal correlation, so the relevant observation interval matters. Slow source-frequency drift during a long scan can yield a different envelope from that seen during short repeated measurements. One practical separation is to record forward and

backward delay scans and monitor a fixed, nonzero delay between them. A drifting carrier changes fringe phase there even when the single-scan visibility remains high; a broad instantaneous spectrum reduces the visibility within each scan. The two effects require different physical interpretations despite potentially resembling one another in a long-time spectrum.

Study the full derivation

Photon counting, bunching and the time resolution of a coincidence test · Optical fields, instruments and observations

An intensity trace can be steady while individual detector clicks show correlations. Derive the observable of a two-detector Hanbury Brown–Twiss measurement, quantify how background and bin width change it, and decide whether a dip near zero delay belongs to the source or to a detector.

Problem sheet 3 · Polarisation and coherence

Use Lectures 5 and 6 to diagnose a dim interference signal with separate polarisation, intensity and bandwidth measurements.

Measure a birefringent sample

A linearly polarised 2.0 mW beam of wavelength 500 nm enters a lossless plate at 45° to its fast axis. The crossed analyser transmits 0.50 mW. Find the smallest positive retardance and the corresponding thickness if $\Delta n = 0.020$. Explain why one reading cannot determine an arbitrary thickness.

▼ **Invert the analyser transmission**

The measured fraction is 0.25, giving $\sin^2(\delta/2) = 1/4$. The first positive solution is

$$\delta = \pi/3.$$

Thus

$$d = \delta\lambda_0/(2\pi\Delta n) = 4.17\,\mu\mathrm{m}.$$

Larger retardances with the same sine squared give the same reading. A thickness bound or a second wavelength can distinguish some of these branches. If the incident beam fails to make 45° to the plate, the inferred retardance is also biased.

Separate loss of coherence from unequal powers

Two interferometer arms separately deliver 0.90 and 0.10 mW at a detector. Together their fringes have visibility 0.30. Find the maximum visibility for perfect mutual coherence, estimate $|g^{(1)}|$ and predict the new visibility after attenuating the brighter beam to match the weaker one.

▼ **Normalise by the unequal-arm ceiling**

The perfect-coherence ceiling is

$$2\sqrt{I_1 I_2}/(I_1 + I_2) = 0.60.$$

Thus the measured modulus of the normalised mutual coherence is 0.50, assuming matching polarisation and spatial overlap. With balanced arms the ceiling rises to one, so visibility becomes 0.50 if attenuation is uniform and the coherence is preserved. The mean power falls to 0.20 mW. Better contrast and poorer photon statistics can occur together.

Identify a line separation from a mirror scan

A source with two equal, narrow spectral lines has its first contrast minimum 0.30 mm from equal arm lengths on a Michelson mirror stage. Find the separation in hertz, the next stage displacement giving full contrast,

and the effect of orthogonal polarisations in the two arms.

▼ **Convert motion into delay first**

The optical path difference at the first dip is 0.60 mm, or a delay of 2.00 ps. For the doublet $\Delta\nu = 1/(2\tau) = 250$ GHz. The first revival occurs at twice the delay and hence 0.60 mm of stage motion from equal paths. With mutually orthogonal arm polarisations their vector fields have zero cross term at a polarisation-insensitive detector. A common output analyser with nonzero projections onto both can recover fringes, provided the original temporal coherence remains.

Explore these relations

- For the birefringent plate, vary retardance and orientation and inspect analyser transmission. Open the live calculation.
- For coherence, vary spectral width and compare the coherence envelope with delay. This continuous rectangular spectrum is a contrast case for the two-line measurement. Open the live calculation.

Chapter 7 · Diffraction as a Fourier transform

Derive the far-field pattern from a finite aperture and identify which features control its scale and its sidelobes.

FROM FIRST PRINCIPLES

Sum the field across a uniformly illuminated slit

A distant screen receives contributions from across the whole aperture. Consider monochromatic scalar illumination normal to a long slit of width a. In the far field, paths from two points separated by distance dx differ by approximately $dx \sin\theta$ in direction θ. Distance-dependent amplitudes are nearly constant across the slit, leaving a linear phase variation. Put $k = 2\pi/\lambda$ and choose the origin at the slit centre.

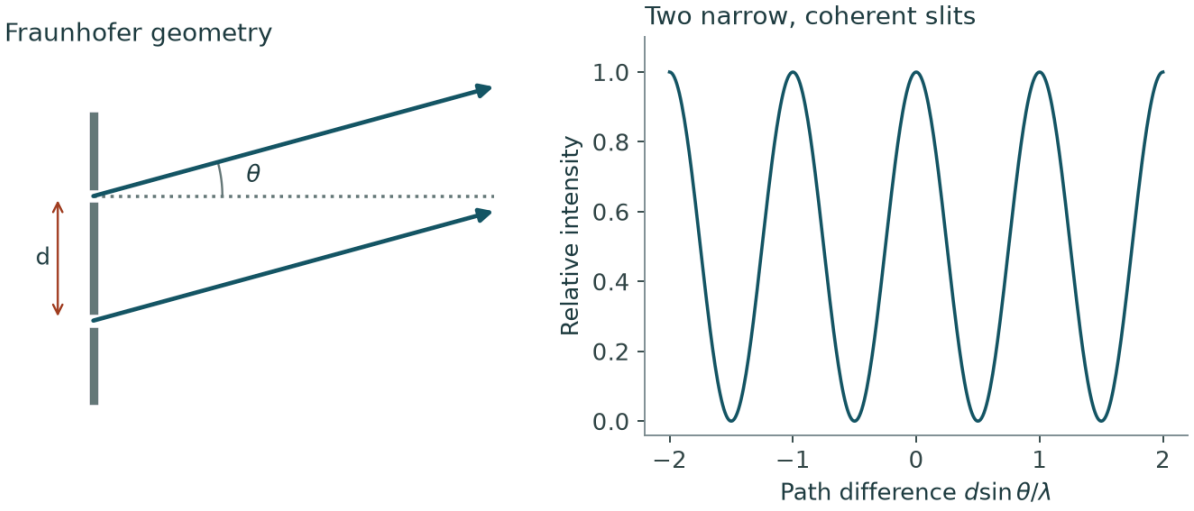
$$U(\theta) = C \int_{-a/2}^{a/2} e^{-ikx \sin\theta}\, dx.$$

Integrating the exponentials from the two edges and normalising to forward amplitude yields a sinc function. The central limit is one by continuity even though the displayed ratio takes the form zero divided by zero there.

$$\frac{U(\theta)}{U(0)} = \frac{\sin z}{z},$$

$$z = \frac{\pi a \sin\theta}{\lambda}.$$

Intensity is the squared magnitude of the coherent field sum. The first noncentral zero therefore occurs at $a|\sin\theta| = \lambda$. Widening the slit narrows its central angular peak. For a general transmission profile the same weighted integral is its Fourier transform sampled at transverse wavenumber $k \sin\theta$.



Read fringe order from the geometry

Illuminate two narrow slits normally so they emit equal, in-phase amplitudes with the same polarisation. Projecting their separation onto the far-field observation direction gives a path difference d sin θ. A path difference of half a wavelength adds the second amplitude with the opposite sign. A finite slit width introduces a separate envelope over these fringes.

Explore the related parameter study beside the computed figure.

DERIVATION AND APPLICATIONS

Locate missing interference orders in a double slit

A pair of finite slits can erase an interference maximum even when the detector is perfectly aligned. Consider two identical uniformly illuminated slits of width a, centre separation $d > a$, and long dimension perpendicular to the observation plane. Assume coherent monochromatic light of wavelength λ, scalar diffraction and far-field detection. At observation angle θ define transverse spatial frequency $q = 2\pi \sin\theta/\lambda$. Each point at coordinate x contributes phase e^{-iqx}. A centred slit therefore contributes

$$U_1(q) = C \int_{-a/2}^{a/2} e^{-iqx}\, dx = Ca\, \frac{\sin(qa/2)}{qa/2}.$$

Translating a slit multiplies its transform by a phase factor. The two translated copies have centres at $\pm d/2$. Add their fields, square the magnitude, and divide by the central intensity. The aperture width and slit separation now enter different factors.

$$\frac{I(\theta)}{I(0)} = \left[\frac{\sin(qa/2)}{qa/2}\right]^2 \cos^2(qd/2).$$

The centres of the ideal two-point interference orders satisfy

$$d \sin\theta = m\lambda.$$

Evaluate the single-slit envelope at those positions. The order disappears exactly when its location coincides with a noncentral zero of the envelope.

$$a \sin\theta = p\lambda,$$

$$m = p\frac{d}{a},$$

$$p = \pm 1, \pm 2, \ldots.$$

If d/a is a rational number, only those choices of p producing integer m erase interference orders. Finite width can also shift the actual maxima of the product slightly away from the nominal interference centres. The coincidence of a zero with an order centre remains exact in this ideal model.

WORKED EXAMPLE

Explain an absent green fringe

Change the inputs and investigate this derivation.

Two slits 50.0 µm wide have centre spacing 200 µm. At $\lambda = 500$ nm the fourth order lies at $\sin\theta = 4\lambda/d = 0.0100$. This is the first zero of the single-slit amplitude. A lens of focal length 200 mm would place that missing fringe about 2.00 mm from the centre. The first order samples an envelope factor $[\sin(\pi/4)/(\pi/4)]^2 = 0.811$.

SECOND APPLICATION

Reconstruct slit dimensions from a pattern

Change the inputs and investigate this derivation.

A different apparatus uses 600 nm light and a 300 mm focal length. Near the axis, the measured nominal interference spacing is 0.900 mm, giving $d = f\lambda/\Delta x = 200$ µm. The first envelope zero lies 2.25 mm from the centre, giving $a = f\lambda/x_1 = 80.0$ µm. Their ratio is 2.50. The first envelope zero lies halfway between two nominal orders, while the second erases order five at 4.50 mm.

> **Start from this chapter.** Use slit separation and wavelength to predict the first minimum before reading the finite-distance pressure field. Repeat at the same frequency with a different receiver radius to test the far-field assumption.
>
> OneSimulation laboratory · Find the dark fringes of a double slit

Where a geometrical shadow breaks down

Illuminate a thin screen at normal incidence with a monochromatic plane wave of vacuum wavelength λ in air. Light transmitted through different parts of an opening reaches a distant observation point along slightly different paths. Their phases matter even where geometrical rays would draw a hard shadow. Use scalar diffraction, suited to features much larger than the wavelength and modest angles. Represent the complex field just beyond the screen by an aperture function $A(x, y)$, incorporating amplitude and phase.

For a far-away observation direction $\mathbf{n}$, the propagation distance from aperture point $\mathbf{r}'$ is $r - \mathbf{n} \cdot \mathbf{r}'$ to first order. Replacing the inverse distance in the wave amplitude with $1/r$ while retaining this linear phase yields

$$U(\theta_x, \theta_y) \propto \iint A(x, y) e^{-ik(x \sin\theta_x + y \sin\theta_y)}\, dx\, dy.$$

The angular field samples a spatial Fourier transform. The neglected quadratic path term produces a phase of order $\pi a^2/(\lambda r)$ for aperture half-width a. The free-space far-field approximation requires this quantity to be small. A lens supplies a quadratic phase that creates the Fourier pattern in its back focal plane even at a practical laboratory distance.

Add the contributions of a uniform slit

For a very tall slit of width w, study the horizontal cross-section. Uniform amplitude across the width gives an integral of the exponential between $-w/2$ and $w/2$. The result is $w \sin u/u$ with $u = \pi w \sin\theta/\lambda$. Taking squared magnitude and normalizing at the origin yields

$$I(\theta)/I(0) = \left(\frac{\sin u}{u}\right)^2.$$

The first zeros satisfy

$$\sin\theta = \pm\lambda/w.$$

A broader opening makes the central spot narrower because a smaller angular change suffices to accumulate a full cycle across it. The oscillating tails arise from the abrupt edges. A gradual transmission taper reduces those sidelobes at the cost of a wider central maximum. Peak brightness additionally depends on the transmitted amplitude and collection geometry.

Measure width in a focal plane

Illuminate a 120 µm slit with 600 nm light and use a focal-length 200 mm lens. In the paraxial focal plane $X \simeq f\sin\theta$. The first minima occur at $X = \pm 1.00$ mm, giving a central width of 2.00 mm. At half that distance $u = \pi/2$, so the intensity is $4/\pi^2 = 0.405$ of its central value. These two measurements separately test angular scale and profile shape.

If the slit is moved sideways, the Fourier shift theorem supplies only a phase factor in the far field, and the intensity pattern remains centred in this ideal arrangement. If it is illuminated at a small angle instead, the incident field carries a linear phase across the slit and the whole pattern moves. This distinction becomes useful when diagnosing whether a displaced focal-plane spot comes from a decentered opening or a tilted incoming beam.

A narrow receiving slit or detector pixel also integrates across part of the pattern. If it straddles a predicted zero, it collects light from either side and reports a positive count even for perfect diffraction. Reducing detector width tests this averaging interpretation.

Observe an aperture before its far field forms

A finite opening can produce bright and dark features even on its axis as the detector moves away. Consider a uniformly illuminated circular aperture of radius a in an opaque thin screen. Use a monochromatic scalar plane wave and a paraxial on-axis detector at distance z large compared with a, yet not necessarily large compared with a^2/λ. At aperture radius r the extra path is $\sqrt{z^2 + r^2} - z \simeq r^2/(2z)$. The Fraunhofer approximation drops this quadratic phase; Fresnel propagation keeps it.

Relative to the unobstructed advancing plane wave, the on-axis field is proportional to $\int_0^a (2\pi r/i\lambda z) e^{i\pi r^2/(\lambda z)}\, dr$. The factor $r\,dr$ makes the substitution $u = \pi r^2/(\lambda z)$ immediate. Using the free propagation normalization gives

$$\frac{U(0,z)}{U_{\text{free}}(0,z)} = 1 - e^{i\pi N_F},$$

$$\frac{I(0,z)}{I_{\text{free}}} = 4\sin^2\left(\frac{\pi N_F}{2}\right),$$

$$N_F = \frac{a^2}{\lambda z}.$$

Successive annuli adding half a wavelength of path alternately reinforce and cancel on axis. An even integer Fresnel number gives perfect cancellation in the ideal model. An odd integer makes the on-axis intensity four

times the unobstructed value. There is no excess total power: interference redistributes light in the transverse plane. The result concerns local on-axis intensity at a specified distance rather than integrated transmission.

Predict successive axial extrema

Choose $a = 0.500$ mm and $\lambda = 500$ nm. Then

$$a^2/\lambda = 0.500\,\mathrm{m}.$$

A detector at $z = 0.250$ m sees $N_F = 2.00$, predicting an axial null. Moving it to $z = 0.500$ m gives $N_F = 1.00$ and a fourfold maximum. The aperture-to-distance ratio is at most 0.002 in these two measurements, supporting the small-angle path expansion even though the far-field criterion is not satisfied.

At $z = 5.00$ m the Fresnel number is 0.100 and the exact Fresnel expression gives

$$I/I_{\mathrm{free}} = 0.0979.$$

Its far-field expansion is $(\pi N_F)^2 = 0.0987$, now within about one percent. This limit recovers the inverse-square on-axis intensity at fixed opening and incident irradiance. A real photodiode averages several transverse points, which can fill an axial null, and a broadband source averages slightly different Fresnel numbers. Repeating the axial scan with a smaller active area or a narrower spectral band tests which averaging effect dominates the observed minimum.

A complementary screen can help verify that the observed oscillations arise from aperture interference. Within the same scalar boundary approximation, replacing the circular opening by an opaque disk makes the two complex transmitted fields add to the unobstructed field. On axis the disk's normalized field is therefore $e^{i\pi N_F}$ in this model, with unit intensity and a distance-dependent phase. A bright point behind an opaque disk may seem surprising from geometric rays, but neighbouring angles still redistribute the blocked power. Finite coherence, source size and a detector averaging over the narrow central feature can strongly reduce its measured contrast.

From Fresnel zones to a diffractive lens

The axial null from a clear circular opening results from destructive addition of neighbouring half-period zones. A zone plate reverses that outcome by deliberately blocking every other annulus. At desired focal distance f, a ray from the edge of zone number m has path excess $m\lambda/2$ relative to the centre. Solving $\sqrt{f^2 + r_m^2} - f = m\lambda/2$ gives

$$r_m^2 = m\lambda f + \frac{m^2\lambda^2}{4} \simeq m\lambda f.$$

The approximation requires the outer zone radius to remain small compared with focal distance. Consecutive radii therefore grow approximately as the square root of zone number, and outermost zones are narrowest. Leaving odd zones transparent and even zones opaque removes contributions with the opposite phase at the design focus. Power is also distributed into other diffraction orders and an undiffracted background, so this simple plate does not concentrate all incident light into one focus.

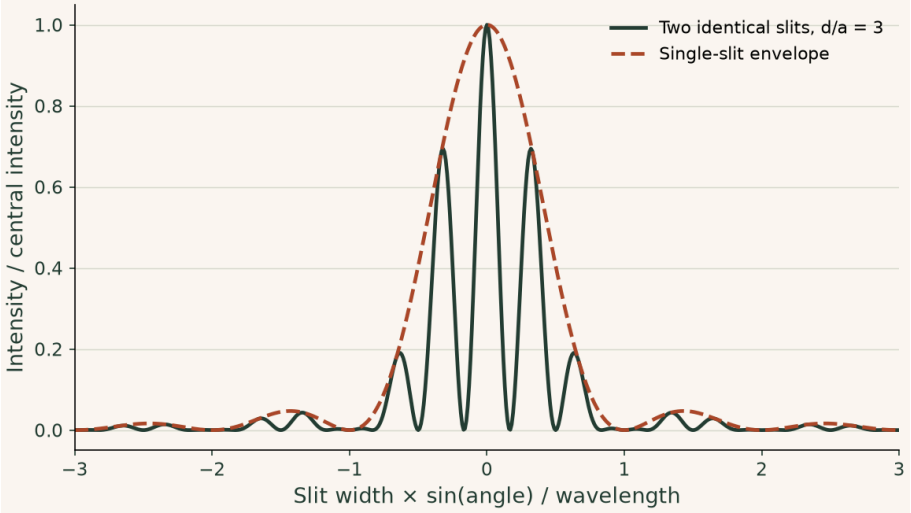
Set $\lambda = 500$ nm and design for $f = 0.200$ m with twenty zones. The radius of the first zone is $\sqrt{\lambda f} = 0.316$ mm, and the twentieth radius is 1.414 mm. The width of that outer zone is $\sqrt{\lambda f}(\sqrt{20} - \sqrt{19}) = 35.8$ µm, setting a fabrication scale substantially finer than the overall aperture. If the plate is used at 600 nm without changing its radii, its first-order focal length becomes

$$r_1^2/\lambda = 0.167\,\mathrm{m}.$$

This shortening with increasing wavelength follows from the required relative phase and is an observable distinction from simply changing the clear aperture of a refracting lens. Scanning on-axis intensity at two colours tests both the zone layout and its predicted chromatic shift.

Which interference orders disappear under diffraction?

Two uniformly illuminated slits have width a and centre spacing $3a$. The Fraunhofer field is the product of the single-aperture transform and a two-source phase sum. Normalise intensity at zero angle and use $u = a \sin\theta/\lambda$.



The nominal third interference maximum occurs at u = 1 and is extinguished by the first zero of the slit envelope. A missing order diagnoses the width-to-spacing ratio; it need not indicate missing or blocked illumination.

Try it · Double slit

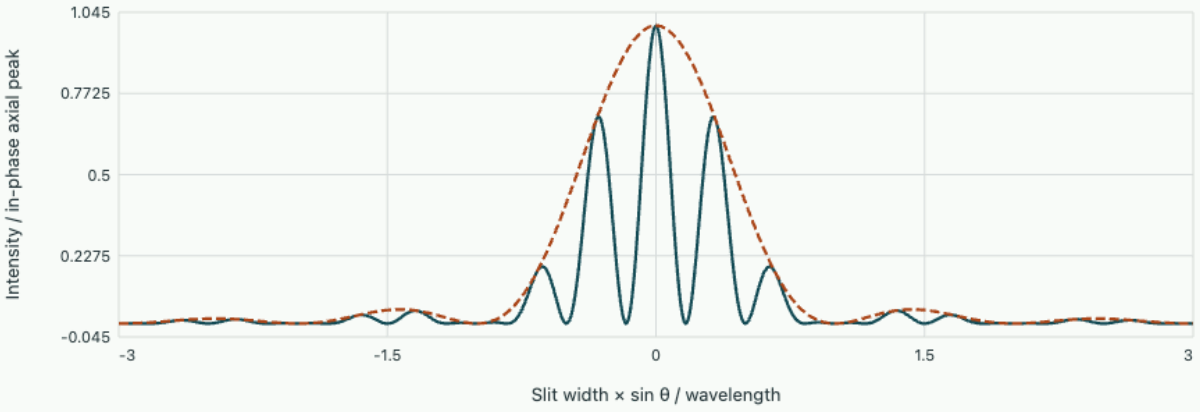
Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Slit spacing / slit width **3**

Relative incident phase (degrees) **0**



Two finite slits Single-slit envelope

For zero phase and integer d/a, the corresponding interference order is removed by the first envelope zero. Fraunhofer scalar model with equal slit illumination.

▶ **Read values from the plot**

Study the full derivation

Recover an optical phase from one off-axis intensity frame · Optical fields, instruments and observations

A camera measures intensity and discards field phase. A known tilted reference wave can shift a cross term away from the intensity background so the missing complex field becomes accessible. Derive how much camera bandwidth that costs, and convert the recovered phase into an optical-path measurement with a stated ambiguity.

Chapter 8 · Images through a finite pupil

Separate single-point diffraction, two-point resolution and spatial-frequency contrast for a circular imaging aperture.

FROM FIRST PRINCIPLES

Relate the pupil field to a point image

Treat an on-axis distant point as a uniform plane wave over a lens pupil. Assume scalar, monochromatic, paraxial propagation and an ideal thin lens of focal length f. The lens cancels the quadratic propagation phase at focus. Moving a small distance x across the focal plane introduces a nearly linear phase across pupil position ξ. For a one-dimensional pupil function $P(\xi)$ the field sum is

$$U(x) \propto \int P(\xi) \exp\left[-i\frac{2\pi x\xi}{\lambda f}\right] d\xi.$$

The recorded point-spread function h is proportional to $|U|^2$ and can be normalised to unit integral. For a uniform slit pupil of width D, the first zero follows directly from the phase difference across its edges.

$$h(x) \propto \left[\frac{\sin(\pi D x/\lambda f)}{\pi D x/\lambda f}\right]^2,$$

$$|x_{\text{first}}| = \lambda f/D.$$

An extended incoherent object consists of mutually uncorrelated point emitters, so their shifted intensity patterns add. Where the point response is independent of field position this gives $I_{\text{image}} = I_{\text{object}} * h$ in common image coordinates. A wider pupil narrows the blur and passes finer object structure.

Background · Numerical aperture

$$\text{NA} = n \sin\theta$$

The largest accepted angle controls which transverse spatial frequencies reach the pupil. Resolution also depends on the optical imaging and coherence assumptions.

DERIVATION AND APPLICATIONS

Relate a defocus error to the on-axis point intensity

A displaced detector spreads the intensity of an unresolved point even if the lens is otherwise perfect. Consider a uniformly illuminated circular pupil of radius a and a monochromatic point on the optical axis. Use scalar paraxial diffraction in air and a clear pupil. Let ϕ be the phase error at the rim relative to the centre. Defocus contributes a quadratic phase $\phi r^2/a^2$. At the on-axis image point the in-focus contributions are in phase, so the defocused complex amplitude relative to its in-focus value is the pupil-area average

$$H(0) = \frac{1}{\pi a^2}\int_0^a e^{i\phi r^2/a^2} 2\pi r\, dr = \int_0^1 e^{i\phi u}\, du,$$
$$u = r^2/a^2.$$

Equal increments of u label equal pupil areas. Integrating these phasors gives a phase factor times a real amplitude. Its squared magnitude is the axial intensity at the intended point relative to the ideal peak.

$$H(0) = e^{i\phi/2}\frac{\sin(\phi/2)}{\phi/2},$$
$$S = |H(0)|^2 = \left[\frac{\sin(\phi/2)}{\phi/2}\right]^2.$$

For a detector displaced axially by magnitude $|\Delta z|$ from a focus of focal length f, the change in quadratic propagation phase has magnitude $k|\Delta z|r^2/(2f^2)$ to first order in $|\Delta z|/f$. With paraxial numerical aperture $\mathrm{NA} = a/f$ this gives

$$|\phi| \simeq \frac{\pi\,\mathrm{NA}^2|\Delta z|}{\lambda}, \qquad S \simeq 1 - \frac{\phi^2}{12} \quad (|\phi| \ll 1).$$

Total transmitted power stays constant. A loss of axial intensity represents redistribution across the image plane and the same axial value can occur for displacements on either side of best focus.

WORKED EXAMPLE

Set a permissible detector displacement

Change the inputs and investigate this derivation.

A microscope working in air has $\mathrm{NA} = 0.100$ and $\lambda = 500$ nm. An axial displacement of 25.0 µm corresponds to $|\phi| = \pi/2$. The exact pupil average gives

$$S = [\sin(\pi/4)/(\pi/4)]^2 = 0.811.$$

A detector that measures only the central intensity would therefore show a 19 percent loss even with ideal transmission.

SECOND APPLICATION

Infer numerical aperture from a through-focus scan

Change the inputs and investigate this derivation.

A separate lens at 640 nm shows its first axial zero 80.0 µm from best focus. The first nonzero root occurs at $|\phi| = 2\pi$, so

$$\mathrm{NA} = \sqrt{2\lambda/|\Delta z|} = 0.126.$$

If the effective focal length is 50.0 mm, the inferred pupil radius is 6.32 mm. The scan thus measures the effective illuminated aperture; partial or graded illumination would change the pupil weighting in the first integral.

Try it · Defocus strehl

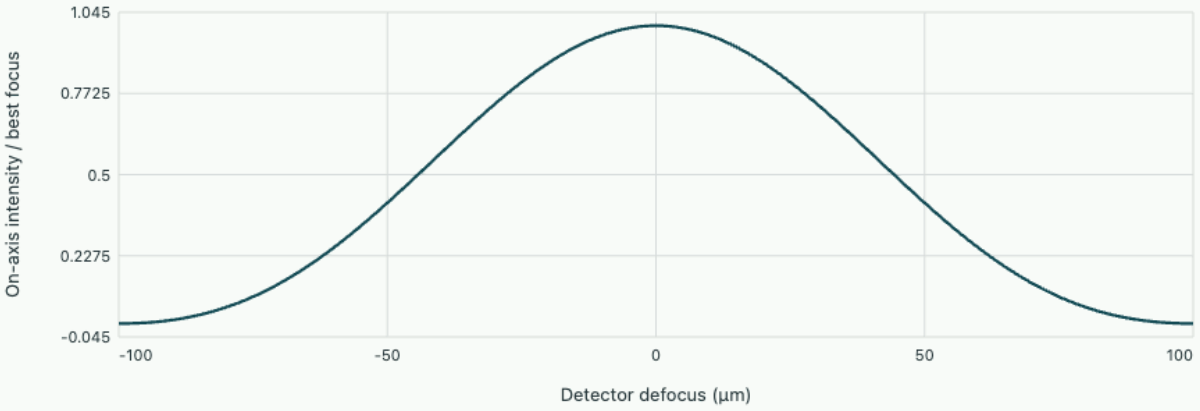
Vary the physical parameters in the derived relation. Use the units and assumptions in the readout to connect the curves to the two worked examples. A parameter study can test the forward prediction; the same relation can then be inverted to infer a parameter from a measurement.

Open this live calculation

Try the calculation

Numerical aperture **0.1**

Wavelength (nm) **500**



Uniform circular pupil

Pupil phase span ϕ=πΔz NA²/λ. First axial zero at |Δz|=100 µm. Scalar paraxial diffraction with no other aberrations.

A point becomes an extended image

A thin lens focuses a distant point into its back focal plane. An ideal geometrical ray diagram predicts a mathematical point. The finite entrance pupil limits the transverse field that can contribute. Define the point-spread function as the image intensity from a single on-axis point. Restrict attention to an aberration-free, circular, uniformly illuminated pupil, a narrow spectral band and an image region sufficiently small that shifting the source translates the same point-spread function.

The preceding aperture transform now runs over a disk of radius a. For an observation angle θ the transverse wavenumber is $q = k \sin\theta$. Performing the angular integral of the plane-wave phase gives $2\pi J_0(qr)$. The radial integral then uses $d[zJ_1(z)]/dz = zJ_0(z)$, so the field normalized at the centre is $2J_1(qa)/(qa)$. Squaring gives the Airy intensity pattern. Its first zero is at $qa = 3.8317$, hence

$$\sin\theta_1 = 1.22\lambda/D,$$
$$D = 2a.$$

Combine two point sources

For statistically independent emitters the fluctuating relative phase averages away. Their point-spread intensities add. The Rayleigh convention places the centre of one Airy spot on the first zero of its neighbour. It provides a useful separation scale for two comparable emitters, though achievable accuracy also depends on photon counts, background and whether the measurement estimates a known model. Two coherent emitters instead add complex fields before squaring. Their relative phase can brighten or extinguish the midpoint.

For an extended incoherent object each point contributes a shifted intensity pattern. Superposition gives convolution of object intensity with the intensity point-spread function. Taking a Fourier transform converts convolution to multiplication. The resulting optical transfer function is the autocorrelation of the pupil in spatial-frequency coordinates. Its support reaches twice the radius of the coherent field passband because displaced pupil copies overlap until their centres are separated by one diameter. In object space the incoherent cutoff in cycles per metre is $2\,\mathrm{NA}/\lambda_0$, where $\mathrm{NA} = n\sin\alpha$ and λ_0 is vacuum wavelength.

Choose the object-side sampling

Use an air objective of numerical aperture 0.40 at 520 nm. The Rayleigh scale is $0.61\lambda_0/\mathrm{NA} = 0.793$ µm. The incoherent cutoff is 1.54 cycles/µm, corresponding to a limiting sinusoidal period of 0.650 µm with contrast vanishing at the cutoff. These describe different target geometries. A grating near the cutoff can have weak contrast even though its nominal period exceeds the reciprocal cutoff.

To sample the full image band without aliasing, the object-side pixel pitch must be at most $1/(2f_{\mathrm{cut}}) = 0.325$ µm. A detector with 6.5 µm pixels therefore needs at least $20\times$ lateral magnification. A still higher magnification gives more samples of the same optical image. Recovering finer physical structure would require a larger numerical aperture, a shorter wavelength, or additional measurement information.

A separate point-source image is a practical calibration target. Its recorded radial profile can be convolved with a proposed object model to predict the measured image, including overlapping spots, before attempting an inversion of noisy data.

Derive the contrast transfer from two overlapping pupils

A cutoff frequency specifies where contrast disappears, but it does not tell how much contrast survives below cutoff. For incoherent imaging with a clear circular pupil, the normalized optical transfer function is the overlap area of two pupil disks displaced in reciprocal space, divided by one disk area. Consider one spatial-frequency magnitude ν and normalize it by the incoherent cutoff, $q = \nu/\nu_c$. At $q = 0$ the disks coincide; at $q = 1$ their centres are one full diameter apart. For pupil radius a the centre separation is therefore $2aq$.

The overlap lens consists of two circular segments. In either disk the half-angle between radii to the intersection points is $\arccos q$. The sector area minus the area of its isosceles triangle is

$a^2[\arccos q - q\sqrt{1-q^2}]$. Double this result and divide by πa^2. For ideal focus the transfer is real and nonnegative,

$$H(q) = \frac{2}{\pi}[\arccos q - q\sqrt{1-q^2}], \qquad 0 \le q \le 1.$$

At zero spatial frequency $H = 1$, expressing conservation of a uniform background after suitable calibration. At the cutoff only tangential contact remains and $H = 0$. If the object intensity is $I_0[1 + m\cos(2\pi\nu x)]$, linear shift-invariant incoherent imaging leaves the average unchanged while multiplying its modulation depth by $H(q)$. This establishes a direct experimental meaning for a frequency-dependent contrast loss.

Separate objective and detector contrast losses

An air microscope at 500 nm has NA 0.50, giving $\nu_c = 2\mathrm{NA}/\lambda = 2.00$ cycles/µm in object space. A target with period 1.00 µm has $q = 0.500$. The circular overlap formula gives

$$H = (2/\pi)(\pi/3 - 0.5\sqrt{0.75}) = 0.391.$$

An object modulation $m = 0.600$ therefore appears optically with modulation 0.235, even under ideal focus.

Suppose each detector pixel integrates over object-side width $p = 0.250$ µm, with contiguous active pixels. Averaging the sinusoid over a width p gives another multiplier $\sin(\pi\nu p)/(\pi\nu p) = 0.900$ at this target frequency. The recorded modulation is consequently about 0.211. This decrease occurs before sampling values at the pixel centres and differs from aliasing. The detector Nyquist frequency is $1/(2p) = 2.00$ cycles/µm, just enough to cover the optical band. Reducing pixel pitch can recover its own averaging loss but leaves the objective's 0.391 transfer unchanged. Calibrating with several grating periods distinguishes an optical contrast curve from a simple constant gain error.

This test also explains why visibility estimates near the optical cutoff are noisy. Recovering an object modulation from an image would require dividing by the optical and pixel transfer factors. At the example period their product is about 0.352, already amplifying a small modulation measurement error by nearly three. Close to cutoff the corresponding inverse becomes much larger. More exposure time reduces photon-counting uncertainty but cannot recover a component at which the transfer is exactly zero. A known calibration target is preferable when testing the transfer curve, since otherwise a weak observed fringe can arise from either the specimen or the imaging system.

Study the full derivation

What an objective must collect to resolve a specimen · Optical fields, instruments and observations

A microscope can transmit plenty of light and still erase fine detail. Treat the specimen as a weak grating, follow its spatial harmonics through the objective, and determine why changing illumination coherence changes the transfer of contrast.

Problem sheet 4 · Fourier pupils and image sampling

Reconstruct a slit, predict a shifted diffraction image and specify a sampled image for an incoherent microscope.

Recover an aperture width

A 250 mm lens gives first minima separated by 2.40 mm under 480 nm light. Find the single-slit width. Predict the intensity fraction 0.60 mm from centre and the new separation if a 600 nm source is substituted.

▼ **Show the transform and its scale**

The first minimum lies at $X_1 = f\lambda/w = 1.20$ mm, so $w = 100$ µm. At $X_1/2$, the sinc argument is $\pi/2$ and intensity is $4/\pi^2 = 0.405$ of peak. Keeping the same lens and aperture, angular width scales with wavelength. At 600 nm the first-zero separation becomes

$$2.40(600/480) = 3.00\,\mathrm{mm}.$$

Shift or tilt?

The same slit moves sideways by 50 µm. Derive the resulting far-field phase factor and identify the intensity change. Then restore the slit position and tilt the incoming plane wave by 2.0 mrad. Predict the position of the central spot and compare both manipulations.

▼ **Show the shift theorem**

Writing the translated aperture as $A(x - x_0)$ and putting $u = x - x_0$ in the transform gives

$$U(q) = e^{-iqx_0}U_0(q).$$

Its squared magnitude is unchanged. Tilt instead multiplies the aperture field by $e^{ik\alpha x}$, replacing the pattern by $U_0(q - k\alpha)$. The peak moves to $X = f\alpha = 0.500$ mm for the chosen sign. Either statement presumes the illuminating beam covers the translated slit uniformly and the lens pupil does not clip the pattern.

Pixel pitch or optical resolution?

An air objective with NA 0.50 images independent emitters at 600 nm onto 5.0 µm pixels with magnification 10. Calculate the Rayleigh scale, optical cutoff and detector Nyquist frequency referred to object space. Find the minimum magnification sampling the full passband.

▼ **Show the independent limits**

The Rayleigh scale is $0.61(0.600)/0.50 = 0.732$ µm. The incoherent cutoff is $2(0.50)/0.600 = 1.67$ cycles/µm. Each pixel spans 0.500 µm in the object plane, so Nyquist frequency is $1/(2 \times 0.500) = 1.00$ cycle/µm. To capture all optical frequencies without aliasing set object pitch at most 0.300 µm, requiring magnification

$M \geq 5.0/0.300 = 16.7$. Better sampling preserves more optical contrast but cannot transmit spatial frequencies outside the optical passband.

Explore these relations

- For the aperture transform, vary slit width and inspect the diffraction envelope; the two-slit interference factor is an additional feature absent from a single slit. Open the live calculation.
- For pupil-limited imaging, vary aperture and wavelength and compare the diffraction response under defocus. Open the live calculation.

Infer slit spacing from a sampled angular minimum

The first minimum is measured at 9.75° on the 0.90 m receiver arc, with a 0.25° angular sample spacing. Assuming a far-field two-slit model, infer the dimensionless spacing d/λ. Propagate half a sampling interval as an angle uncertainty. Does this source of uncertainty alone cover the designed spacing of 3?

▼ **Work through the solution**

The first destructive condition is $d \sin\theta = \lambda/2$, so

$$d/\lambda = 1/[2\sin(9.75^\circ)] = 2.9525.$$

The fractional change is $|\cot\theta\, d\theta|$. With $0.125^\circ = 0.002182$ radians this is about 1.27%, or 0.0375 in d/λ. The upper estimate is about 2.990, below 3. The sampling interval alone falls short. Finite observation distance and the model of aperture illumination can also move the observed minimum.

Use the OneSimulation laboratory to check your answer

Set separation/wavelength to 3, radius to 0.90 m and width to 0.050 m. Compare the sampled finite-radius minimum with the far-field curve before attributing the discrepancy to angular sampling alone. Open the corresponding live laboratory reference.

Chapter 9 · Multiple-beam interference in a cavity

Sum the returning fields in a plane-parallel optical cavity and turn measured line spacing and width into a mirror diagnostic.

FROM FIRST PRINCIPLES

Sum the repeated transmissions of a cavity

Consider an ideal plane cavity with two identical lossless mirrors of power reflectance $R < 1$, normal incidence and coherent monochromatic light. The first transmitted field has magnitude $(1 - R)$ times the incident field. Each later transmitted copy makes one extra round trip, acquiring an amplitude multiplier R and a phase δ that includes any mirror reflection phases. In steady state the optical fields add as a convergent geometric series.

$$\frac{E_t}{E_i} = (1 - R)e^{i\psi} \sum_{j=0}^{\infty} (Re^{i\delta})^j = \frac{(1 - R)e^{i\psi}}{1 - Re^{i\delta}}.$$

The phase ψ acquired on the first traversal cancels in intensity. Multiply the denominator by its complex conjugate and use $1 - \cos\delta = 2\sin^2(\delta/2)$ to obtain

$$\mathcal{T} = \frac{(1 - R)^2}{(1 - R)^2 + 4R\sin^2(\delta/2)}.$$

At $\delta = 2\pi m$ all transmitted copies arrive in phase and the ideal transmission reaches unity. Increasing reflectance allows more copies to matter, so a smaller change in phase spoils their alignment and narrows the transmission peak. Loss, mirror imbalance or imperfect mode overlap can reduce its height.

DERIVATION AND APPLICATIONS

Measure mirror imbalance from resonant transmission

High finesse alone does not determine the brightness of a resonant cavity. Consider two partially transmitting lossless mirrors at normal incidence, separated by a transparent gap. Their power reflectances are R_1 and R_2, and $T_j = 1 - R_j$. Assume the incident spatial mode is matched, a single polarisation is present and reflected and transmitted fields are normalised to equal external impedances. Absorb mirror reflection phases into the round-trip phase δ. The first transmitted field has magnitude $\sqrt{T_1 T_2}$ relative to the incident field. Each extra circulation multiplies it by $\sqrt{R_1 R_2}e^{i\delta}$.

$$\frac{E_t}{E_i} = e^{i\psi}\sqrt{T_1 T_2} \sum_{n=0}^{\infty} \left(\sqrt{R_1 R_2}e^{i\delta}\right)^n = \frac{e^{i\psi}\sqrt{T_1 T_2}}{1 - \sqrt{R_1 R_2}e^{i\delta}}.$$

The common single-pass phase ψ disappears on taking the power ratio. At a resonance the phase gained per round trip is an integer multiple of 2π and every contribution adds in phase. The resulting peak transmission is

$$\mathcal{T}_{\mathrm{res}} = \frac{(1-R_1)(1-R_2)}{(1-\sqrt{R_1R_2})^2}.$$

There is no internal loss, so the untransmitted fraction exits through the input port. Subtracting the numerator from the denominator and factoring the difference gives the reflected power on resonance.

$$\mathcal{R}_{\mathrm{res}} = 1 - \mathcal{T}_{\mathrm{res}} = \left(\frac{\sqrt{R_1}-\sqrt{R_2}}{1-\sqrt{R_1R_2}}\right)^2.$$

Identical mirrors cancel the net reflected field on resonance even when each isolated mirror transmits only a small fraction. The result is symmetric under exchange of the two mirrors. A power scan by itself cannot determine which mirror is the more reflective one.

WORKED EXAMPLE

Predict the output of unequal mirrors

Change the inputs and investigate this derivation.

Let the input and output mirror field-reflection magnitudes be 0.900 and 0.800. Their power transmissions are 0.190 and 0.360. The peak transmitted fraction is

$$(0.190)(0.360)/(1-0.720)^2 = 0.872.$$

For 5.00 mW incident in the matched mode this gives 4.36 mW transmitted and 0.638 mW reflected. The sum equals the incident power to rounding.

SECOND APPLICATION

Identify a replacement output mirror

Change the inputs and investigate this derivation.

A second cavity has a known input field-reflection magnitude 0.800 and an output mirror independently known to be less reflective. Its measured resonant reflected power is $1/16$ of incident power. Taking the appropriate square root gives

$$(0.800 - r_2)/(1 - 0.800 r_2) = 0.250.$$

Solving yields $r_2 = 0.6875$ and $R_2 = 0.473$. The predicted peak transmitted fraction is $15/16 = 0.9375$, providing a separate measurement to test the lossless and mode-matching assumptions.

Try it · Asymmetric cavity

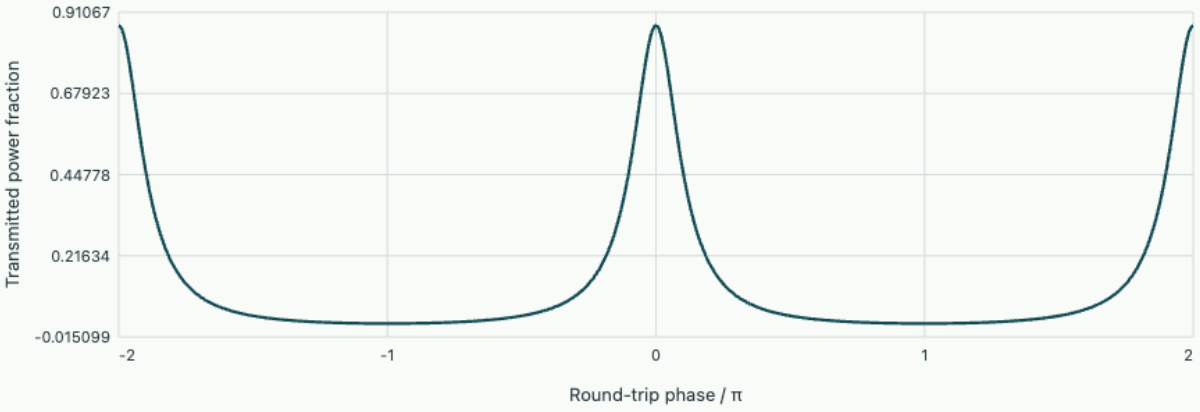
Vary the physical parameters in the derived relation. Use the units and assumptions in the readout to connect the curves to the two worked examples. A parameter study can test the forward prediction; the same relation can then be inverted to infer a parameter from a measurement.

Open this live calculation

Try the calculation

Input field reflectivity **0.9**

Output field reflectivity **0.8**



Unequal lossless cavity mirrors

Peak transmission 0.87245. Peak reflection 0.12755. Matched mirror amplitudes recover unit peak transmission.

Many transmitted copies

A transparent gap of length L and refractive index n sits between two identical plane mirrors. Illuminate it at normal incidence with a coherent, monochromatic plane wave. Assume the mirrors are lossless, their intensity reflectivity R is constant over the band, and the outgoing beams overlap in space. Each trip between mirrors adds phase. Each completed round trip also reflects the field twice. Because one mirror reflection has amplitude magnitude $\sqrt{R}$, the field magnitude after both is multiplied by R.

Choose a phase convention absorbing constant reflection phases into the resonance condition. A single transmitted pass has amplitude magnitude $(1 - R)$ times the incident amplitude. The pass emerging after one additional round trip acquires $Re^{i\delta}$ with $\delta = 4\pi nL/\lambda_0$. Summing the geometric series and squaring the magnitude gives

$$\mathcal{T} = \frac{(1-R)^2}{|1 - Re^{i\delta}|^2} = \frac{1}{1 + F\sin^2(\delta/2)},$$

$$F = \frac{4R}{(1-R)^2}.$$

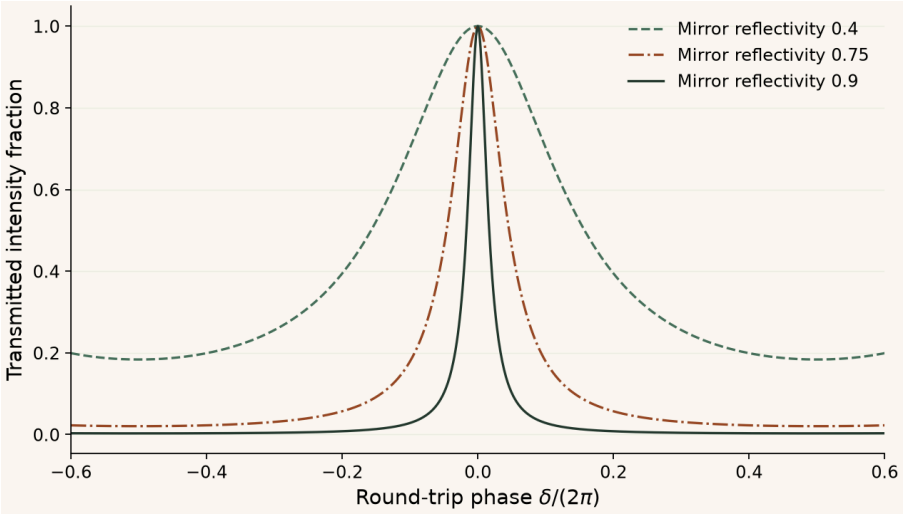
Resolving the peaks

Transmission reaches unity at $\delta = 2\pi m$ in this symmetric lossless model. Coherent addition of multiply reflected fields simultaneously cancels the returning field at the input. With unequal mirrors or intracavity loss, peak transmission is generally lower. The resonance frequencies are $\nu_m = mc/(2nL)$ when dispersion is negligible, giving a free spectral range $\mathrm{FSR} = c/(2nL)$. If n varies appreciably, differentiation of round-trip phase replaces it with the group index.

The half-power points satisfy $|\sin(\delta/2)| = 1/\sqrt{F}$ near resonance. Their full separation in phase is $4\arcsin(1/\sqrt{F})$ when $F \geq 1$. Dividing the 2π phase interval by that width yields finesse,

$$\mathcal{F} = \pi/[2\arcsin(1/\sqrt{F})].$$

For large F this reduces to $\pi\sqrt{R}/(1-R)$. Finesse counts how many linewidths fit between adjacent orders. It is distinct from resolving power $\nu/\Delta\nu$, which also depends on the order.



For a symmetric lossless cavity, stronger reflection narrows the allowed range of round-trip phase while the on-resonance transmission stays at unity.

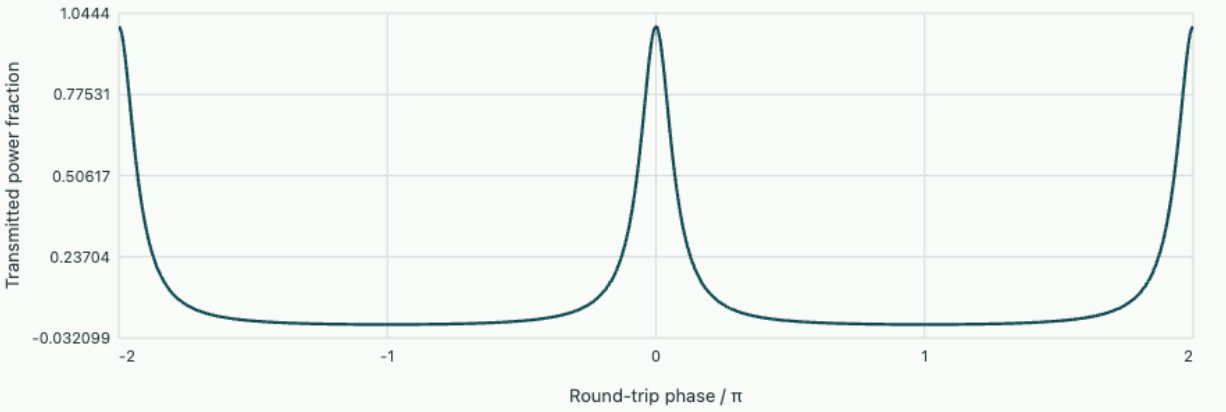
Try it · Fabry perot

Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Mirror power reflectivity **0.8**



Symmetric lossless Fabry–Pérot

Finesse coefficient 4R/(1–R)² = 80. Stronger reflection narrows each resonance while the peak remains one.

Inspect a scanned cavity

An air gap of 15.0 mm with $R = 0.900$ has a free spectral range of 9.99 GHz. The exact finesse is about 29.8, so adjacent half-power frequencies across one peak differ by 335 MHz. At a 600 nm carrier that width corresponds to about 0.00040 nm. Two lines one free spectral range apart nevertheless reach resonance together. Very narrow peaks and unambiguous identification across a broad input spectrum are different instrument requirements.

Mirror tilt makes different parts of a beam see different round-trip phases. Finite source divergence has a similar effect, since oblique rays accumulate a different axial phase. If raising the nominal mirror reflectivity no longer narrows a measured peak as predicted, compare the line shape at two illuminated spot sizes or angular apertures. Such a test separates instrumental phase averaging from a change in the source spectrum.

Measuring a frequency scan in both directions supplies a useful consistency test. Thermal expansion of the spacer or actuator hysteresis can change the calibration between scan position and optical frequency. An

independently known reference line or sideband establishes that calibration without relying on the actuator displacement alone.

Derive cavity linewidth from the decay after switch-off

A cavity that transmits a narrow frequency range also releases stored energy slowly. This connection can be demonstrated without fitting another steady interference pattern. Let $t_{\mathrm{rt}} = 2nL/c$ be the round-trip time in a nondispersive empty-gap model and let ρ be the intensity survival fraction per completed round trip, including both mirrors and all distributed losses. After turning off a resonant input, the energy remaining after j round trips is $U_j = U_0\rho^j$. For decay slow compared with one trip, replace the discrete sequence by its continuous envelope,

$$U(t) = U_0 e^{-t/\tau_U},$$

$$\tau_U = -t_{\mathrm{rt}}/\ln\rho.$$

The output power leaking through a fixed mirror follows this energy envelope. A field amplitude instead decays as $e^{-t/(2\tau_U)}$. Fourier transforming that causal amplitude gives a response proportional to $[1/(2\tau_U) - i2\pi(\nu-\nu_m)]^{-1}$ close to a resonance. Squaring the magnitude gives a Lorentzian with full width at half maximum $\Delta\nu \simeq 1/(2\pi\tau_U)$. This single-line approximation requires many round trips during a lifetime, or well-separated narrow resonances. It links a time-domain loss measurement to the spectral width obtained by scanning the cavity.

For two identical mirrors of intensity reflectivity R and negligible internal absorption, $\rho = R^2$. In the high-reflectivity limit $\ln\rho \simeq -2(1-R)$, reproducing the familiar finesse dependence of linewidth. Other loss channels can shorten the lifetime even if the two measured peak frequencies keep the same separation.

Determine an additional internal loss

An air cavity has length 0.300 m, giving $t_{\mathrm{rt}} = 2.00$ ns and free spectral range 500 MHz. Both mirrors have independently calibrated $R = 0.9950$. With no other losses the expected survival is 0.990025 and energy lifetime is

$$-2.00\,\mathrm{ns}/\ln(0.990025) = 200\,\mathrm{ns}.$$

A fast detector instead records an intensity $1/e$ time of 150 ns. Effective survival is then

$$\exp(-2.00/150) = 0.98676.$$

Dividing by R^2 gives an internal survival factor 0.9967 per round trip, an extra loss of about 0.33 percent.

The inferred full linewidth is $1/(2\pi \times 150\,\mathrm{ns}) = 1.06$ MHz, well below the free spectral range. A frequency sweep should agree if the driving laser is narrower than this line and the scan is sufficiently slow to reach steady response. If a measured line is broader while the ring-down lifetime remains unchanged, source jitter or mechanical length fluctuations can broaden the scan without absorbing additional circulating energy. At very low reflectivity the separate-resonance model fails, and individual delayed output pulses provide a more direct interpretation.

The extra loss can be turned into a material attenuation estimate if its path length is known. If all the inferred internal loss came from a sample traversed over 20.0 mm of material per round trip, its intensity attenuation coefficient would be $-\ln(0.9967)/0.0200 = 0.165$ m⁻¹. This attribution assumes that inserting the sample does not add appreciable surface reflection or scattering. Measuring otherwise identical samples of two thicknesses separates a loss growing with path length from a nearly constant contribution at the surfaces, using the same ring-down protocol.

Study the full derivation

Measure rotation with two waves following the same path · Optical fields, instruments and observations

A stationary optical loop gives equal return times for opposite propagation directions in a reciprocal medium. When the loop rotates the beams meet their starting beamsplitter at different times. Derive the measurable differential phase and compare a passive interferometer with a ring resonator.

Chapter 10 · How a grating separates wavelengths

Derive the angular dispersion and resolving power of a finite grating and diagnose an order overlap.

FROM FIRST PRINCIPLES

Derive a grating peak from equal phase steps

Take N equally spaced, equally illuminated grooves of pitch d under normal monochromatic illumination. Observe in the far field at angle θ to the grating normal. Neighbouring positions produce a path difference $d\sin\theta$. Ignoring the common single-groove envelope for the moment, each groove contributes an equal field with phase increment δ. Their sum is

$$\delta = \frac{2\pi d}{\lambda}\sin\theta,$$

$$U = U_1\sum_{j=0}^{N-1} e^{ij\delta} = U_1 e^{i(N-1)\delta/2}\frac{\sin(N\delta/2)}{\sin(\delta/2)}.$$

The fields line up at $\delta = 2\pi m$, giving $d\sin\theta = m\lambda$. The nearest zero requires an additional phase step of $2\pi/N$. Convert that change to a small angle and compare with the angular separation of two neighbouring wavelengths at fixed order.

$$\Delta\theta_{\text{zero}} \simeq \frac{\lambda}{Nd\cos\theta},$$

$$\Delta\theta_{\lambda} \simeq \frac{|m|\Delta\lambda}{d\cos\theta}.$$

Equating the separations gives the ideal Rayleigh resolving power $\lambda/\Delta\lambda = |m|N$ for nonzero order. It is the finite illuminated groove count that fixes the intrinsic width; optical aberrations, slit width and pixel sampling enter a measured line profile separately.

DERIVATION AND APPLICATIONS

Convert an entrance-slit width into spectral broadening

An ideal grating can distinguish wavelengths separated by about $\lambda/(mN)$ in order m with N illuminated grooves. A finite entrance slit adds another width because light from different points on the slit arrives at different incidence angles. Consider a slit of full width w in the focal plane of a collimator of focal length f_c. Use small angular deviations around incidence α and diffraction angle β, both defined in the convention $m\lambda = d(\sin\alpha + \sin\beta)$ for groove spacing d. Differentiating at fixed wavelength gives

$$0 = d(\cos\alpha\, d\alpha + \cos\beta\, d\beta),$$

$$|d\beta| = \frac{\cos\alpha}{\cos\beta}|d\alpha|.$$

The angular interval incident on the grating is w/f_c. A camera lens of focal length f_{cam} converts the outgoing angular spread into a full image width W at its focal plane. Measure local displacements perpendicular to the

central diffracted ray.

$$W = f_{\mathrm{cam}} \frac{\cos\alpha}{\cos\beta} \frac{w}{f_c},$$

$$\left|\frac{dx}{d\lambda}\right| = \frac{f_{\mathrm{cam}} m}{d\cos\beta}.$$

The second relation follows by varying wavelength at fixed incidence. Dividing the slit image width by this local dispersion gives the wavelength interval corresponding to the full geometric slit image.

$$\Delta\lambda_{\mathrm{slit}} = \frac{d\cos\alpha}{m} \frac{w}{f_c}.$$

The camera focal length cancels because it magnifies both the slit image and the spectrum equally. Actual recorded lines are convolutions of the projected slit, diffraction response and detector response. Their distinct width definitions must be retained when fitting line profiles.

WORKED EXAMPLE

Identify the dominant width in a green spectrometer

Change the inputs and investigate this derivation.

For $d = 1.00$ µm, $\alpha = 0$, $m = 1$, $f_c = 100$ mm and $w = 50.0$ µm, the geometric width is 0.500 nm. At 500 nm, $N = 10{,}000$ illuminated grooves give a diffraction resolution scale of 0.0500 nm. The slit width corresponds to ten such intervals. Doubling the camera focal length would double the image width in pixels while leaving the geometric 0.500 nm wavelength interval unchanged.

SECOND APPLICATION

Size a slit for an angled input

Change the inputs and investigate this derivation.

A separate instrument uses $d = 2.00$ µm, $\alpha = 60.0^\circ$, $m = 2$ and $f_c = 200$ mm. To obtain a full geometric spectral width of 0.100 nm, invert the result: $w = m f_c \Delta\lambda / (d\cos\alpha) = 40.0$ µm. At $\lambda = 600$ nm the grating equation gives $\sin\beta = 0.600 - 0.866 = -0.266$, a physically accessible direction. This angle check matters when changing high orders or incidence settings.

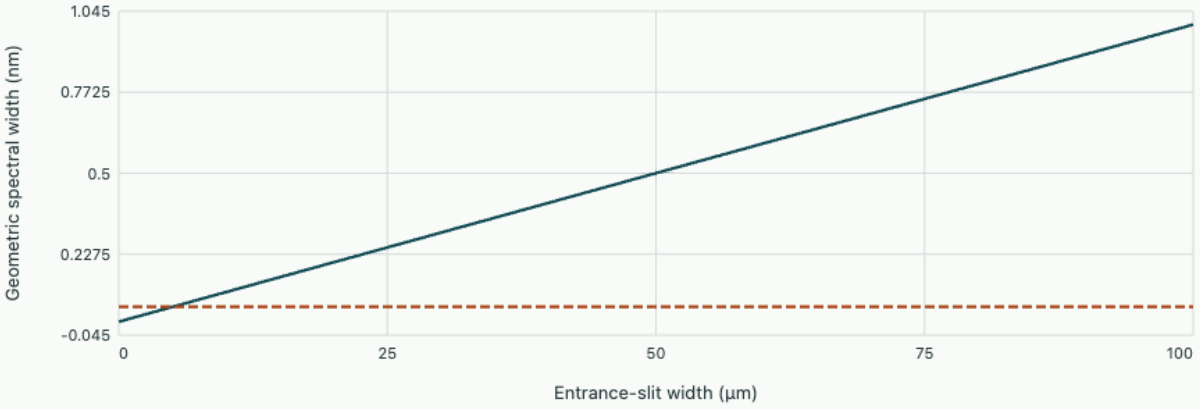
Try it · Slit resolution

Vary the physical parameters in the derived relation. Use the units and assumptions in the readout to connect the curves to the two worked examples. A parameter study can test the forward prediction; the same relation can then be inverted to infer a parameter from a measurement.

Open this live calculation

Try the calculation

Collimator focal length (mm)	100
Incidence angle (degrees)	0
Diffraction order	1



1 µm grating spacing Diffraction scale: 500 nm, 10000 grooves

Geometric width d cosα w/(m fc); diffraction gives λ/(mN). Reducing the slit width eventually reaches the diffraction floor rather than an unlimited resolving power.

Start from this chapter. Treat the two slits as the two-element limit of an array. Check how changing wavelength moves its first minimum. The numerical model has two apertures throughout, so many-groove resolving power remains a separate prediction.

OneSimulation laboratory · Find the dark fringes of a double slit

Interference across equally spaced openings

Illuminate N identical narrow slits separated by distance d with normally incident monochromatic light. In the far field every slit produces the same single-slit envelope. At observation angle θ adjacent contributions differ in optical path by $d\sin\theta$, so their phases differ by $\phi = 2\pi d\sin\theta/\lambda$. The array field equals the single-slit field multiplied by a sum of N unit phasors. Taking its geometric sum and squared magnitude gives the array intensity factor

$$G(\phi) = \left[\frac{\sin(N\phi/2)}{\sin(\phi/2)}\right]^2.$$

At $\phi = 2\pi m$ the expression has a removable limit N^2. Those directions satisfy the grating equation $d\sin\theta_m = m\lambda$. Each peak collects contributions from all illuminated slits in phase. Some orders can nevertheless be faint or missing if the single-slit envelope is small or zero there. Transmission efficiency into a selected order thus depends on the groove or opening profile in addition to the grating period.

Dispersion and finite width

At fixed order different wavelengths peak at different angles. Differentiate the grating equation to obtain

$$d\theta_m/d\lambda = m/(d\cos\theta_m).$$

A large angular dispersion aids spatial separation at the detector. To assess resolution, also compute the width of each peak. The adjacent zeros of the numerator lie one phase increment $2\pi/N$ either side of the maximum. Converting the increment back to angle gives a one-sided width $\lambda/(Nd\cos\theta_m)$.

Choose the conventional criterion that one wavelength's peak falls on the neighbour's first zero. Equate the angular separation due to $\Delta\lambda$ to this one-sided width and obtain $\lambda/\Delta\lambda = |m|N$. Broad illumination of a grating therefore improves the ideal resolving power even with unchanged dispersion. The derivation assumes uniform coherent illumination, collimation, adequate detector sampling, and an entrance slit narrow enough to preserve this diffraction-limited width.

A small spectrometer

A 600-lines/mm grating has $d = 1.667$ µm. A collimated 10.0 mm beam illuminates 6000 lines. At $\lambda = 600$ nm first order occurs at $\arcsin(0.360) = 21.1^\circ$. The ideal smallest separation is

$$600/6000 = 0.100\,\text{nm}.$$

The angular dispersion is 6.43×10^{-4} rad/nm. A camera of focal length 200 mm centred on that order maps a 0.100 nm change to about 12.9 µm near its axis. Detector pixels that are large compared with this separation reduce the measurable resolution.

At that same angle, 300 nm light in second order has the same value of $m\lambda$. Increasing the beam width narrows both peaks without separating their centres. An order-selecting filter can restrict the allowed band before measurement. A second grating or prism dispersing perpendicular to the first can instead separate overlapping orders into different traces. This illustrates why calibration with only one visible line cannot certify the identity of every line in a wide-band spectrum.

Test the finite-array argument with the two-slit acoustic field. For two equally excited slits the array factor has its second positive-angle zero at $d\sin\theta = 3\lambda/2$. With $d/\lambda = 3$, predict 30.0° and compare with the sampled 30.50° at receiver radius 0.90 m. At $d/\lambda = 2$ predict 48.59°. Compare the measured angle at both receiver radii to see how the finite observation distance affects the far-field limit.

Map overlapping orders in a high-order spectrometer

Using high diffraction order increases the ideal resolving power for a fixed number of illuminated grooves. It also reduces how much spectrum belongs unambiguously to a single order. Consider a reflection grating in a geometry where incidence and diffraction are on the same side of the normal in the usual grating-angle convention. The path difference between neighbouring grooves is $d(\sin\alpha + \sin\beta)$, so peaks obey $m\lambda = d(\sin\alpha + \sin\beta)$. In a Littrow setting the selected diffracted beam retraces the incident direction and $\alpha = \beta = \theta$.

At fixed angle two wavelengths overlap exactly if $m\lambda_m = (m+1)\lambda_{m+1}$. Solving for their separation gives $\lambda_m - \lambda_{m+1} = \lambda_m/(m+1)$, approximately λ_m/m at high order. This is the scale of a free spectral range in wavelength. Combined with resolving power mN for N uniformly illuminated grooves, the number of independently resolved spectral elements per order is approximately N. Higher order makes each element narrower while proportionately reducing the unambiguous range at a fixed spatial coordinate.

Real echelle gratings use inclined groove facets to redirect much of the reflected power into the chosen high orders. The groove spacing sets peak positions and order overlap; the groove profile sets an efficiency envelope. A second dispersive element ordinarily separates the superposed orders along another detector coordinate. Its angular dispersion can be lower than the primary grating's because it only needs to separate different orders, not resolve the closest spectral lines within one.

Design the spectral window at one detector position

Take groove spacing $d = 31.6$ µm, order $m = 100$ and central wavelength 550 nm. The Littrow relation requires $\sin\theta = m\lambda/(2d) = 55.0/63.2 = 0.870$, hence $\theta = 60.5^\circ$. Illuminating 1000 grooves gives ideal wavelength resolution $550/(100 \times 1000) = 0.00550$ nm. The wavelength in order 101 landing at the same position is

$$550(100/101) = 544.55\,\text{nm}.$$

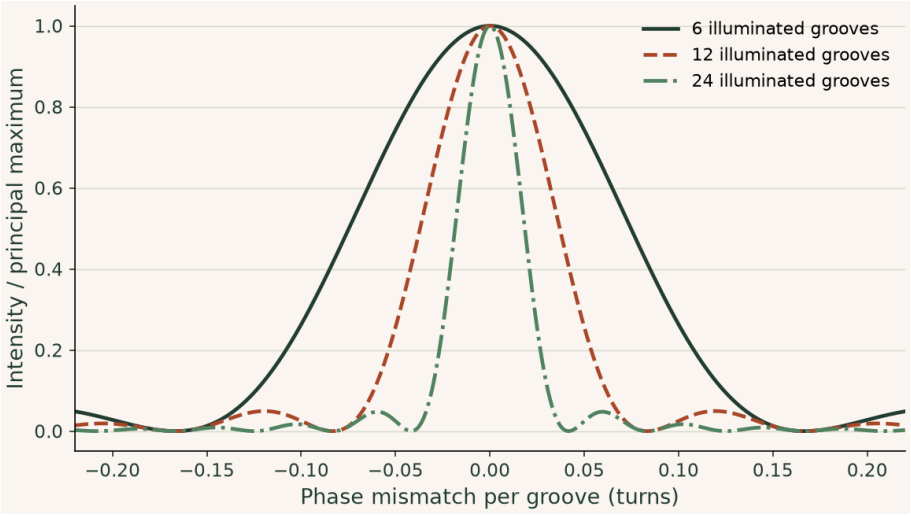
Thus the separation between these coincident-order wavelengths is 5.45 nm, about 990 resolution intervals.

A narrow-band filter passing 548 to 552 nm would reject that particular neighbouring order at the central position while retaining many resolvable lines near 550 nm. An instrument meant to record hundreds of nanometres at once needs cross-dispersion or several separate exposures. As incidence approaches grazing, simple angular-dispersion estimates can become very large, but finite beam footprint, usable groove count, diffraction efficiency and detector size limit the physical gain. Accounting for order separation, illuminated footprint and sampling is necessary before describing two nearby spectral features as resolvable in the laboratory.

A rough detector-format estimate follows directly from this example. If each independent spectral resolution interval requires at least two samples, one full free spectral range near the chosen order needs about two thousand useful pixels along the high-dispersion direction. That count ignores additional margins, optical distortion and the need to sample the entrance-slit profile adequately. The perpendicular axis must also have room for the traces of neighbouring orders and for a background estimate between them. Resolution, spectral coverage and detector format therefore impose independent constraints on what a single exposure can realistically measure.

Separate groove count from grating period

Keep the grating period and incident wavelength fixed. Adding N equal amplitudes as a geometric series gives a central peak whose first zeros occur at phase mismatch $\pm 1/N$ in turns. Each plotted peak is normalised to its own height.



Doubling illuminated groove count halves the first-zero width without moving the principal order. Normalising hides the N^2 increase in peak intensity for fixed incident field per groove; spectral width and absolute throughput answer different questions.

Try it · Grating

Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Illuminated grooves 20



Finite uniform grating

First zeros at ±2π/N = ±0.31416 radians. Absolute peak intensity scales as N² at fixed field per groove.

▶ **Read values from the plot**

Problem sheet 5 · Orders, linewidths and spectral identity

Use cavity and grating data to separate linewidth, resolving power and overlap of different spectral orders.

Infer mirror reflectivity

A symmetric, lossless air cavity 30.0 mm long has a transmission peak full width of 100 MHz. Determine the free spectral range and finesse. Estimate R using the high-finesse formula and predict how doubling cavity length changes width.

▼ **Show the cavity inference**

The free spectral range is $c/(2L) = 5.00$ GHz and finesse is approximately 50. Writing $x = \sqrt{R}$, the approximation $\mathcal{F} = \pi x/(1 - x^2)$ gives

$$x = [-\pi + \sqrt{\pi^2 + 4\mathcal{F}^2}]/(2\mathcal{F}) = 0.969.$$

Thus $R \simeq 0.939$. For unchanged mirrors and negligible dispersion the finesse stays constant. Doubling length halves both free spectral range and linewidth, to 2.50 GHz and 50 MHz.

Select enough grooves

A first-order grating must separate 589.0 and 589.6 nm by the Rayleigh criterion. Determine the minimum illuminated number of grooves and the width at 1200 grooves/mm. Calculate the first-order diffraction angle near the doublet centre.

▼ **Show the finite-array calculation**

Using 589.3 nm as representative wavelength gives $N \geq 589.3/0.6 = 982.2$, so illuminate at least 983 grooves. Their span is approximately $983/1200 = 0.819$ mm. The period is 833.3 nm. The grating equation gives $\sin\theta = 589.3/833.3 = 0.7072$, hence $\theta \simeq 45.0^\circ$. Detector sampling, entrance-slit size and unequal line intensities may require a larger illuminated width in practice.

Identify two peaks in one setting

A cavity with 8.0 GHz free spectral range transmits two optical frequencies separated by exactly one order at around 600 nm. Determine their wavelength separation to first order. Find the first-order grating resolving power needed to distinguish them and explain what separate measurements determine.

▼ **Show the order ambiguity**

From $\nu = c/\lambda$, $|\Delta\lambda| \simeq \lambda^2\Delta\nu/c = 0.0096$ nm. Their ratio $\lambda/|\Delta\lambda| = 62\,500$ requires about 62,500 uniformly illuminated grooves in first order under the ideal criterion. A narrow cavity scan can measure lines within one order very accurately but maps frequencies apart by the free spectral range to the same phase. A

sufficiently resolving independently calibrated grating can distinguish the coincident orders. A second cavity of different spacing can also change their relative alignment.

Explore these relations

- For the mirror inference, vary reflectivity and compare the cavity linewidth and free spectral range. Open the live calculation.
- For groove counts and overlapping orders, vary groove number and compare the interference peaks. Open the live calculation.

Chapter 11 · Gaussian beams and focusing

Relate beam waist, wavefront curvature and divergence, and estimate the focus of a collimated laser beam.

FROM FIRST PRINCIPLES

Obtain a Gaussian divergence from its waist profile

Take a monochromatic fundamental beam in a uniform medium at a waist plane with flat phase front. Let the transverse field be $U(x,0) = U_0 e^{-x^2/w_0^2}$ on one diameter. Scalar diffraction represents this field as a superposition of plane waves with transverse wavenumber q. The Fourier transform of a Gaussian is again a Gaussian, with reciprocal width.

$$\widetilde{U}(q) \propto \int_{-\infty}^{\infty} e^{-x^2/w_0^2} e^{-iqx}\, dx \propto e^{-q^2 w_0^2/4}.$$

For a paraxial plane wave $q = k\sin\theta \simeq k\theta$, with $k = 2\pi/\lambda$. In the far field each angle addresses its corresponding plane-wave amplitude. The $1/e$ field half-angle, also the $1/e^2$ intensity half-angle, therefore satisfies

$$\frac{k^2\theta_0^2 w_0^2}{4} = 1,$$

$$\theta_0 = \frac{\lambda}{\pi w_0}.$$

The characteristic propagation length is

$$z_R = w_0/\theta_0 = \pi w_0^2/\lambda.$$

The paraxial Gaussian solution gives $w(z) = w_0\sqrt{1 + (z/z_R)^2}$, approaching $\theta_0|z|$ far from the waist. A tighter waist thus brings both larger angular spread and shorter depth of focus.

Background · Gaussian Rayleigh range

$$z_R = \pi w_0^2/\lambda$$

Here the waist is the one-over-e field radius and the wavelength is in the medium. At one Rayleigh range its squared beam radius doubles.

DERIVATION AND APPLICATIONS

Measure a Gaussian beam with a translating knife edge

A camera measures an intensity profile directly. A power meter behind a straight edge instead integrates the light that passes the edge. For a centred, circular fundamental Gaussian at one measurement plane, write $I(x,y) = I_p \exp[-2(x^2+y^2)/w^2]$, where w is the $1/e^2$ intensity radius. Assume that the power meter collects all transmitted light, the edge is opaque and straight, and diffraction at the edge does not send appreciable power outside the detector. With the opening extending from $x = b$ to infinity, integration over y first gives

$$P(b) = I_p w \sqrt{\frac{\pi}{2}} \int_b^{\infty} e^{-2x^2/w^2}\, dx,$$

$$P_{\text{tot}} = \frac{\pi w^2 I_p}{2}.$$

Change variables to $u = \sqrt{2}x/w$ and normalise to the unobstructed reading. This step removes the absolute intensity calibration and gives a cumulative Gaussian curve in the measured edge position.

$$\frac{P(b)}{P_{\text{tot}}} = \frac{1}{2}\operatorname{erfc}\left(\frac{\sqrt{2}\,b}{w}\right).$$

At the half-power point the edge passes through the beam centre. At the ten and ninety percent positions the arguments of the complementary error function have equal magnitudes and opposite signs. Their separation is consequently independent of the arbitrary origin of the translation stage.

$$b_{10} - b_{90} = \sqrt{2}\, w\ \operatorname{erf}^{-1}(0.8) = 1.28155\, w.$$

The derivative of the measured power curve is proportional to the intensity integrated along the blade, which gives another way to inspect asymmetry. For an elliptical Gaussian aligned with the axes, translation in x measures the x radius, and rotating the blade allows the other radius to be tested.

WORKED EXAMPLE

Infer a beam radius without a calibrated camera

Change the inputs and investigate this derivation.

An edge produces 90.0 percent transmission at 1.24 mm and 10.0 percent at 1.88 mm. Their separation gives

$$w = 0.640/1.28155 = 0.499\,\text{mm}.$$

The centre lies at their midpoint, 1.56 mm. An unobstructed reading of 2.00 mW then gives $I_p = 2P_{\text{tot}}/(\pi w^2) = 5.11$ kW/m² for a circular beam.

SECOND APPLICATION

Locate a useful aperture radius

Change the inputs and investigate this derivation.

A different circular beam has $w = 0.250$ mm. A straight edge at $b = w$ would transmit $\operatorname{erfc}(\sqrt{2})/2 = 0.0228$ of its power. A centred circular aperture of that radius transmits $1 - \exp(-2a^2/w^2) = 0.865$ after radial integration. To pass 99.0 percent, its radius must instead be $a = w\sqrt{\ln 100/2} = 0.379$ mm. The different answers follow from the areas collected by a half-plane and a disk.

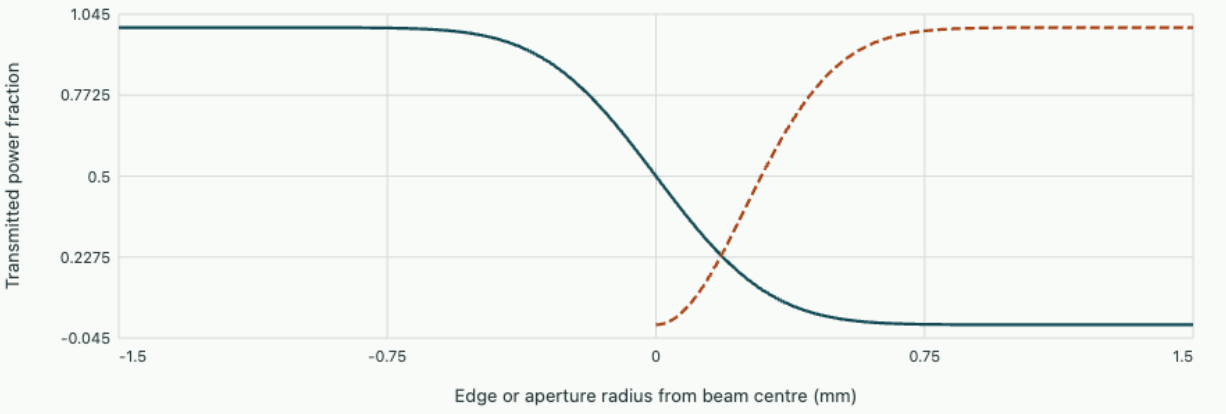
Try it · Knife edge

Vary the physical parameters in the derived relation. Use the units and assumptions in the readout to connect the curves to the two worked examples. A parameter study can test the forward prediction; the same relation can then be inverted to infer a parameter from a measurement.

Open this live calculation

Try the calculation

Gaussian field radius (mm) **0.5**



Knife-edge 10–90% width = 1.2815516w = 0.64078 mm. A circular aperture integrates a different geometry from a straight edge.

Measure the phase that a waist adds to a Gaussian beam

Two axial planes on opposite sides of a waist can have identical intensity and different phase. An interferometer comparing the field with a calibrated reference can distinguish them. Predict the sign and size of the residual phase before drawing the quadrature traces.

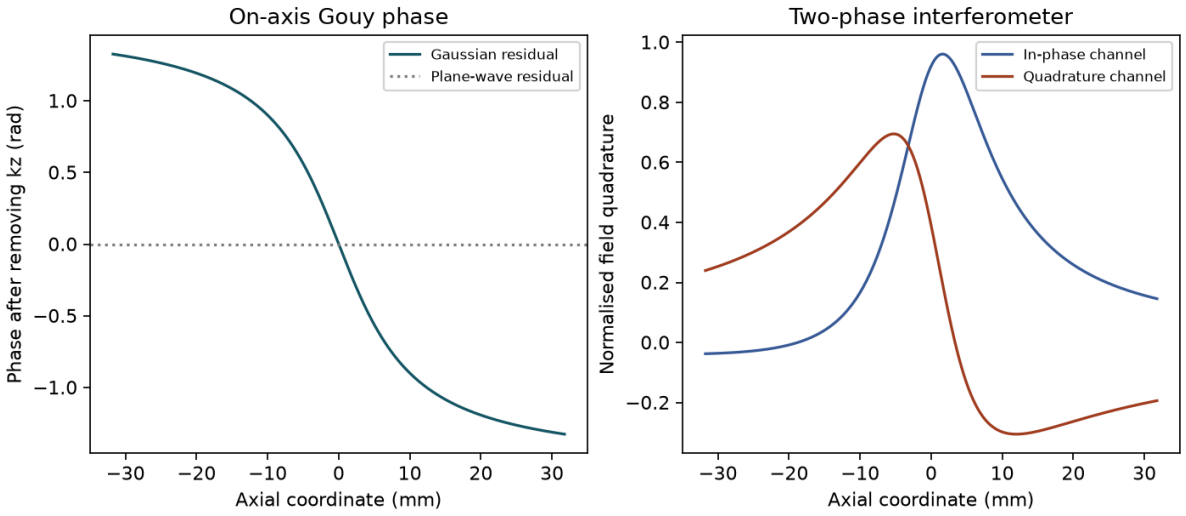
Factor out ordinary plane-wave propagation

For a fundamental paraxial Gaussian beam with waist field proportional to $\exp(-r^2/w_0^2)$, the transverse angular spectrum is a Gaussian in transverse wave number. Each component propagates with an additional paraxial phase $\exp[-izk_\perp^2/(2k)]$ after removing the common e^{ikz} carrier. On the axis one integrates over both transverse wave-vector components. Each Gaussian integral contributes an inverse square root of a complex width; their product gives

$$\frac{U(0,z)}{U(0,0)e^{ikz}} = \frac{1}{1+iz/z_R},$$

$$z_R = \pi w_0^2/\lambda.$$

The residual phase with this travelling-wave convention is $-\arctan(z/z_R)$. Its difference between positions very far upstream and downstream approaches $-\pi$. Its difference across a finite focus is generally smaller. At $z = \pm z_R$ the residual phases are $\mp\pi/4$; relative to a plane wave the focused beam loses a total of $\pi/2$ across those two points. The intensity, by contrast, is $1/[1+(z/z_R)^2]$ in units of the waist intensity, which has identical values at the two positions.



The same waist size sets the longitudinal scale in both panels. The readout phase offset rotates the two quadratures without changing the physical axial phase difference.

Try it · Gouy

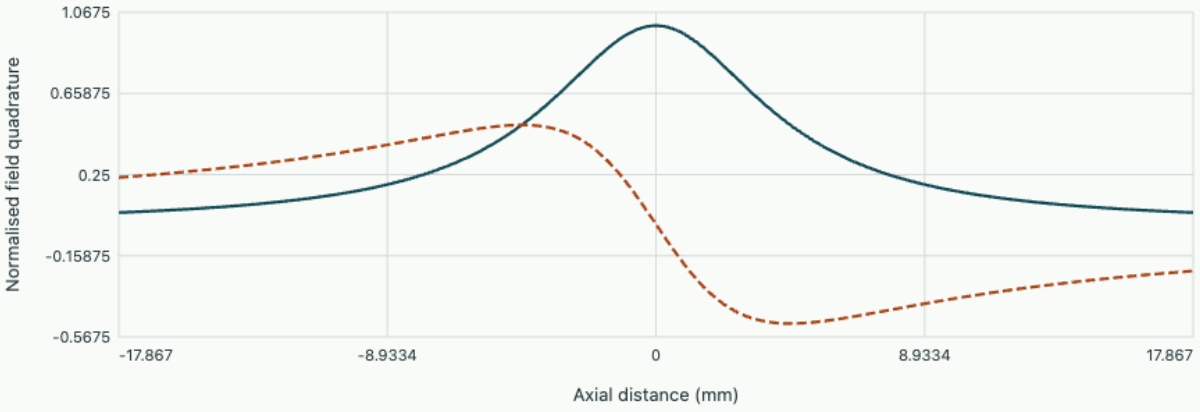
Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Waist radius (µm) **30**

Reference phase (degrees) **0**



Wavelength 633 nm; Rayleigh range 4.4667 mm. After removing kz the phase is –atan(z/zR); reference phase rotates the measured quadratures.

Download the corresponding editable notebook. The browser recalculates the reduced equations directly; the notebook exposes the full calculation and its checks.

For $w_0 = 40$ micrometres and $\lambda = 0.633$ micrometre in the medium, $z_R = 7.94$ mm. The axial phase gradient at the waist is smaller than the carrier wave number by $1/z_R$, about $126\,\mathrm{rad/m}$. An ordinary path-length comparison over millimetres must remove many carrier cycles to see this residual accurately. Measuring visibility alone will not do so. With a reference field E_r, the interference term in intensity is $2\,\mathrm{Re}(E_s E_r^*)$. Shifting the reference phase by a known quarter turn gives the complementary quadrature. After subtracting backgrounds and calibrating gains one reconstructs phase by the signed two-argument arctangent and unwraps it through adjacent positions.

The notebook applies an adjustable constant phase offset to both synthetic channels. It disappears from a calibrated phase difference. Doubling the waist multiplies the Rayleigh length by four. The total residual phase across the complete focus is still $-\pi$, but a scan over a fixed physical distance sees a smaller fraction of it. Wavefront curvature also contributes a phase away from the axis. Keep the detector on axis, or model its transverse position and aperture before attributing a measured phase entirely to the Gouy shift.

The residual Gouy phase changes by $-\pi/3$ between symmetric planes $z = -5$ mm and $z = +5$ mm. Infer the Rayleigh length and the waist radius at 0.633 micrometre wavelength.

▼ **Work through the solution**

Symmetry gives $2\arctan(5\ \mathrm{mm}/z_R) = \pi/3$, hence

$$z_R = 5/\tan(\pi/6) = 8.66\,\mathrm{mm}.$$

The waist radius is $\sqrt{\lambda z_R/\pi}$, about 41.8 micrometres. This inference uses the one-over-e field radius of a fundamental circular Gaussian beam and a phase referenced to the same plane-wave propagation.

Open editable notebook Download .ipynb

Diffraction during propagation

A narrow parallel-looking laser beam eventually spreads. Use a scalar monochromatic field $U = u(x, y, z)e^{ikz}$, with u changing slowly on the scale of a wavelength. Substitution into the homogeneous Helmholtz equation cancels the largest longitudinal terms. Dropping $\partial_z^2 u$ relative to $2ik\partial_z u$ gives the paraxial equation

$$2ik\,\partial_z u + \nabla_\perp^2 u = 0.$$

This approximation requires a small angular spectrum about the optical axis. The transverse Fourier component with wavevector $\mathbf{q}$ acquires phase $\exp[-iq^2 z/(2k)]$. A Gaussian waist field $u(r, 0) = u_0 e^{-r^2/w_0^2}$ has a Gaussian angular spectrum proportional to $\exp(-w_0^2 q^2/4)$. On propagation, multiplication by the quadratic phase keeps that spectrum Gaussian with a complex width.

Invert the Gaussian transform

Collect its exponent as $-q^2(w_0^2/4 + iz/2k)$ and invert the transverse Fourier transform. The reciprocal complex width determines both the real amplitude decay in r and the phase curvature. Separating the real and imaginary parts gives

$$w(z) = w_0\sqrt{1 + (z/z_R)^2},$$

$$R(z) = z[1 + (z_R/z)^2],$$

$$z_R = \pi w_0^2/\lambda.$$

The convention here defines w at $1/e$ of field amplitude and $1/e^2$ of on-axis intensity. The on-axis intensity falls as w_0^2/w^2 so the transverse integral of intensity stays constant. At a Rayleigh range from the waist the area doubles and the axial intensity halves. At large distance the half-angle divergence approaches $\theta = \lambda/(\pi w_0)$. Waist size and angular spread therefore cannot be chosen independently for this fundamental mode.

OneSimulation · Two identical intensity photographs, two different waists

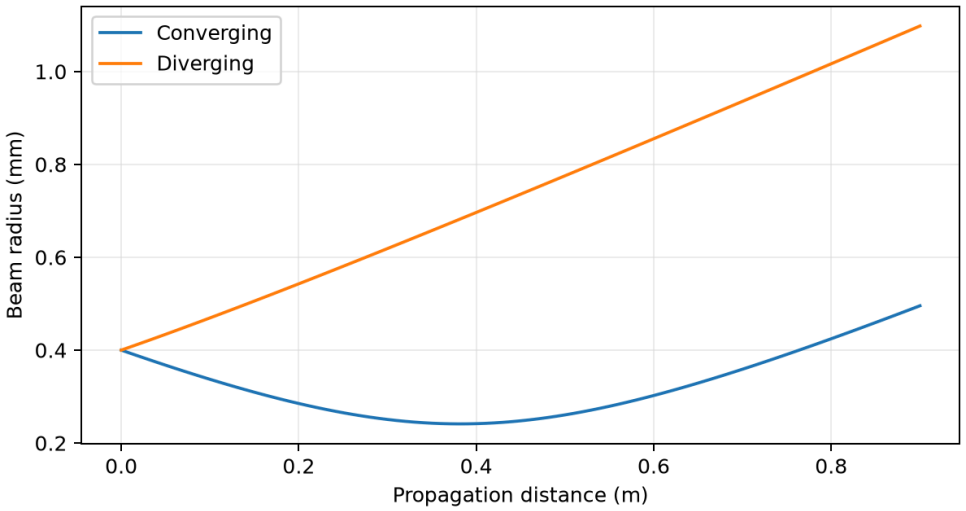
Give a 633 nm Gaussian beam a 0.40 mm field radius and take an intensity photograph at $z = 0$. Reversing the sign of its quadratic phase leaves every pixel of that photograph unchanged. The signed wavefront radius R, however, sets the local tilt of the wave normal. In the paraxial limit the predicted radius is

$$w^2(z) = w_0^2\left[\left(1 + \frac{z}{R}\right)^2 + \left(\frac{z}{z_R}\right)^2\right],$$

$$z_R = \frac{\pi w_0^2}{\lambda}.$$

Here w_0 names the radius in the photograph. The true waist can lie elsewhere. Propagate both complex fields with OneSimulation's angular spectrum. Each sampled transverse spatial frequency acquires its own longitudinal phase. Measure the radius as twice the square root of the intensity-weighted second moment and compare with the paraxial prediction.

```
"""Propagate two Gaussian fields with the same measured input intensity."""
import matplotlib.pyplot as plt
import numpy as np
from onesim.physics.optics.propagation.numerics.angular_spectrum import (
    AngularSpectrumPropagation, longitudinal_wavenumber,
)

wavelength, width = 633e-9, 0.40e-3  # m; width is the 1/e field radius
k = 2 * np.pi / wavelength
rayleigh = np.pi * width**2 / wavelength
x = np.linspace(-6e-3, 6e-3, 2048, endpoint=False)
kx = 2 * np.pi * np.fft.fftfreq(len(x), d=x[1] - x[0])
propagate = AngularSpectrumPropagation(longitudinal=longitudinal_wavenumber(kx[:, None], k),
    forward=np.fft.fft, inverse=np.fft.ifft, reference_wavenumber=k)
z = np.linspace(0, 0.90, 91)
fig, ax = plt.subplots(figsize=(6.6, 3.5), layout="constrained")
for curvature, label in [(-0.60, "Converging"), (0.60, "Diverging")]:
    field = np.exp(-(x / width)**2 + 1j * k * x**2 / (2 * curvature))
    radii = []
    for distance in z:
        intensity = np.abs(propagate(field, distance))**2
        radii.append(2 * np.sqrt(np.average(x**2, weights=intensity)))
    reference = width * np.sqrt((1 + z / curvature)**2 + (z / rayleigh)**2)
    assert np.allclose(radii, reference, rtol=2e-5, atol=0)
    print(f"{label}: narrowest sampled radius {1000*min(radii):.3f} mm at {z[np.argmin(radii)]:.2f} m")
    ax.plot(z, 1000 * np.asarray(radii), label=label)
ax.set(xlabel="Propagation distance (m)", ylabel="Beam radius (mm)")
ax.grid(alpha=0.2)
ax.legend()
fig.savefig("waves-optics.png", dpi=180)
```



Wavefront radii of opposite sign and equal magnitude 0.60 m give different downstream widths. The converging beam reaches its narrowest sampled section 0.38 m beyond the camera.

Try it · Gaussian curvature

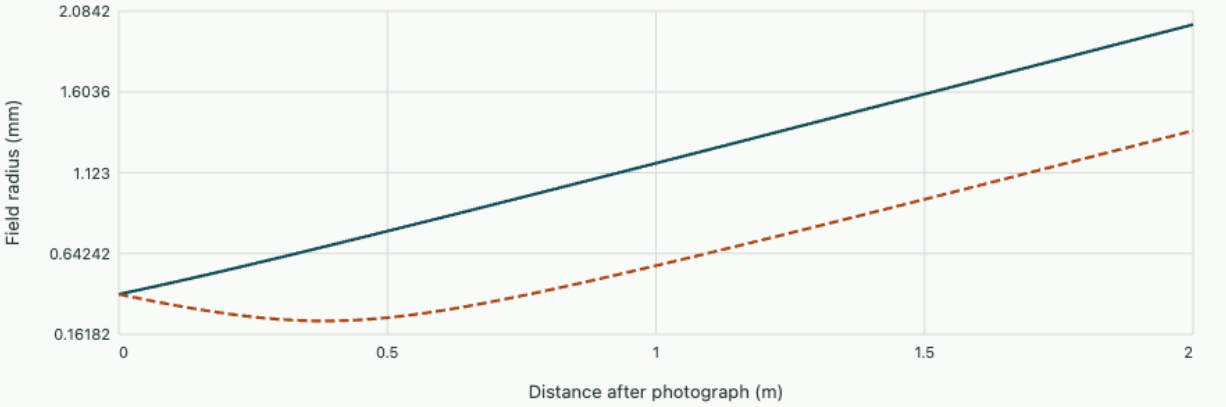
Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Photographed field radius (mm) **0.4**

Magnitude of wavefront radius (m) **0.6**



Initially diverging phase Initially converging phase

Identical initial intensities, opposite quadratic phases. Converging waist at z = 0.38194 m. Paraxial Gaussian model at 633 nm.

Download the Python example

When does a beam propagate without diffracting?

The Gaussian beam spreads freely because its angular components develop different longitudinal phases. A dielectric guide can instead admit a transverse field that reproduces itself times a single propagation phase. Take a symmetric slab of core index n_1, cladding index $n_2 < n_1$, and core extending from $x = -a$ to a. For electric polarization parallel to the interfaces write $E_y = X(x)e^{i(\beta z - \omega t)}$. The scalar wave equation gives

$$X'' + (n^2 k_0^2 - \beta^2)X = 0.$$

For a guided mode require oscillation in the core and decay in the cladding, so $n_2 k_0 < \beta < n_1 k_0$.

The even solution is proportional to $\cos(hx)$ in the core and $e^{-\gamma(x-a)}$ in the right cladding. Continuity of X and its derivative for equal magnetic permeabilities requires

$$h \tan(ha) = \gamma.$$

Define $u = ha$, $w = \gamma a$ and $V = k_0 a \sqrt{n_1^2 - n_2^2}$. Then $u^2 + w^2 = V^2$ and $u \tan u = w$. The first even solution exists for any positive V and becomes weakly confined as V tends to zero. Odd solutions instead obey $-u \cot u = w$, with the first appearing above $V = \pi/2$.

For $\lambda_0 = 1.00$ µm, $n_1 = 1.50$ and $n_2 = 1.40$, choose $a = 0.328$ µm. This gives $V \simeq 1.11$, for which $u = w \simeq \pi/4$ satisfies both equations. The effective index is $\beta/k_0 = \sqrt{n_2^2 + (w/ak_0)^2} = 1.451$, between the two material indices. The cladding amplitude decays over $a/w = 0.418$ µm. This mode can have a fixed width along a uniform guide. An arbitrarily shaped input can excite radiation or several modes and need not retain its initial profile.

Practise on sheet 6 · Choose a slab width for one transverse TE mode. Further reading · Massachusetts Institute of Technology · Haus and Melcher, Electromagnetic Fields and Energy · §13.5 Dielectric Waveguides: Optical Fibers, Figures 13.5.1–13.5.4.

A focused spot with finite depth

A collimated ideal Gaussian at 633 nm has $1/e^2$ intensity radius $w_{\mathrm{in}} = 1.00$ mm at a lens of focal length 100 mm. Assume negligible clipping and a nearly flat incident wavefront. The lens converts input radius into convergence angle $w_{\mathrm{in}}/f = 0.0100$ rad. Equating this angle to the far-field Gaussian divergence gives $w_0 \simeq \lambda f/(\pi w_{\mathrm{in}}) = 20.1$ µm. The focused Rayleigh range is 2.01 mm. At axial displacement 4.02 mm the spot radius grows to about 45.0 µm and the axial intensity drops to a fifth.

Increasing the input radius by a factor of two halves the waist while reducing Rayleigh range by four. This trade matters when illuminating a sample extending several millimetres along the beam. A hard aperture that clips substantial power reshapes the angular spectrum, so this simple product no longer predicts the entire spot. A measured waist and divergence with product exceeding λ/π can also signal imperfect mode quality. Radius measurements must use the same intensity convention before making that comparison.

Two radius measurements equidistant from a proposed waist should give the same result for an ideal beam. If they do not, first refine the estimate of waist position. Multiple axial samples can be fitted to $w^2(z) = w_0^2[1 + (z - z_0)^2/z_R^2]$ to infer the waist plane and divergence simultaneously. Power falling through

an aperture, however, also depends on the beam centre. A beam walking sideways past a fixed detector can imitate a growing radius unless beam position is measured independently. Recording transverse profiles instead of only total collected power makes this distinction observable.

Propagate a Gaussian that already has curved wavefronts

An input to a lens is often converging or diverging before reaching the lens. Its radius alone then cannot determine the new waist. Combine wavefront curvature and intensity radius in a complex beam parameter q defined by $1/q = 1/R - i\lambda/(\pi w^2)$. In homogeneous space the solution to the paraxial wave equation has $q = z - z_0 + i z_R$, where z_0 is the waist plane and $z_R > 0$ the Rayleigh range. Straight propagation through distance d therefore adds d to q.

A thin positive lens multiplies the field by a quadratic phase that changes curvature according to $1/R_{\mathrm{out}} = 1/R_{\mathrm{in}} - 1/f$ and leaves transverse intensity unchanged at the lens. Combining its two effects in the definition gives

$$\frac{1}{q_{\mathrm{out}}} = \frac{1}{q_{\mathrm{in}}} - \frac{1}{f}.$$

More generally free propagation and lenses act through the same real ray matrices used for first-order geometrical optics. Composing their actions on q gives

$$q_{\mathrm{new}} = (A q + B)/(C q + D).$$

To locate a subsequent waist in free space, propagate until the real part of q is zero. Its positive imaginary part then supplies z_R, hence the new waist radius. This construction retains finite diffraction and avoids predicting a singular geometric focus.

Reconstruct a focus downstream of a displaced lens

An ideal beam of wavelength 633 nm has an initial Rayleigh range 100 mm. Its input waist is 50.0 mm upstream of a lens of focal length 100 mm. At the lens $q_{\mathrm{in}} = 50.0 + i100$ mm, so $1/q_{\mathrm{in}} = 0.00400 - i0.00800$ mm⁻¹. Subtracting $1/f = 0.0100$ mm⁻¹ gives $1/q_{\mathrm{out}} = -0.00600 - i0.00800$ mm⁻¹. On inversion $q_{\mathrm{out}} = -60.0 + i80.0$ mm. The waist is 60.0 mm after the lens and has Rayleigh range 80.0 mm.

At this new waist $w_0 = \sqrt{\lambda z_R/\pi} = 0.127$ mm. At the lens the outgoing parameter predicts $w = w_0\sqrt{1 + (60/80)^2} = 0.159$ mm, the same incident radius obtained from the original 100 mm Rayleigh range and 50 mm propagation. This consistency check tests both the signs and the continuity of intensity. Treating the incident beam as fully collimated would place the waist close to the focal plane and miss its measured position by tens of millimetres. In the limit $f \to \infty$ the reciprocal correction vanishes and the original waist and curvature are recovered. A measured beam profile at several axial planes can distinguish this finite-curvature effect from a simple error in the stated lens focal length.

The complex parameter can be recovered experimentally from intensity alone if enough axial profiles are recorded. Fit squared radius to a quadratic in position, locate its minimum to obtain the waist plane, and infer Rayleigh range from how rapidly that squared radius increases on either side. A separately measured

wavelength then predicts the relation between fitted waist radius and divergence for a fundamental Gaussian. If the fitted values violate that relation beyond uncertainty, the more likely problems include a higher-order mode contribution, clipping or an inconsistent radius definition. Simply adjusting the lens focal length cannot repair all three discrepancies at once.

Study the full derivation

Build a time lens from dispersion and a quadratic temporal phase · Optical fields, instruments and observations

A temporal imaging system can stretch the time separation of two optical features without requiring them to arrive more slowly at the source. Derive the imaging condition from delay and frequency, calculate the sign of time magnification, and estimate the finite temporal aperture that limits which features can be recovered.

Chapter 12 · Measuring a star with interference

Infer angular diameter or binary separation from the spatial coherence of light at two separated collectors.

FROM FIRST PRINCIPLES

Connect a projected baseline to source coherence

Take two collectors observing a distant thermal source in a narrow wavelength band. Let θ be the small angular offset of one source element along the projected baseline of length B. Its geometric path difference between collectors, after removing the delay to the field centre, is approximately $B\theta$. At wavelength λ this produces the phase difference

$$\phi(\theta) = 2\pi B\theta/\lambda.$$

Different source elements fluctuate independently, so their correlations add with weights proportional to brightness. If $S(\theta)$ is brightness integrated in the perpendicular angular direction, dividing by total brightness gives normalised complex coherence.

$$\gamma(B) = \frac{\int S(\theta) e^{i2\pi B\theta/\lambda}\, d\theta}{\int S(\theta)\, d\theta}.$$

For two equally bright point sources at projected angles $\pm s/2$, this sum has just two phasors. Their opposite phases make the normalised visibility real.

$$\gamma(B) = \tfrac{1}{2}(e^{i\pi Bs/\lambda} + e^{-i\pi Bs/\lambda}) = \cos(\pi Bs/\lambda).$$

The first contrast zero occurs at $B = \lambda/(2s)$ for equal arm intensities. Rotating the projected baseline changes the separation component being measured. A source extended perpendicular to one baseline can therefore look unresolved there and extended on another.

DERIVATION AND APPLICATIONS

Infer the orientation-dependent size of a Gaussian source

A star surrounded by elongated emission can have a visibility that decays without crossing zero. Model its sky brightness as an elliptical Gaussian with full widths at half maximum Θ_x, Θ_y along known perpendicular sky axes. Assume spatially incoherent emission, a narrow spectral channel of wavelength λ, small angles in radians and far-field collectors. At projected baseline components (B_x, B_y), the normalised visibility is the two-dimensional Fourier transform of the brightness. The Gaussian factors separate along the two axes.

$$S(x,y) = S_0 \exp\left[-4\ln 2\left(\frac{x^2}{\Theta_x^2} + \frac{y^2}{\Theta_y^2}\right)\right],$$

$$u = \frac{B_x}{\lambda},$$

$$v = \frac{B_y}{\lambda}.$$

For each axis complete the square in the Fourier integral. Dividing by the integral at zero baseline cancels the normalisation of the Gaussian. Taking the product of the two factors then gives a positive real visibility for a source centred at the coordinate origin.

$$\frac{\int e^{-ax^2} e^{2\pi i u x}\, dx}{\int e^{-ax^2}\, dx} = e^{-\pi^2 u^2/a},$$

$$|\gamma| = \exp\left[-\frac{\pi^2(\Theta_x^2 B_x^2 + \Theta_y^2 B_y^2)}{4\ln 2\, \lambda^2}\right].$$

For a baseline of length B at angle ψ to the x axis, define the projected squared width by weighting the two squared widths by the directional cosines squared. Solving for the baseline that halves the visibility gives

$$\Theta_{\text{proj}}^2 = \Theta_x^2 \cos^2\psi + \Theta_y^2 \sin^2\psi,$$

$$B_{1/2} = \frac{2\ln 2}{\pi}\frac{\lambda}{\Theta_{\text{proj}}}.$$

A larger projected source loses fringe contrast at a shorter baseline. The half point remains available as an observable even though an ideal Gaussian has no exact visibility zero at finite baseline.

WORKED EXAMPLE

Read two angular widths from a rotating baseline

Change the inputs and investigate this derivation.

At 1.60 µm, calibrated visibility reaches 0.500 at 40.0 m along one sky axis and at 80.0 m along the perpendicular axis. The widths are $0.4413(1.60 \times 10^{-6})/40.0 = 1.77 \times 10^{-8}$ rad and half that value. Multiplication by 2.063×10^{8} milliarcseconds per radian gives 3.64 and 1.82 milliarcseconds. The longer half-visibility baseline corresponds to the narrower source axis.

SECOND APPLICATION

Predict a diagonal baseline reading

Change the inputs and investigate this derivation.

For another source, independently measured half-visibility baselines are 30.0 m and 60.0 m on the principal axes in a single spectral channel. At 45.0° the reciprocal squared half-baseline is $(1/30^2 + 1/60^2)/2$, giving 37.9 m. For a 60.0 m diagonal baseline the predicted visibility is

$$2^{-(60.0/37.9)^2} = 0.177.$$

This diagonal reading tests the elliptical model against information beyond the two fitted widths.

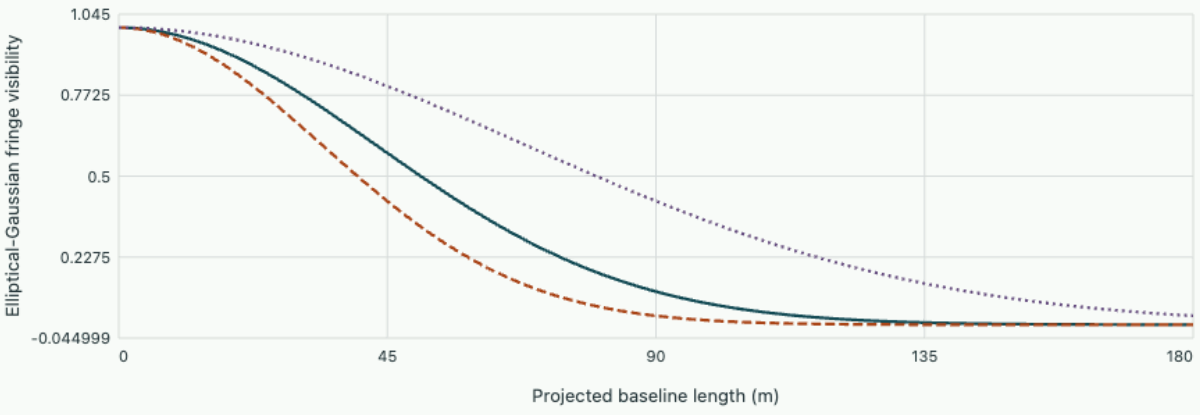
Try it · Stellar ellipse

Vary the physical parameters in the derived relation. Use the units and assumptions in the readout to connect the curves to the two worked examples. A parameter study can test the forward prediction; the same relation can then be inverted to infer a parameter from a measurement.

Open this live calculation

Try the calculation

Half-visibility baseline on x (m)	40
Half-visibility baseline on y (m)	80
Baseline direction (degrees)	45



Half-visibility baseline along selected direction 50.596 m. Gaussian intensity widths and projected baseline enter through their squared principal-axis components.

When two copies cease to interfere

A narrow-band point source sends essentially the same waveform to two distant collectors after the proper path delay. Two extended sources can light both collectors equally while producing different fringe patterns. The measurable quantity is the complex degree of coherence γ, normalized by the geometric mean of the two intensities. If the two collected beams have equal intensities, controlled recombination with variable phase produces fringe visibility $|\gamma|$. Unequal intensities reduce visibility by an additional known balance factor.

Consider collectors separated by projected baseline B transverse to their line of sight. A source element at a small angular coordinate θ along the baseline has geometric delay $B\theta/c$ and phase $2\pi B\theta/\lambda$. Different

surface elements of a thermal source radiate independently on average, so cross terms between distinct elements disappear. Each element still interferes with itself through the two collectors. Adding their complex interference contributions, weighted by angular intensity S, gives

$$\gamma(B) = \frac{\int S(\theta) e^{2\pi i B\theta/\lambda}\, d\theta}{\int S(\theta)\, d\theta}.$$

Here S is angular brightness integrated perpendicular to the baseline. Successive baselines sample different Fourier components of the projected brightness profile. The narrow-band, far-field conditions let one use a single wavelength and a planar angular phase across the entire source.

Two stars and a disk

For equally bright unresolved points at angles $\pm s/2$, the normalized sum is $\gamma = \cos(\pi B s/\lambda)$. Its first zero occurs at $B = \lambda/(2s)$. Rotating the baseline changes the projected separation and can restore visibility. If one point is brighter, the unequal phasor lengths cannot cancel completely. For a uniform circular disk of angular diameter Θ the two-dimensional brightness integral instead gives $2J_1(x)/x$ with $x = \pi B\Theta/\lambda$. The first zero occurs at $B = 1.22\lambda/\Theta$. The same Bessel function appeared for a filled circular aperture, now describing coherence as a function of collector separation.

Interpret a visibility measurement

At 550 nm a uniform disk first loses its fringes at projected baseline 50.0 m. The inferred diameter is $1.22(550 \times 10^{-9})/50.0 = 1.34 \times 10^{-8}$ rad or 2.77 milliarcseconds. At distance 10.0 parsecs its linear diameter would be 0.0277 astronomical units, about 4.14×10^{9} m. If independent spectroscopy suggested a binary made of two unresolved equals instead, the same null would correspond to projected separation 1.13 milliarcseconds. One zero by itself does not determine which brightness distribution produced it.

Measurements at several baseline lengths distinguish the cosine from the disk transform, and observations after rotation examine orientation. A broad optical passband averages different λ values and can fill a sharp null. In addition, uncontrolled fluctuations of the optical path during an exposure reduce contrast even for a point source. Observing a nearby unresolved calibrator at the same instrumental setting separates that loss from the target’s physical angular extent.

The measured baseline is a projection perpendicular to the line of sight. Earth’s rotation changes both its effective length and orientation for a fixed pair of collectors, providing additional Fourier samples over a night.

Remove telescope-dependent phase errors with three collectors

Two telescopes measure a complex visibility: its magnitude describes contrast while its argument contains the position of brightness features. Atmospheric delay at each telescope can overwhelm that phase. Use three telescopes and orient the measured baselines cyclically from 1 to 2, 2 to 3 and 3 to 1. If the instantaneous phase delay at telescope j is ϕ_j, measured visibility on baseline ij is $V_{ij}^{\mathrm{obs}} = V_{ij} e^{i(\phi_i - \phi_j)}$. Multiply the three measured visibilities. Every telescope phase then appears once with each sign and cancels. The argument of the product,

$$\Phi_{\mathrm{cl}} = \arg[V_{12} V_{23} V_{31}],$$

is called the closure phase. The three oriented baseline vectors sum to zero. Translating the entire source on the sky introduces a linear phase proportional to baseline, which cancels in the same product. Closure phase thus probes relative brightness structure while losing sensitivity to an absolute position offset. A brightness distribution symmetric about a point has real visibilities after such a translation; its closure phase is zero or π away from zeros of visibility.

Cancellation requires phases measured over the same atmospheric realization or combined in a suitable short-time correlation. A separate instrumental phase error specific to one baseline is not expressible as a difference of telescope phases and will remain in the product. Observing a point-source calibrator can detect such a residual.

Detect asymmetry in an unequal binary

Place a star of unit brightness at the origin and a companion of relative brightness one half at angular separation s along a line of projected baselines. Adopt the Fourier sign convention yielding $V(\phi) = (1 + 0.5 e^{i\phi})/1.5$ with $\phi = 2\pi B s/\lambda$. Choose collectors so that the first two oriented baselines each give phase $\pi/2$. The closing baseline then gives $-\pi$. Their complex visibilities are $(1 + 0.5i)/1.5$, $(1 + 0.5i)/1.5$ and $(1 - 0.5)/1.5 = 1/3$.

Only the first two terms contribute nonzero phase. Their summed argument is $2 \arctan(0.5) = 53.1^\circ$. Reflect the binary about the origin and all visibility phases reverse, giving -53.1° while every two-telescope visibility magnitude is unchanged. At 600 nm and projected separation 3.00 milliarcseconds, the first two projected baseline lengths are

$$B = \lambda/(4s) = 10.3\,\mathrm{m}.$$

For a vanishingly faint companion each visibility tends to one and closure phase tends to zero. Combining several independent triangles with magnitude observations can therefore distinguish asymmetric brightness distributions that reproduce an identical contrast on a single baseline.

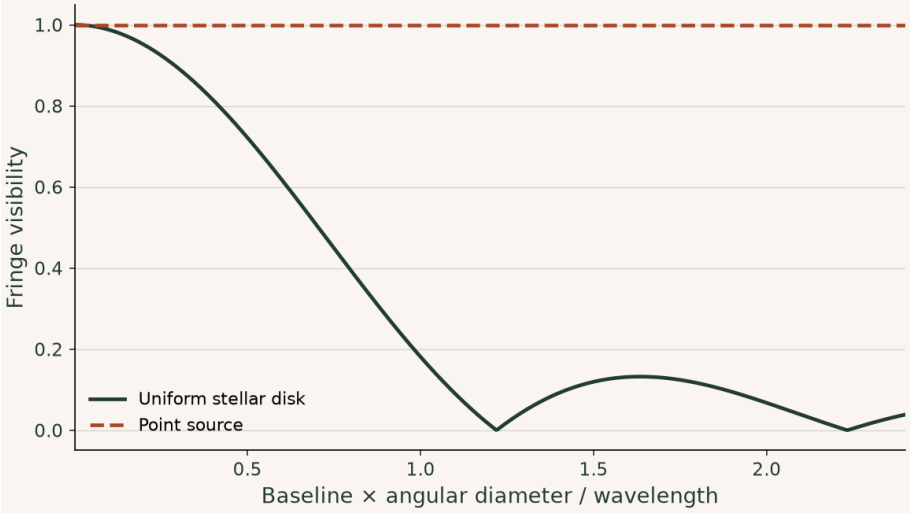
The chosen one-dimensional triangle can be realized with three equally spaced collectors. Its closing baseline then has twice the magnitude of either short baseline and points in the opposite direction under cyclic orientation. In a general two-dimensional array the vectors still form a closed triangle, even though their lengths no longer have this simple ratio. Reversing every baseline convention conjugates the triple product and reverses the reported phase. Astrometric orientation and instrument phase conventions must therefore be calibrated before interpreting a phase sign as the direction of a companion on the sky.

Infer an angular diameter from a visibility null

For a monochromatic, uniformly bright circular disk the normalised mutual coherence is $2J_1(s)/s$,

$$s = \pi B \theta / \lambda.$$

A two-aperture instrument measuring contrast sees its magnitude. Seek a minimum across a range of baselines.



The first null is at $B\theta/\lambda \approx 1.22$. At 600 nm with a 100 m baseline this corresponds to an angular diameter of about 1.51 milliarcseconds. Limb darkening or finite spectral bandwidth can shift the diameter inferred with a uniform-disk fit.

Try it · Stellar disk

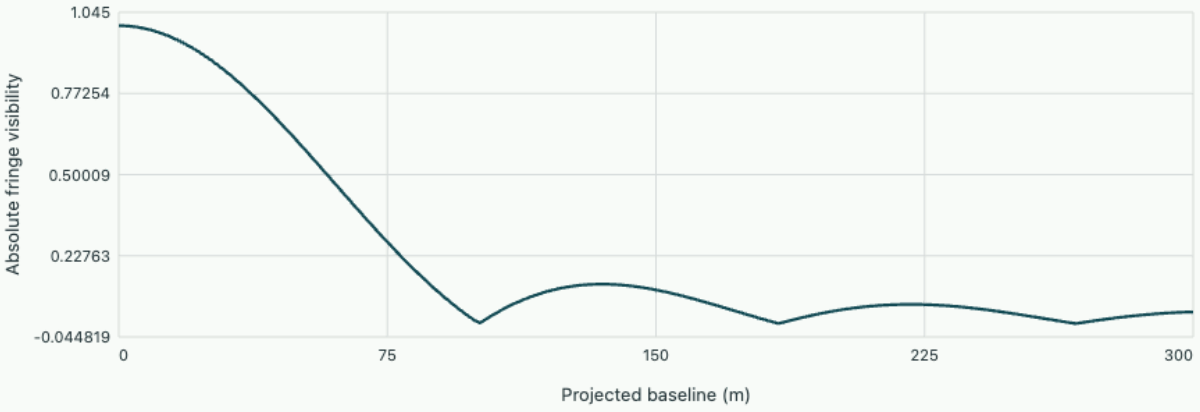
Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Angular diameter (milliarcseconds) **1.5**

Wavelength (nm) **600**



Uniform circular stellar disk

First visibility zero near 100.63 m. Monochromatic uniform disk; limb darkening and bandwidth are omitted.

▶ **Read values from the plot**

Study the full derivation

Photon counting, bunching and the time resolution of a coincidence test · Optical fields, instruments and observations

An intensity trace can be steady while individual detector clicks show correlations. Derive the observable of a two-detector Hanbury Brown–Twiss measurement, quantify how background and bin width change it, and decide whether a dip near zero delay belongs to the source or to a detector.

Problem sheet 6 · Beam transport and spatial coherence

Estimate a focus, diagnose beam quality and determine what a partially resolved binary can tell an interferometer.

Set the depth of focus

An ideal 532 nm Gaussian should maintain at least half its peak axial intensity throughout a 4.00 mm sample centred on the waist. Find the smallest allowed waist radius and the required collimated input radius for a 75.0 mm lens forming that waist.

▼ **Show the Gaussian design**

The half-intensity planes lie at $\pm z_R$, so require $z_R \geq 2.00$ mm. The smallest waist is $\sqrt{\lambda z_R/\pi} = 18.4$ µm. For a flat incoming wavefront, $w_{\text{in}} \simeq \lambda f/(\pi w_0) = 0.690$ mm. A smaller focus would raise central intensity at fixed power while decreasing the end-to-centre ratio below the required half.

Compare a beam with the fundamental mode

At 633 nm a beam has waist radius 40.0 µm and far-field half-angle 7.50 mrad. Find the product relative to the ideal Gaussian value. Under the convention $M^2 = \pi w_0 \theta/\lambda$, estimate the minimum collimated input radius to make a 25.0 µm waist with a 100 mm lens.

▼ **Show the quality and focusing estimates**

The measured product is 3.00×10^{-7} m rad and the ideal value is $\lambda/\pi = 2.01 \times 10^{-7}$ m rad. Their ratio is $M^2 = 1.49$. In the paraxial, approximately Gaussian beam-quality model $\theta = M^2\lambda/(\pi w_0)$, giving $w_{\text{in}} \simeq f\theta = 1.20$ mm for 25.0 µm. This comparison needs consistent radius and divergence definitions, and a non-clipping lens.

A binary with unequal brightness

Two unresolved stars have flux ratio $3:1$ and projected angular separation 4.00 milliarcseconds. Derive the minimum visibility and find its first baseline at 600 nm. What happens if the baseline rotates perpendicular to their separation?

▼ **Show the phasor addition**

Normalize one flux to 3 and the other to 1. Their complex degree of coherence has magnitude $|3 + e^{i\phi}|/4$, with $\phi = 2\pi B s/\lambda$. Its first minimum occurs at $\phi = \pi$ and equals

$$(3 - 1)/4 = 0.50.$$

The separation is $4.00 \times 10^{-3}/206265 = 1.94 \times 10^{-8}$ rad, giving $B = \lambda/(2s) = 15.5$ m. A perpendicular baseline gives zero projected separation, equal phases and unit intrinsic visibility if individual stellar disks remain unresolved.

Can a camera record what its objective resolves?

A 550 nm incoherent image is formed in air by an objective of numerical aperture 0.25 and lateral magnification 20. The detector pitch is 6.5 µm. Compare the Rayleigh separation, the object-side incoherent cutoff and the detector Nyquist limit. Predict the apparent object frequency if a 0.85 cycle/µm sinusoidal target is sampled without prefiltering.

▼ **Worked solution**

The Rayleigh distance for two point emitters is 0.61λ/NA = 1.34 µm. For a sinusoidal intensity pattern the relevant quantity is the incoherent cutoff 2NA/λ = 0.909 cycle/µm. These refer to different targets. Taking the reciprocal of the Rayleigh distance would give the wrong cutoff.

At the object plane each detector pitch corresponds to 6.5/20 = 0.325 µm. Its Nyquist frequency is 1/(2 × 0.325) = 1.54 cycles/µm, above the optical cutoff. The 0.85 cycle/µm target remains at 0.85 cycle/µm. The pupil transfers it weakly because it is close to cutoff. A finite pixel further reduces its contrast by averaging.

Now reduce the magnification to 10. The effective pitch becomes 0.65 µm, with sampling frequency 1.538 cycles/µm. Since 0.85 exceeds half this frequency, the sampled pattern aliases to |0.85 – 1.538| = 0.688 cycle/µm. The optics are unchanged. Choosing M ≥ 2 × 6.5 × 0.909 ≈ 11.8 places even the optical cutoff within the detector passband.

Choose a slab width for one transverse TE mode

At 1.55 µm use $n_1 = 1.50$ and $n_2 = 1.45$. Find the full slab thickness at the first odd-mode cutoff. Classify full thicknesses 1.60 and 2.20 µm for TE polarization. What changes just above cutoff?

▼ **Use the half-thickness convention consistently**

With full thickness $d = 2a$,

$$V = \pi d \sqrt{n_1^2 - n_2^2}/\lambda_0.$$

The odd threshold $V = \pi/2$ gives $d_c = \lambda_0/[2\sqrt{1.50^2 - 1.45^2}] = 2.02$ µm. The 1.60 µm slab supports only the fundamental TE mode, while 2.20 µm also supports the first odd TE mode. Just above threshold its cladding decay constant is small and its field extends far outside the core. The count here excludes TM polarization.

Revisit chapter 11 · When does a beam propagate without diffracting?.

Explore these relations

- For the focus and beam quality, compare Gaussian propagation around the waist and its phase evolution. Open the live calculation.
- For the single-mode guide, vary the slab width and compare the available transverse TE modes. Open the live calculation.

Notebook practice · Measure the phase that a waist adds to a Gaussian beam

Chapter 13 · Reflection, polarisation and the field beyond a reflecting surface

A ray in glass can reflect completely at an air boundary. A second piece of glass brought close to that boundary can nevertheless change the reflection. The field outside the glass follows from the same equations that determine the reflected intensity.

Match the waves where the materials meet

Put the interface at $z = 0$, with transparent nonmagnetic media of refractive indices n_1, n_2. Write the incoming wave as $e^{i(k_x x + k_z z - \omega t)}$. To satisfy Maxwell's boundary conditions at every point on the surface, the incident, reflected and transmitted waves need the same k_x and ω. With $k_0 = \omega/c$ this gives

$$n_1 \sin\theta = n_2 \sin\theta_t,$$

where angles are measured from the normal. The normal component may differ on either side.

For s polarisation the electric field points along the interface perpendicular to the incidence plane. Continuity of tangential electric and magnetic fields gives

$$1 + r_s = t_s,$$
$$n_1 \cos\theta (1 - r_s) = n_2 \cos\theta_t \, t_s.$$

The minus sign on the reflected magnetic term follows from reversing the normal wavevector. Elimination of t_s gives the first expression below. For p polarisation it is the full magnetic field that points perpendicular to the incidence plane. Resolving the electric field into tangential components supplies the interchanged index weights in the second expression. A reversal of the reflected p basis reverses the sign of r_p without changing the measurable reflectance.

$$r_s = \frac{n_1 \cos\theta - n_2 \cos\theta_t}{n_1 \cos\theta + n_2 \cos\theta_t},$$
$$r_p = \frac{n_2 \cos\theta - n_1 \cos\theta_t}{n_2 \cos\theta + n_1 \cos\theta_t},$$
$$R_{s,p} = |r_{s,p}|^2.$$

At normal incidence glass of index 1.50 against air reflects $[(1.50 - 1)/(1.50 + 1)]^2 = 0.040$ for either polarisation. Setting the p numerator to zero and using Snell's law gives $\tan\theta_B = n_2/n_1$, here 33.69°. Only the p reflection vanishes there. At higher angles both reflected intensities approach one.

Continue the transmitted wave through the critical angle

The transmitted normal wavenumber satisfies

$$k_{z,2}^2 = k_0^2 (n_2^2 - n_1^2 \sin^2\theta).$$

Beyond $\theta_c = \arcsin(n_2/n_1) = 41.81^\circ$ it becomes imaginary. Choose $k_{z,2} = +i\kappa$, the root giving a field that decays into the upper half space. The tangential phase still propagates. In the s Fresnel amplitude the

denominator and numerator are now complex conjugates. Its modulus is exactly one, but its phase depends on angle. The p wave also has unit reflectance with its own phase.

$$\frac{1}{\kappa} = \frac{\lambda_0}{2\pi\sqrt{n_1^2 \sin^2\theta - n_2^2}}.$$

For free-space wavelength $500\,\mathrm{nm}$ at 50°, the electric-field amplitude decays by e^{-1} over approximately 141 nm. Its squared magnitude falls by the same factor in just 70 nm. A semi-infinite, lossless second medium carries no time-averaged normal flux in this regime. A nearby second boundary changes the solution by introducing another evanescent component, enabling optical tunnelling.

Locate the two special angles

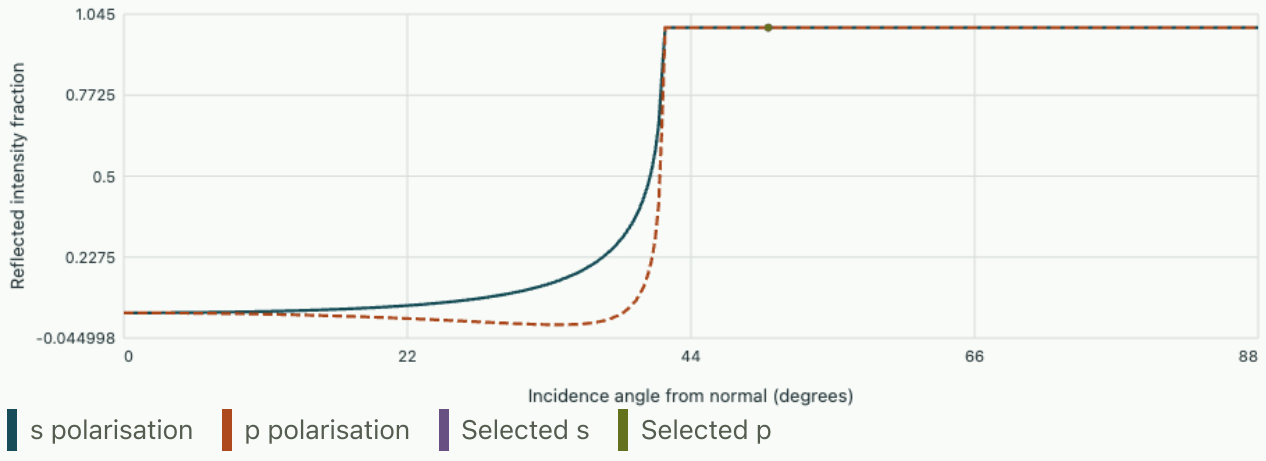
Compare s and p reflectances as the glass index changes. Read the predicted critical and Brewster angles before moving the angle slider. Near the critical angle an ideal plane wave has a very long penetration length. A finite beam has an angular spread, so its measured response mixes nearby incidence angles.

Open this experiment online

Try the calculation

Glass refractive index **1.5**

Angle of incidence (degrees) **50**



Brewster angle 33.69°, critical angle 41.81°. Field penetration at 500 nm is 140.6 nm.

Tutorial on a liquid cladding

Replace air with a liquid of index 1.33. For glass of index 1.50 determine whether the same 50° beam still undergoes total internal reflection. At what angle does the p reflection vanish?

▼ **Work through the solution**

The critical angle increases to $\arcsin(1.33/1.50) = 62.46^\circ$. At 50° a propagating wave now enters the liquid. Its p-polarised reflected amplitude is zero at $\arctan(1.33/1.50) = 41.56^\circ$. The sensitivity of the critical angle to the surrounding medium underlies a refractometer.

At 50° and free-space wavelength 500 nm, a dilute, weakly responding layer samples intensity 70 nm above the glass–air interface. Estimate the local intensity relative to its value immediately outside the glass.

▼ **Work through the solution**

The field penetration length is 140.6 nm. Squaring the decay factor gives

$$\exp[-2(70)/140.6] = 0.369.$$

This ratio concerns local squared field; it does not give the net normal power crossing a lossless interface. A strongly perturbing layer would require recomputing the field.

Estimate the critical-angle shift for a small change of ambient refractive index near air. Use fixed glass index 1.50 and an increase $\delta n_2 = 0.01$.

▼ **Work through the solution**

Differentiating $\arcsin(n_2/n_1)$ gives $d\theta_c/dn_2 = 1/\sqrt{n_1^2 - n_2^2}$ in radians. Near air the predicted shift is 0.00894 rad or 0.512°. This local linear estimate becomes less accurate as the indices approach matching.

Further reading · Delft Interactive Optics, Fresnel coefficients and evanescent waves.

Chapter 14 · How a dielectric slab chooses its guided modes

A plane interface supports a decaying field outside glass under total reflection. Put two interfaces around a thin core and a new question appears. Which transverse patterns reproduce themselves as the wave propagates along the guide?

Separate longitudinal and transverse variation

Take a uniform slab of half-width a, core index n_c, and equal claddings of lower index n_s. A transverse-electric mode can be written $E_y = F(x)e^{i(\beta z-\omega t)}$. Substitution into the scalar Helmholtz equation in each homogeneous region leaves $F'' + (n^2k_0^2 - \beta^2)F = 0$. A guided wave must oscillate in the core and decay outside. Those requirements give the interval $n_s k_0 < \beta < n_c k_0$. Equivalently the effective phase index β/k_0 lies between the material indices.

Introduce $u = a\sqrt{n_c^2k_0^2 - \beta^2}$ and $w = a\sqrt{\beta^2 - n_s^2k_0^2}$. Adding their squares eliminates the unknown β. The possible pairs lie on a circle of radius $V = ak_0\sqrt{n_c^2 - n_s^2}$. Here a always denotes the half-width. Using the full thickness in V would move every numerical cutoff by a factor of two.

Enforce continuity at the core boundary

Choose the even fundamental mode with $F(0) = 1$. The core solution is $\cos(ux/a)$. In the upper cladding write $F = \cos u\, e^{-w(x-a)/a}$, which has the same value at $x = a$ and is bounded at infinity. Nonmagnetic media require both E_y and the tangential magnetic field, proportional to dF/dx, to be continuous. Equating the two derivatives gives the mode equation

$$u\tan u = w = \sqrt{V^2 - u^2}, \qquad 0 < u < \min(V, \pi/2).$$

On this interval the left side increases from zero, and the right side decreases. At the far end either the tangent diverges or the square root reaches zero. There is exactly one root, however thin the symmetric core. For the first odd mode, the core uses $\sin(ux/a)$, and the derivative condition instead gives

$$-u\cot u = w.$$

Its first branch starts at $u = \pi/2$. Consequently $V > \pi/2$ is needed for that mode to become bound.

Resolve the tail with a root solver

At $V = 1$, bisection gives $u = 0.7391$, $w = 0.6736$. Substitution yields the same number on the two sides of the mode equation. The amplitude tail beyond each boundary extends over $a/w = 1.48a$ before dropping by a factor e. At $V = 2$, the fundamental has $u = 1.0299, w = 1.7145$, so the decay length falls to $0.583a$. The first odd mode can now also exist. A guided mode may therefore spread substantially beyond the nominal guide material.

Vary V while the graph keeps the same half-width units. The display solves the transcendental equation afresh, using its monotonic branch to bracket the root. Compare the apparent width at half the field amplitude with the decay length reported in the readout. The former includes part of the oscillatory core, the latter

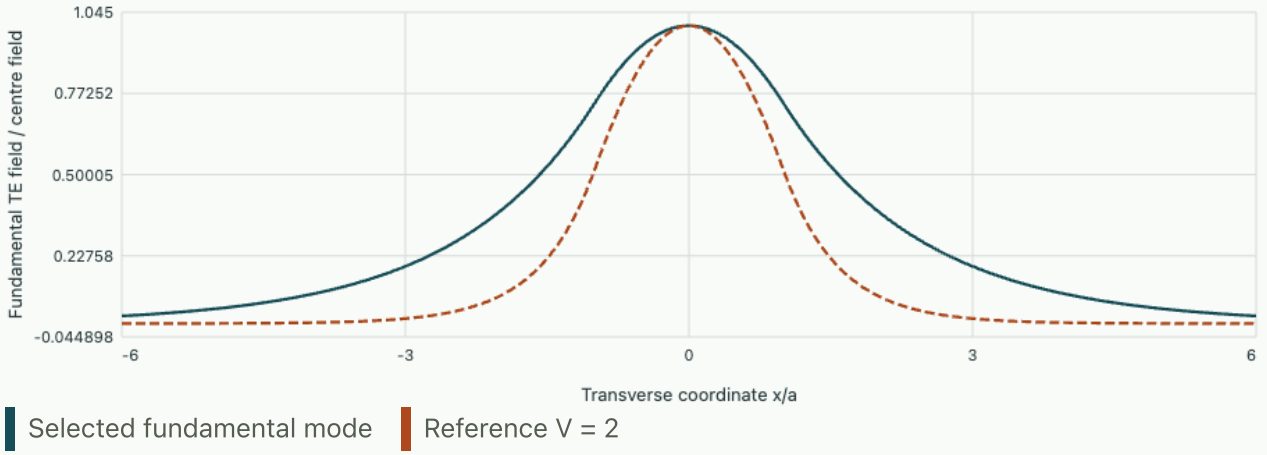
describes only the tail. A detector measuring local intensity responds to $|F|^2$, which has twice the logarithmic decay slope.

Open this experiment online

Try the calculation

Normalised half-width V

1



Root u = 0.7391, cladding exponent w = 0.6736. Field amplitude decays by 1/e over 1.485 half-widths outside the core.

Tutorial on a fabrication tolerance

At a free-space wavelength of $1.55\,\mu\mathrm{m}$, take $n_c = 1.50$ and $n_s = 1.45$. What is the full thickness at the first odd TE cutoff? How would decreasing the wavelength affect the number of modes?

▼ **Work through the solution**

At cutoff set $ak_0\sqrt{n_c^2 - n_s^2} = \pi/2$. The full thickness is $2a = \lambda_0/[2\sqrt{n_c^2 - n_s^2}] \simeq 2.02\,\mu\mathrm{m}$. At fixed geometry and approximately fixed indices, decreasing the wavelength increases V. Additional oscillatory branches can then intersect the available circle and become guided.

Express the squared effective index in terms of w/V. Then estimate it at $V = 1$, $n_c = 1.50, n_s = 1.45$.

▼ **Work through the solution**

From the definitions $w^2/V^2 = (\beta^2/k_0^2 - n_s^2)/(n_c^2 - n_s^2)$. With $w = 0.6736$, the normalised squared-index fraction is 0.454. Substitution gives

$$n_{\mathrm{eff}} = \sqrt{1.45^2 + 0.454(1.50^2 - 1.45^2)} = 1.473.$$

This is a phase index obtained with the mode solution.

How does the fundamental cladding tail scale in a very thin symmetric guide?

▼ **Work through the solution**

At small V, u is small and $u \tan u \simeq u^2$, hence $w \simeq u^2$. The constraint $u^2 + w^2 = V^2$ then gives

$$u \simeq V,$$

$$w \simeq V^2.$$

The field decay length in units of the core half-width grows as V^{-2}, so the mode exists mathematically while occupying mostly cladding.

Further reading · MIT slab dielectric waveguides, TE modes and confinement.

Chapter 15 · Why opening a circular aperture can darken its centre

The on-axis intensity behind a round opening may decrease when the opening grows. Following the phase contributed by each annulus gives a quantitative prediction for this near-field diffraction effect.

Add the complex contributions from the pupil

Illuminate a circular hole of radius a in a thin opaque screen with a uniform monochromatic plane wave. Observe at distance z on the symmetry axis. An annulus at radius r has area $2\pi r\,dr$. Its path to the detector is $\sqrt{z^2+r^2} = z + r^2/(2z) - r^4/(8z^3) + \cdots$. When a/z is small, slowly varying amplitude and inclination factors can be taken at the centre. Retain the quadratic phase but demand that the discarded quartic phase $ka^4/(8z^3)$ be small. The quadratic phase itself can span several cycles.

With the Fresnel propagator $e^{ikz}/(i\lambda z)$, the field relative to an unobstructed plane wave at the same point is therefore an elementary integral. Use $v = r^2$, so $2r\,dr = dv$. The increasing annular area exactly compensates the decreasing radial width of successive equal-phase intervals.

$$\frac{U}{U_{\text{free}}} = \frac{2\pi}{i\lambda z}\int_0^a r\,e^{i\pi r^2/(\lambda z)}\,dr = 1 - e^{i\pi N},$$

$$N = \frac{a^2}{\lambda z}.$$

Taking the squared magnitude gives

$$I/I_{\text{free}} = 2 - 2\cos(\pi N) = 4\sin^2(\pi N/2).$$

The phase convention can reverse the exponent while leaving the intensity unchanged. Odd integer N gives intensity four times the unobstructed intensity on axis. Even integer N gives an ideal zero. A local enhancement is compatible with energy conservation because interference redistributes intensity across the observation plane.

Divide the opening into half-period zones

Boundaries $r_j = \sqrt{j\lambda z}$ separate annuli whose outer and inner points differ by phase π. All complete zones have area $\pi\lambda z$ and alternating complex contributions under this approximation. Opening one complete zone gives a bright axial spot. Opening its neighbour provides the opposite amplitude. If a small detector instead averages over off-axis positions, the phases no longer have perfect azimuthal symmetry and the measured minimum fills in.

For $\lambda = 500\,\text{nm}$ and $z = 0.50\,\text{m}$, the first boundary has radius $0.500\,\text{mm}$. The next has radius $0.707\,\text{mm}$. At fixed opening $a = 0.500\,\text{mm}$ decreasing the screen distance to $0.25\,\text{m}$ also changes N from one to two. Thus an axial scan with a fixed pupil tests the same interference argument as changing the aperture. The small ratio a/z at these distances makes the approximations internally consistent.

Compare amplitude addition with area scaling

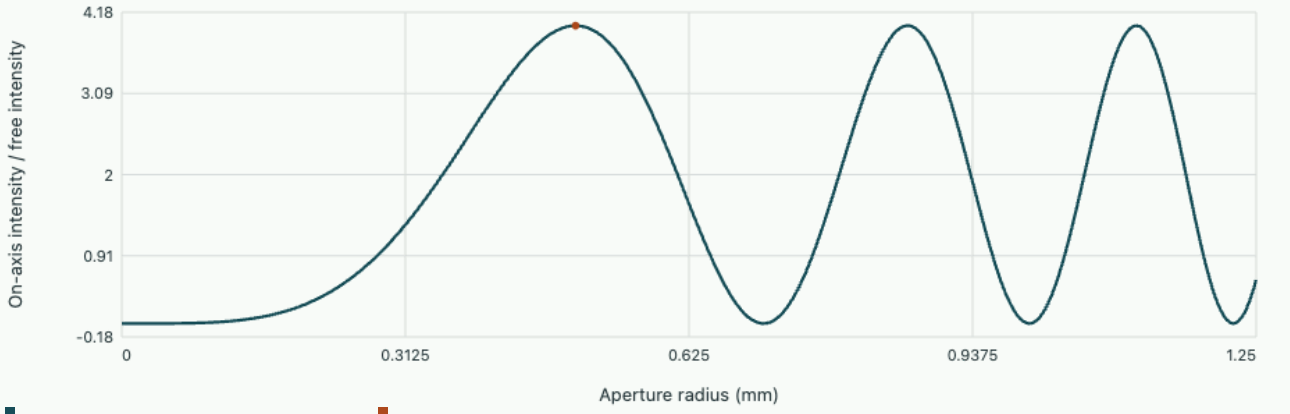
At small N, expand the exponential to find $U/U_{\text{free}} \simeq -i\pi N$, hence intensity grows as N^2 or a^4. This follows because all small-pupil contributions arrive nearly in phase. The small-opening expansion fails long before the paraxial integral does. Vary the radius and distance below. Compare the position of the first minimum with the radius at the first maximum. The display evaluates intensity as a function of physical radius so that the square-root spacing becomes visible. Real propagation includes the pupil transmission profile, spectral bandwidth and detector size.

Open this experiment online

Try the calculation

Aperture radius (mm) **0.5**

Screen distance (m) **0.5**



The selected opening spans 1 half-period zones and gives normalised intensity 4. First complete zone radius 0.5 mm.

Tutorial on a partially blocked zone

Keep only the annulus between the first and second half-period boundaries. What intensity would it produce on axis by itself? What if that annulus and the central disc both transmit, but the annulus transmits only half the electric-field amplitude?

▼ **Work through the solution**

At $N = 1$ the central disc contributes $2U_{\text{free}}$. At $N = 2$ the total contribution vanishes. The annulus therefore supplies $-2U_{\text{free}}$ and by itself has normalised intensity four. With amplitude transmission one half its contribution is $-U_{\text{free}}$. Adding the disc gives U_{free}, or normalised intensity one.

Behind a 0.50 mm opening at 500 nm, find the first three on-axis maxima as the detector approaches from large positive z.

▼ **Work through the solution**

Maxima occur at odd $N = 1, 3, 5$. Since $a^2/\lambda = 0.50\,\mathrm{m}$, the distances are 0.50, 0.1667 and 0.100 m. The spacings shrink toward the aperture because the controlled phase is proportional to $1/z$. Continuing the sequence requires checking the neglected quartic phase.

At the $N = 2$ null, increase the aperture radius by 1 percent with everything else fixed. Estimate the residual relative intensity from the leading nonzero term.

▼ **Work through the solution**

The new Fresnel number is

$$2(1.01)^2 = 2.0402.$$

Near an even integer, $4\sin^2(\pi\delta N/2) \simeq \pi^2(\delta N)^2$, giving 0.0160. The first derivative of intensity vanishes at the ideal null but a small geometric error still creates a measurable finite signal.

Further reading · Delft Interactive Optics, Fresnel propagation and paraxial approximation.

Chapter 16 · When do two curved mirrors support a confined beam?

A ray launched slightly away from the axis may stay near it for many round trips or walk off a mirror. A single round-trip matrix predicts which behaviour occurs and selects the Gaussian beam that can reproduce its wavefront.

Multiply one complete round trip

Use a paraxial ray vector (x, θ), with $\theta = dx/dz$. Free propagation through distance L changes x by $L\theta$, leaving θ fixed. In unfolded coordinates reflection from a concave spherical mirror of radius $R > 0$ changes θ by $-2x/R$. The factor two arises from the change in direction of the local surface normal. Represent propagation and reflection by

$$P = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}, \qquad M = \begin{pmatrix} 1 & 0 \\ -2/R & 1 \end{pmatrix}.$$

For two identical mirrors a round trip is $C = (MP)^2$ starting just after a reflection. Each constituent has determinant one. The half-trip has trace $2 - 2L/R = 2g$, so the identity $\operatorname{tr}(A^2) = (\operatorname{tr} A)^2 - 2\det A$ gives

$$\tfrac{1}{2}\operatorname{tr} C = 2g^2 - 1.$$

Its eigenvalues are reciprocal. If their trace lies strictly between −2 and +2 they are complex conjugates of modulus one and repeated round trips remain bounded. If the absolute trace exceeds two, a component grows exponentially. This produces the usual symmetric resonator range $0 < L < 2R$. The exact confocal point $L = R$ has $C = -I$ and bounded rays despite the repeated eigenvalue. Real alignment and apertures also influence losses.

Match the field curvature to the mirrors

For a symmetric cavity a fundamental Gaussian has its waist at the midpoint. At distance z from that waist its wavefront radius is $R_w(z) = z + z_R^2/z$, where $z_R = \pi w_0^2/\lambda$. At the mirror $z = L/2$, a reproducing wave must have the spherical mirror curvature. Equating $R_w(L/2)$ to R gives

$$z_R^2 = \frac{L}{2}\left(R - \frac{L}{2}\right),$$

$$w_0 = \sqrt{\frac{\lambda z_R}{\pi}},$$

$$w_m = w_0\sqrt{1 + (L/2z_R)^2}.$$

The same interval emerges from requiring a positive squared Rayleigh length. The radius w is the distance at which intensity has fallen to e^{-2} of the on-axis value. For $R = 0.50\,\mathrm{m}$, $L = 0.50\,\mathrm{m}$ and $\lambda = 600\,\mathrm{nm}$, $z_R = 0.25\,\mathrm{m}$, $w_0 = 0.219\,\mathrm{mm}$ and $w_m = 0.309\,\mathrm{mm}$. These scale as the square root of the wavelength at fixed geometry.

Move towards the concentric limit

Keep the same mirror radius and increase the separation to $0.90\,\mathrm{m}$. The Rayleigh length becomes $0.150\,\mathrm{m}$, the waist contracts to $0.169\,\mathrm{mm}$, and the mirror spot grows to $0.535\,\mathrm{mm}$. A smaller focus can therefore require a larger mirror aperture. The ideal spherical paraxial solution ceases to describe an actual beam if diffraction clipping or large ray angles become important.

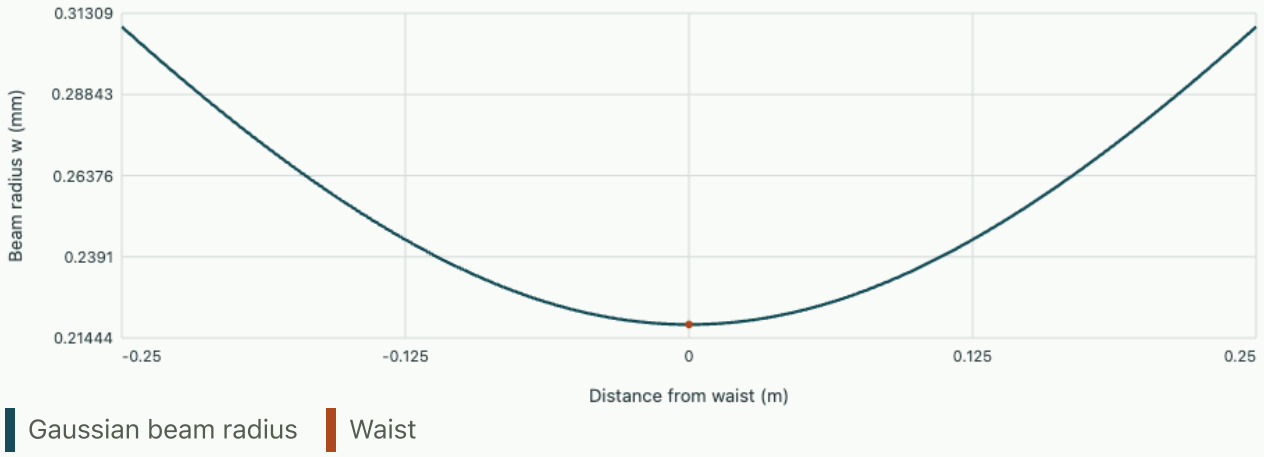
Change L/R and inspect the beam radius across one pass. Compare settings equidistant from the confocal ratio, such as 0.4 and 1.6. Their waist radii agree because $L(2R - L)$ is symmetric. Their mirror spots differ because the distances from the waist differ.

Open this experiment online

Try the calculation

Mirror separation / curvature radius **1**

Wavelength (nm) **600**



Rayleigh length 0.25 m, waist radius 0.2185 mm, mirror spot radius 0.309 mm.

Tutorial on an aperture measurement

A symmetric air cavity has $L = 0.40\,\mathrm{m}$ and measured $w_m/w_0 = \sqrt{2}$. Find R and, at $\lambda = 632.8\,\mathrm{nm}$, the waist radius.

▼ **Work through the solution**

The width ratio gives $L/(2z_R) = 1$, hence $z_R = 0.20\,\mathrm{m}$. The curvature there is

$$R = 0.20 + 0.20^2/0.20 = 0.40\,\mathrm{m}.$$

The cavity is confocal, and $w_0 = \sqrt{632.8 \times 10^{-9} \times 0.20/\pi} = 0.201\,\mathrm{mm}$.

At fixed curvature radius R and wavelength, which symmetric separation maximises the waist radius? Find that maximum and check it against the confocal example.

▼ **Work through the solution**

Maximising w_0 is equivalent to maximising $z_R^2 = L(2R - L)/4$. Differentiation sets $L = R$, with $z_R = R/2$. Thus $w_{0,\max} = \sqrt{\lambda R/(2\pi)} = 0.219\,\mathrm{mm}$ for the parameters in the chapter. Near either end of the ideal stability window the waist shrinks instead.

A ray has $x = 0.1\,\mathrm{mm}$, $\theta = 1\,\mathrm{mrad}$ just after a reflection in an exactly confocal cavity. Where is the ray in the same section after one and after two round trips?

▼ **Work through the solution**

At $L = R$ the round-trip matrix is $-I$, so after one return the coordinates are $-0.1\,\mathrm{mm}, -1\,\mathrm{mrad}$. After two trips they return to their initial values. This exact symmetry of identical mirrors explains why a degenerate trace does not automatically imply secular walk-off.

Further reading · MIT Fundamentals of Photonics, Gaussian beams and resonators.

Chapter 17 · Threshold, gain saturation and the decay of stored light

After establishing a transverse cavity mode, consider its intensity over many round trips. The same mirror losses determine a threshold with pumping and a measurable decay time when gain is removed.

Record every factor in one round trip

Let L be the physical distance between mirrors. Their intensity reflectances are R_1, R_2. Suppose the entire length contains a uniform medium with intensity gain coefficient g and distributed intensity loss coefficient α, both per metre. Traversing length dz multiplies intensity by $1 + (g - \alpha)dz$, which integrates to $e^{(g-\alpha)L}$. During a round trip the light crosses the medium twice and reflects once at each mirror. The multiplier is $\mathcal{M} = R_1 R_2 e^{2(g-\alpha)L}$. Small fluctuations in light intensity grow if $\mathcal{M} > 1$. Setting it to one and taking logarithms gives

$$g_{\mathrm{th}} = \alpha + \frac{1}{2L} \ln \frac{1}{R_1 R_2}.$$

This is a balance for a resonant mode. Phase closure also selects the allowed longitudinal frequencies and only modes overlapping the gain band can meet threshold. If a shorter gain section occupies length $\ell < L$, its exponent is $2g\ell$ while background loss can still act over $2L$. Using an amplitude reflectivity in a formula derived for intensity would introduce a second independent factor-of-two error.

Find the stationary intensity

As stimulated emission increases, the available population inversion typically falls. In a simple homogeneously saturated, single-mode, spatially averaged model use $g(I) = g_0/(1 + I/I_s)$ with small-signal gain g_0 and saturation intensity I_s. A stationary positive intensity requires $g(I) = g_{\mathrm{th}}$, hence $I/I_s = g_0/g_{\mathrm{th}} - 1$. Below threshold the positive solution disappears and the deterministic zero-intensity state is stable. Above threshold a small increase in intensity lowers gain and pushes the system back. Spontaneous emission seeds a real laser and smooths the ideal transition. Relaxation oscillations would require an explicit time equation for the inversion.

For $L = 0.30\,\mathrm{m}$, $\alpha = 0.10\,\mathrm{m}^{-1}$, $R_1 = 0.98$ and $R_2 = 0.90$,

$$g_{\mathrm{th}} = 0.3093\,\mathrm{m}^{-1}.$$

A pump giving $g_0 = 0.50\,\mathrm{m}^{-1}$ therefore sustains $I \simeq 0.617 I_s$. If the output mirror transmitted more light, it could extract a greater fraction of the circulating light per encounter. It would also raise threshold. At a fixed pump these effects compete.

Read the passive lifetime from the same losses

Switch the gain term to zero. After j round trips the intensity is $I_j = I_0 [R_1 R_2 e^{-2\alpha L}]^j$. The time per round trip in a uniform medium of group index n_g is $t_{\mathrm{rt}} = 2 n_g L/c$. Match $I_j/I_0 = e^{-j t_{\mathrm{rt}}/\tau}$ to obtain

$$\tau = -t_{\mathrm{rt}} / \ln(R_1 R_2 e^{-2\alpha L}).$$

In air the preceding example gives $\tau = 10.8\,\mathrm{ns}$. This is the intensity lifetime. The electric-field amplitude would decay twice as slowly.

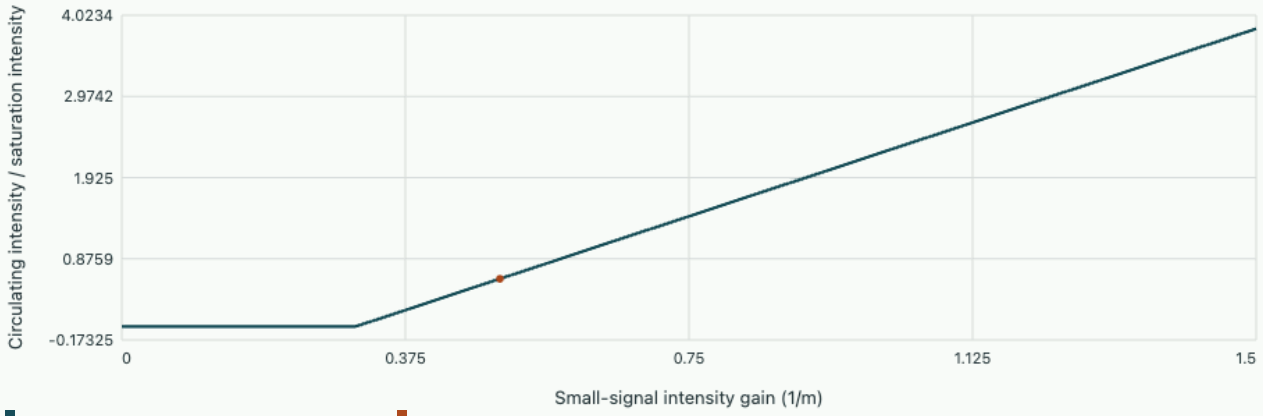
Adjust the output reflectance, then predict which way the threshold and passive lifetime will move. The graph gives the stationary circulating intensity against small-signal gain. Its change of slope is observable even without knowing I_s exactly. The readout evaluates the selected pump point and the passive lifetime independently.

Open this experiment online

Try the calculation

Output mirror reflectance **0.9**

Small-signal gain (1/m) **0.5**



Threshold 0.3093 m⁻¹, selected circulating intensity 0.6167 I_s, passive intensity lifetime 10.79 ns.

Tutorial on inferring a mirror reflectance

A passive air cavity of length $0.30\,\mathrm{m}$ with negligible internal absorption has an intensity lifetime $20.0\,\mathrm{ns}$. Mirror one reflects 99.0 percent. Infer mirror two's intensity reflectance and the uniform gain coefficient required for threshold.

▼ Work through the solution

The round-trip time is $2.00\,\mathrm{ns}$. The round-trip intensity retention is $e^{-2.00/20.0} = 0.9048$, so

$$R_2 = 0.9048/0.990 = 0.914.$$

The threshold is

$$-\ln(0.9048)/(0.60\,\mathrm{m}) = 0.167\,\mathrm{m}^{-1}.$$

Scattering or absorption unaccounted for in the inference would be attributed to mirror two.

Compare the energy lost per round trip with the exponential time constant in the example. Why would treating the logarithm as its first-order expansion bias the lifetime?

▼ **Work through the solution**

The retention is

$$0.98 \times 0.90 e^{-0.06} = 0.8306.$$

The fractional loss is 0.1694, but the exponential log decrement is $-\ln 0.8306 = 0.1856$. Using the fractional loss instead would give $2.00\,\mathrm{ns}/0.1694 = 11.8\,\mathrm{ns}$, longer than the exact 10.8 ns. The difference vanishes only in the small-loss limit.

If an ideal gain crystal fills half the original cavity and all other losses remain, determine the gain coefficient required inside the crystal.

▼ **Work through the solution**

The round-trip gain becomes $2g\ell$ with $\ell = L/2$. The target round-trip log gain is unchanged. Therefore the necessary crystal gain doubles from 0.3093 to $0.6186\,\mathrm{m}^{-1}$. This comparison assumes the gain crystal does not alter reflectance, group velocity or losses elsewhere.

Further reading · Delft Interactive Optics, gain, resonators and laser threshold.

Chapter 18 · Turning invisible phase into visible intensity in a Fourier plane

A transparent specimen can vary the phase of a beam while transmitting nearly constant intensity. A pair of lenses gives independent access to the undiffracted beam and the waves carrying those small phase variations. Recombining them with a controlled relative phase makes the specimen visible.

Locate the orders in a four-focal-length system

Use coherent, normally incident monochromatic illumination and two ideal thin lenses of equal focal length f. Put the specimen one focal length before lens one; its back focal plane coincides with the front focal plane of lens two, and the final image lies one further focal length beyond lens two. A spatial harmonic $e^{i2\pi\nu x}$ leaves the object with transverse wavevector $2\pi\nu$. In the paraxial approximation its propagation angle is $\lambda\nu$. The first lens therefore maps it to $X = f\lambda\nu$ in the common focal plane. The second lens recombines transmitted orders into an inverted image, apart from a common phase and scale.

For grating period $d = 20\,\mu\mathrm{m}$, $f = 100\,\mathrm{mm}$ and $\lambda = 500\,\mathrm{nm}$, the first diffraction orders lie $2.50\,\mathrm{mm}$ to either side of the centre. A mask small enough to cover just the undiffracted spot changes the reference wave independently of those orders if the illumination extent makes each spot sufficiently narrow.

Keep the field complex until the detector

Let a weak pure phase grating have transmission $t(x) = e^{i\epsilon\cos(2\pi x/d)}$, $|\epsilon| \ll 1$. To first order it consists of a constant reference and two imaginary sidebands of amplitude $i\epsilon/2$. The original intensity $|t|^2$ is exactly one. Apply a phase shift φ to the central order alone. To first order the output field is $e^{i\varphi} + i\epsilon\cos(2\pi x/d)$. The inverted coordinate gives the same cosine. Its intensity becomes

$$I(x)/I_0 = 1 + 2\epsilon\sin\varphi\cos(2\pi x/d) + O(\epsilon^2).$$

The cross term comes from multiplying the sidebands by the conjugate of the modified reference. At $\varphi = +\pi/2$ the phase grating gains positive intensity contrast; the opposite shift reverses it. With $\epsilon = 0.10$, the leading extrema are 1.20 and 0.80. Blocking the reference entirely instead leaves a dark-field intensity $\epsilon^2\cos^2(2\pi x/d)$, insensitive at leading order to the sign of the phase.

Retain the exact phase grating for a numerical check

Within an ideal infinite-aperture relay, only the central Fourier coefficient is changed. That coefficient is $c_0 = (2\pi)^{-1}\int_0^{2\pi} e^{i\epsilon\cos u}du$, real because imaginary contributions cancel by symmetry. Thus the exact transmitted image field is $e^{i\epsilon\cos u} + (e^{i\varphi} - 1)c_0$. It reduces exactly to the original grating at zero phase shift. Its period-averaged intensity remains one because a lossless mask changes phases only. When ϵ is no longer small, the image contains distorted harmonics that the linear contrast formula misses.

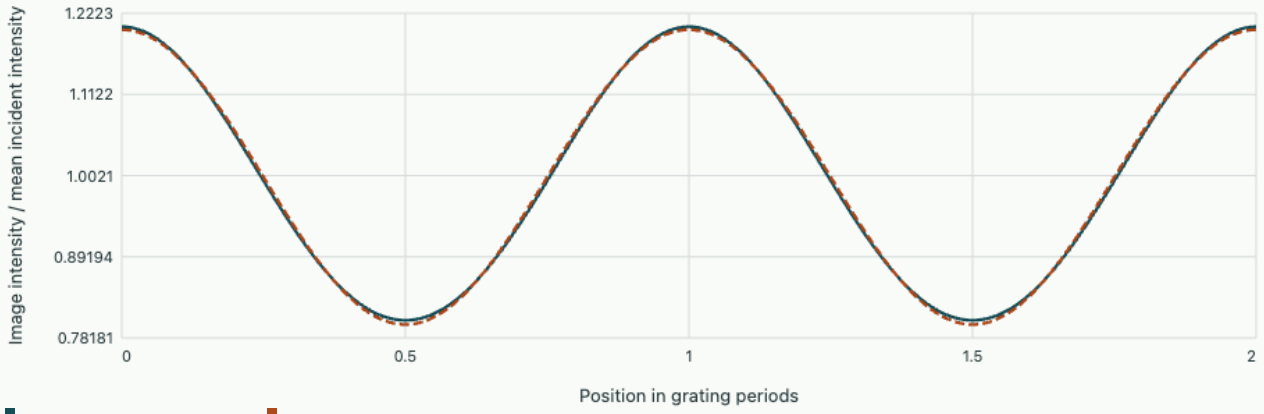
Change the phase amplitude and central phase in the experiment. It uses periodic quadrature for c_0, computes the exact intensity, and compares the linear prediction. Looking at both minima and maxima exposes weak-phase breakdown even when the overall image looks periodic. Aberrations, partial coherence and finite objective aperture can modify real measurements by affecting additional Fourier components.

Open this experiment online

Try the calculation

Specimen phase amplitude ε (rad) **0.1**

Central order shift ϕ (degrees) **90**



Zero-order amplitude 0.9975. Peak-to-trough intensity difference 0.3983; linear prediction 0.4.

Tutorial on the geometry of a phase plate

Change to a $10\ \mu\mathrm{m}$ grating at unchanged optics and then halve the focal length. Where do the first orders lie in the final configuration relative to their original spots?

▼ **Work through the solution**

Halving d doubles ν, while halving f cancels that change. Both spots again sit 2.50 mm from centre. Although the spot positions agree, the two physical gratings have different spatial periods in their respective image planes.

A specimen gives normalised peak-to-trough intensity difference 0.32 with an ideal quarter-cycle central shift. Infer its phase amplitude in the weak model. How could you check for an amplitude grating contaminating the signal?

▼ **Work through the solution**

The leading difference is $4|\epsilon|$, giving $|\epsilon| = 0.080$ rad. Reverse the central phase shift and repeat. The pure-phase contribution changes sign. Also image with no central phase shift, for which a true pure-phase transmission produces constant intensity in an ideal relay.

Further reading · MIT Optics, four-focal-length spatial filters.

Study the full derivation

What an objective must collect to resolve a specimen · Optical fields, instruments and observations

A microscope can transmit plenty of light and still erase fine detail. Treat the specimen as a weak grating, follow its spatial harmonics through the objective, and determine why changing illumination coherence changes the transfer of contrast.

Chapter 19 · Building a dielectric mirror from weakly reflecting interfaces

A single transparent dielectric interface often reflects only a few percent of the optical power. A periodic stack can reflect nearly all incident light over a finite wavelength band. Its response follows from matching both fields at every interface, including every internally reflected wave.

Propagate two continuous tangential fields

At normal incidence use the field vector (E, HZ_0), with free-space impedance Z_0. A forward travelling wave in a nonmagnetic lossless layer of index n has $HZ_0 = nE$; a backward wave has the opposite sign. With time factor $e^{+i\omega t}$, forward propagation contributes e^{-ikz}. Resolve the fields at the exit face into these two waves, propagate back by distance d and recombine. The input fields are the output fields multiplied by

$$M(n,\delta) = \begin{pmatrix} \cos\delta & i\sin\delta/n \\ in\sin\delta & \cos\delta \end{pmatrix}, \qquad \delta = 2\pi n d/\lambda.$$

Its determinant is one. Use a product $M_H M_L$ for a high-index layer encountered first, followed by a low-index layer. The tangential fields are continuous at every interface, so no separate interface matrices are required in this field basis. Changing the order can change the reflection phase and the way the stack matches different external media.

Compute a quarter-wave pair without numerical matrices

Choose layer thicknesses with $n_H d_H = n_L d_L = \lambda_0/4$. At the design wavelength both phases are $\pi/2$. Multiplying the two off-diagonal matrices gives the diagonal pair matrix $\mathrm{diag}(-n_L/n_H, -n_H/n_L)$. A forward transmitted wave in substrate n_s has output admittance $Y_s = HZ_0/E = n_s$. After N identical pairs the entrance admittance becomes

$$Y_{\mathrm{in}} = n_s (n_H/n_L)^{2N}.$$

Matching to the incident plus reflected waves in n_0, $Y_{\mathrm{in}} = n_0(1-r)/(1+r)$, therefore

$$r = \frac{n_0 - Y_{\mathrm{in}}}{n_0 + Y_{\mathrm{in}}},$$

$$R = |r|^2.$$

With incident air, $n_H = 2.0$, $n_L = n_s = 1.50$, and four pairs, $Y_{\mathrm{in}} = 14.98$ and $R = 0.765$. Eight pairs give $Y_{\mathrm{in}} \simeq 149.7$ and $R \simeq 0.974$. Each extra pair multiplies the admittance at the design point, while the reflectance approaches a ceiling of one. The enhancement depends on fixed phase relations across the layers. Adding interface intensities would double-count some energy and lose the interference.

Use the same equations away from design

At a general wavelength the matrices contain both diagonal and off-diagonal terms. For total entries M_{ij},

$$Y_{\mathrm{in}} = (M_{21} + M_{22} n_s)/(M_{11} + M_{12} n_s).$$

One may instead propagate the substrate admittance backwards through each layer by this fractional linear transformation. This is the recursion used below. It avoids introducing any fictitious absorption. The number of periods mainly controls how close the stop-band reflection comes to unity, whereas the index contrast strongly influences the bandwidth. Material dispersion, roughness and optical absorption are excluded from the model.

Set the design wavelength to the fixed 600 nm reference in the experiment. Change the high index or number of pairs. Compare the on-design readout to the diagonal formula, then inspect the finite-stack ripples away from the central band. The display holds quarter-wave optical thicknesses at the design wavelength as indices change, so it represents redesigned layers.

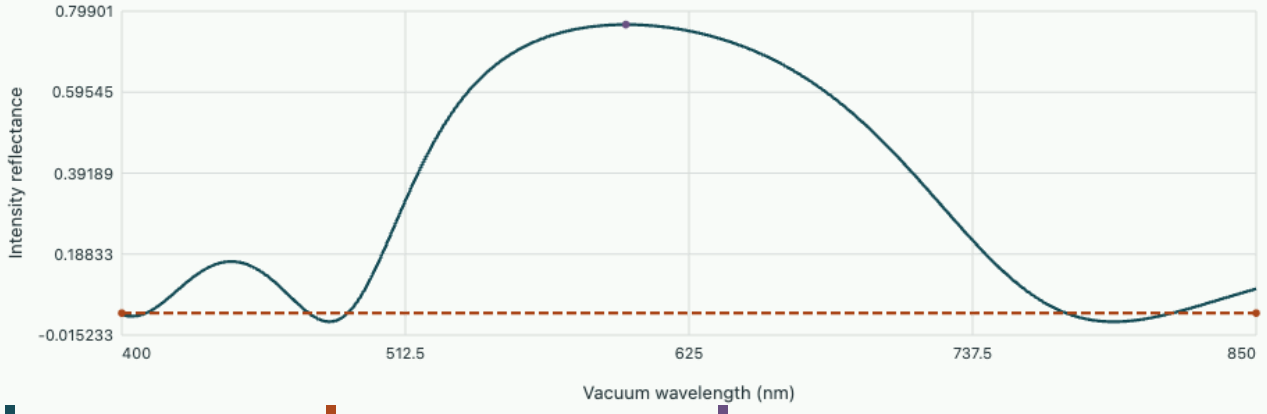
Open this experiment online

Try the calculation

Quarter-wave pairs **4**

High layer index **2**

Probe wavelength (nm) **600**



At 600 nm reflectance is 0.7654. At 600 nm the pair-matrix formula gives 0.7654.

Tutorial on a coating specification

For the worked indices find the minimum pair count with on-design intensity reflectance above 0.99.

▼ **Work through the solution**

Require

$$Y_{\mathrm{in}}/n_0 > (1+\sqrt{0.99})/(1-\sqrt{0.99}) = 398.0.$$

Thus

$$N > \ln(398.0/1.5)/[2\ln(2/1.5)] = 9.70.$$

Ten complete pairs suffice in this ideal lossless calculation.

What happens when the two layer indices become equal? Why can a central wavelength still be defined geometrically yet fail to produce a stop band?

▼ Work through the solution

Every internal interface then has identical fields on both sides with zero reflection. The pair transfer at the design wavelength is $-I$, which leaves the substrate admittance unchanged for any number of pairs. With a substrate of the same index only the external air-to-material mismatch remains. A chosen phase delay alone does not create contrast without a spatial modulation of index.

Further reading · MIT Fundamentals of Photonics, thin-film transfer matrices and Bragg mirrors.

Chapter 20 · Why a grating makes images of itself in empty space

Illuminate a periodic mask with a monochromatic plane wave. A screen placed at certain distances records the original pattern even though there is no imaging lens. Derive the recurrence from the longitudinal phase of each diffraction order, predict the half-period displacement, and use exact propagation to test the paraxial argument.

A near-field measurement that retains the relative phases

Imagine a plane transmission mask made by repeating one profile $t(x)$ every distance d. Coherent, normally incident illumination makes the complex transverse field immediately behind the mask proportional to this transmission. In a far-field experiment one might separate diffraction orders and measure their individual intensities. To reconstruct the mask in free space, however, the orders must illuminate the same region of the screen and preserve their relative phases. Assume an effectively infinite mask and a scalar field with wavelength λ in the propagation medium. A wide finite illuminated region can approximate this idealisation near its centre for a limited distance.

Because $t(x+d)=t(x)$, its Fourier series requires only transverse wave numbers $k_{x,m}=2\pi m/d$ with integer m. Write the input as $\sum c_m e^{2\pi i m x/d}$. The homogeneous Helmholtz equation for a single plane-wave component is $k_x^2+k_z^2=k^2$ with $k=2\pi/\lambda$. Choose $k_z=\sqrt{k^2-k_x^2}$ with nonnegative propagation or decay, so a forward field propagates as $e^{ik_z z}$ and an evanescent component decays. This branch choice is physical; the opposite choice could turn unresolved fine structure into unbounded growth with distance.

Subtract the common carrier e^{ikz}. It has unit modulus in a lossless medium and cannot affect intensity. When $|k_x|\ll k$, expand the square root:

$$k_z-k=-k_x^2/(2k)-k_x^4/(8k^3)+\cdots.$$

Keeping the quadratic part gives a phase proportional to the square of an integer. This arithmetic is the source of the rephasing; it does not require a caustic or refraction at an interface.

Derive the first complete field recurrence

With $\xi=x/d$, the paraxial field envelope a distance z behind the mask is

$$U(\xi,z)=\sum_{m=-\infty}^{\infty} c_m e^{2\pi i m\xi}\exp\!\left(-i\pi\lambda z m^2/d^2\right).$$

Demand that the propagation factor be one for every integer m, including $m=1$. The first positive distance that meets this condition for a generic mask is $z_T=2d^2/\lambda$. Each factor becomes

$$e^{-2\pi i m^2}=1.$$

The complex field is thus restored up to the common carrier, a stronger result than merely recovering an intensity pattern. Some masks have extra symmetries or contain only selected orders and can give earlier intensity revivals. Define the length used in a report explicitly, particularly when shifted self-images are included.

For a 12 micrometre period and wavelength 0.60 micrometre, the full paraxial recurrence is 480 micrometres, short enough that the separation of individual far-field orders would be a misleading picture. Doubling the period at fixed wavelength produces a fourfold distance. Doubling wavelength while holding the period fixed halves the predicted length and at the same time increases the angles of the orders. The latter change makes the paraxial expansion less accurate, so variation of λ tests two effects at once.

Half a recurrence shifts the image; a quarter combines two images

At $z_T/2$ the factor is

$$e^{-i\pi m^2} = (-1)^m.$$

A spatial shift by half a period contributes exactly the same phase to Fourier order m, since $e^{\pi i m} = (-1)^m$. Thus the field becomes $t(\xi + 1/2)$. A photograph at this plane can look identical if one ignores the grating origin; registration against a marked edge discloses the shift. The derivation depends on integers modulo two, not on the detailed transmission of a transparent or absorbing structure.

At $z_T/4$ an even order has factor one and an odd order factor $-i$. Split the field into the components even and odd under a half-period translation. Their respective projectors acting on a pattern are $[t(\xi) + t(\xi + 1/2)]/2$ and $[t(\xi) - t(\xi + 1/2)]/2$. Recombining with the two phases gives

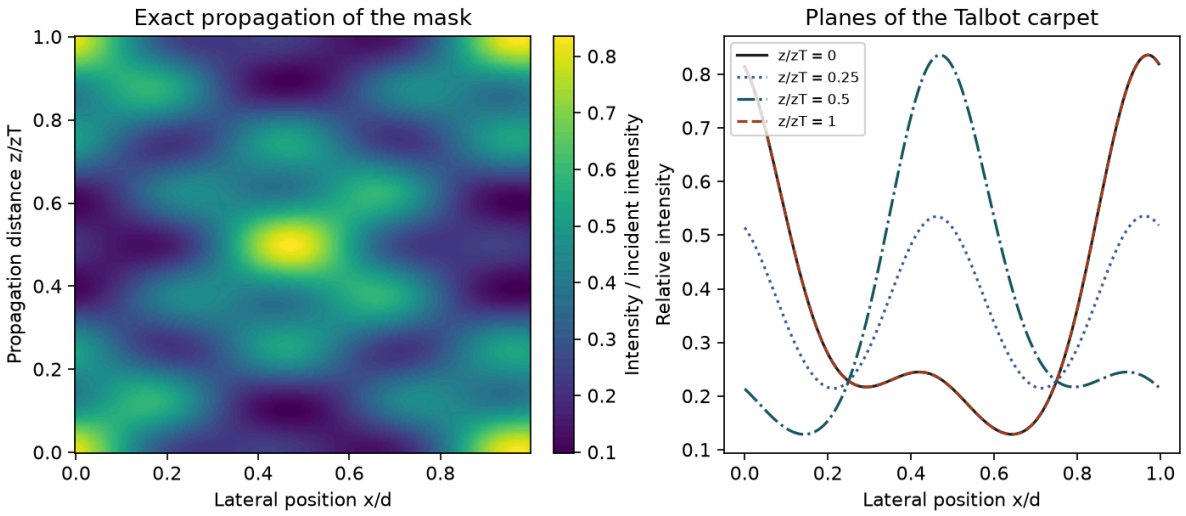
$$U(\xi, z_T/4) = \frac{1 - i}{2} t(\xi) + \frac{1 + i}{2} t(\xi + 1/2).$$

These are complex amplitudes, so they must be summed before taking an absolute square. For a real mask, the coefficients differ in phase by a right angle and the two cross terms cancel in intensity. The quarter-plane intensity is then the average of the intensities in the original and shifted masks. A complex phase mask need not have that cancellation. Replace a real transmission by one with the same modulus and a sinusoidal phase and the input photograph can remain unchanged while the propagated pattern changes markedly.

Open editable notebook Download .ipynb

Compare an exact angular spectrum with the paraxial reconstruction

The notebook uses a deliberately smooth, real mask with a nonzero constant term and two pairs of Fourier orders. Its peak transmission is below one. Starting with only finitely many orders separates propagation error from the truncation error that a numerical square-wave mask would introduce. The code computes the forward Fourier amplitudes, multiplies each by the exact propagator and transforms back. The numerical grid contains many more points than required to represent these five orders. Finer pixel sampling cannot correct a wrong propagation phase.



The carpet uses exact angular-spectrum propagation of a smooth periodic mask. The nearly overlapping original and full-distance curves record revival. A bright fringe crossing the panel is an interference feature, not the path of a particular ray.

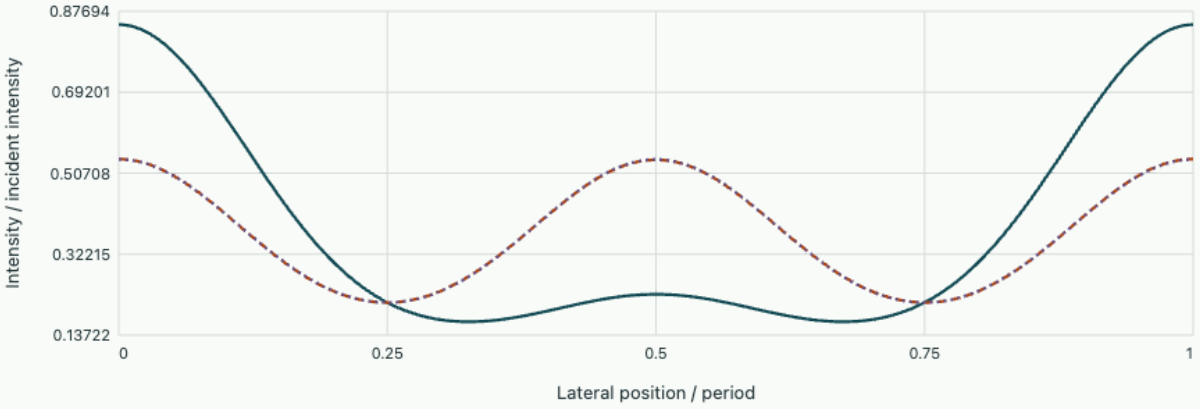
Try it · Talbot

Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Propagation distance / paraxial Talbot length	0.25
Grating period (µm)	10
Second-harmonic mask phase (degrees)	0



Wavelength 633 nm; paraxial Talbot length 315.96 µm. This smooth mask contains only orders 0, ±1, ±2, so their exact propagation is evaluated directly.

Download the corresponding editable notebook. The browser recalculates the reduced equations directly; the notebook exposes the full calculation and its checks.

For an order m the leading phase error of the paraxial approximation at one Talbot distance is $-(\pi/2)m^4(\lambda/d)^2$. This follows by multiplying the neglected quartic term in k_z by $2d^2/\lambda$. A small diffraction angle ensures that the wavenumber expansion is locally accurate, but the propagation distance can still accumulate an appreciable phase, especially for high orders. At $\lambda/d = 0.05$ and $m = 2$, the leading magnitude is 0.0628 rad. The visual distortion remains small for the present mask because this order also has only a modest amplitude. Frequency content and propagation distance both enter an accuracy statement.

Increase the wavelength-to-period ratio and compare the near-coincidence at full distance with the exact curves. Run the included assertions on the complex paraxial field, not only its intensity. Equality at half distance is checked against a discrete half-period shift. At the quarter distance it is checked against the analytic superposition. An optional OneSimulation cell applies its angular-spectrum propagation operator to the identical field and compares the complex output with the explicitly constructed exact spectrum.

Agreement here validates the branch, units and normalisation. It does not make the paraxial approximation exact.

Which aspects of the ideal experiment survive a laboratory source?

A finite wavelength distribution produces a distribution of propagation phases. At a fixed screen location the camera records the intensity averaged over mutually incoherent wavelengths. Orders that agree in phase at the central wavelength accumulate a spread proportional to $m^2 z$. Fine structure requiring higher orders therefore loses contrast first. Averaging wavelengths as fields instead of intensities would represent a phase-locked coherent source and a detector model with different temporal resolution. The distinction matters most when interpreting a blurred revival as evidence of a faulty grating.

A finite angular spread of incident plane waves has a different signature. Under a small tilt θ, the transverse wave number becomes $k\theta + 2\pi m/d$. Squaring produces a common phase, the usual term in m^2, and a term linear in m. A linear Fourier phase shifts the whole pattern laterally by approximately $z\theta$; averaging over unknown incident angles thus convolves it with a distribution of translations. A finite lateral aperture additionally limits the distance over which all orders overlap. These mechanisms suggest separate controls: narrowing spectral width, improving collimation, and enlarging illuminated area alter different parts of the experiment.

For quantitative comparison register the predicted and observed patterns in a common coordinate, subtract camera dark level, and state whether normalisation removes total-power information. An intensity correlation near one may mean that two profiles have the same shape even when one carries much less power. A complex-field discrepancy asks an even stricter question that normally requires phase-sensitive measurements. The numerical field is available in the notebook, but the physical camera does not automatically measure that entire array.

Tutorial on recurrence, fractional images and coherence

A grating of period 20 micrometres makes a full paraxial image 1.60 mm downstream. Determine the wavelength and the location of the shifted image.

▼ **Work through the solution**

Rearranging gives $\lambda = 2d^2/z_T = 0.50$ micrometre. The half-period-shifted pattern appears at 0.80 mm, displaced by 10 micrometres. If the recorded plane were actually the shifted rather than full recurrence, the inferred wavelength would differ by a factor of two. A fixed lateral reference resolves that ambiguity.

Only the constant order and the pair $m = \pm 2$ are present. How soon does the exact paraxial complex envelope recur when distances are expressed using the nominal period d?

▼ **Work through the solution**

The variable propagation factor is $e^{-2\pi i\, 4z/z_T}$. It reaches one first at $z_T/4$. The actual mask constructed from these orders has fundamental period $d/2$, whose ordinary Talbot length $2(d/2)^2/\lambda$ gives exactly the same answer.

A screen is at 1.0 mm and incident directions vary over an rms 1 mrad. Estimate the translation blur for a grating of period 10 micrometres. Which harmonic is most sensitive?

▼ **Work through the solution**

Small-angle translations have rms width about 1.0 micrometre, one tenth of a grating period. For a Gaussian angular distribution, averaging the intensity harmonic with lateral wave number $2\pi n/d$ multiplies its amplitude by $\exp[-(2\pi n z\sigma_\theta/d)^2/2]$. The fundamental retains about 82 per cent and the second harmonic about 45 per cent. Higher spatial frequencies are progressively more sensitive.

Further reading · M. S. Chapman and collaborators · Near-field imaging of atom diffraction gratings

Problem sheet 7 · Choosing an output mirror for a small optical resonator

A laboratory has two interchangeable output mirrors for the same symmetric air cavity. One transmits twice as much light as the other at each reflection. Determine which one can deliver light at the available pump, and use a beam-width measurement and a ring-down trace to test the assumptions.

Predict the geometry before inserting gain

Both mirrors have curvature radius $R = 0.50\,\mathrm{m}$, separation $L = 0.30\,\mathrm{m}$, and operating wavelength $600\,\mathrm{nm}$. Find the Rayleigh length, the waist radius and the spot radius at the mirror. Calculate the fraction of Gaussian power beyond a circular mirror of clear radius $0.60\,\mathrm{mm}$.

▼ **Work through the solution**

Symmetry fixes $z_R = \sqrt{(L/2)(R - L/2)} = 0.2291\,\mathrm{m}$. Thus

$$w_0 = \sqrt{\lambda z_R/\pi} = 0.2092\,\mathrm{mm},$$

$$w_m = w_0\sqrt{1 + (L/2z_R)^2} = 0.2500\,\mathrm{mm}.$$

Integrating the radial Gaussian $r\exp(-2r^2/w_m^2)$ from a to infinity and dividing by the total integral gives

$$e^{-2a^2/w_m^2} = 9.9 \times 10^{-6}.$$

Neglecting clipping against percent-level mirror losses is plausible for this centred mode.

Set the separation-to-radius ratio to 0.60 and verify the mirror spot

Predict the operating point for each mirror

The input mirror has intensity reflectance $R_1 = 0.98$, distributed intensity loss is $\alpha = 0.10\,\mathrm{m}^{-1}$, and uniform small-signal gain is $g_0 = 0.50\,\mathrm{m}^{-1}$. Compare output mirrors with $R_2 = 0.90$ and $R_2 = 0.80$. Use $g(I) = g_0/(1 + I/I_s)$ and estimate transmitted intensity relative to I_s for both.

▼ **Work through the solution**

The thresholds are $\alpha - \ln(R_1R_2)/(2L) = 0.3093$ and $0.5056\,\mathrm{m}^{-1}$. The first mirror supports circulating intensity $I/I_s = 0.50/0.3093 - 1 = 0.617$, with transmitted fraction $0.10 \times 0.617 = 0.0617$ relative to I_s. The second has no positive stationary lasing solution at this pump in the ideal model. Doubling transmission per encounter has raised the required gain above that available.

Run the two reflectance settings at the same pump

Check the losses using a passive decay

For the first mirror combination a passive ring-down fit gives an intensity lifetime of $9.0\,\mathrm{ns}$. Under the uniform model infer the effective attenuation coefficient and compare it to the stated distributed loss. Could the previously estimated clipping of the centred beam explain the discrepancy?

▼ **Work through the solution**

With an air round trip,

$$1/(c\tau) = g_{\mathrm{th,eff}} = 0.3706\,\mathrm{m}^{-1}.$$

Subtract the mirror term $-\ln(0.98 \times 0.90)/0.60 = 0.2093\,\mathrm{m}^{-1}$ to infer $\alpha_{\mathrm{eff}} = 0.1613\,\mathrm{m}^{-1}$. The excess is about $0.0613\,\mathrm{m}^{-1}$, corresponding to $1 - e^{-2L(0.0613)} = 3.61\%$ extra round-trip loss. Two centred aperture losses of order 10^{-5} cannot account for this observation under the assumed mode. Alignment or additional scattering could be examined directly.

Transfer the measurement to a changed geometry

Increase the separation to $0.90\,\mathrm{m}$ with unchanged curvature and wavelength. Estimate the new mirror clipping for $a = 0.60\,\mathrm{mm}$. Explain why using the original measured attenuation as a material coefficient could fail.

▼ **Work through the solution**

Now $z_R = 0.150\,\mathrm{m}$, $w_0 = 0.1693\,\mathrm{mm}$, and $w_m = 0.535\,\mathrm{mm}$. The estimated per-encounter clipped fraction rises to $e^{-2(0.60/0.535)^2} \approx 0.081$. This magnitude is substantial. The field shape may itself be altered by the aperture, so a new ring-down and beam-profile measurement should accompany the estimate. Geometry-dependent clipping can be folded into an inferred attenuation at one geometry without representing distributed absorption.

Problem sheet 8 · Measure a double source with a two-aperture interferometer

Two distant point emitters cannot be separated in a direct image. An observer combines the light collected by two small apertures a variable distance apart. Derive how the visibility of the combined fringes can reveal the angular separation and how a faint companion differs from imperfect temporal coherence.

Add the light from one emitter

A plane wave from a small sky angle θ arrives at apertures separated by projected baseline B. One arm has an adjustable phase delay ϕ. Starting with the complex field amplitude from each aperture, derive the intensity as the delay is scanned. State which phase one can remove by choosing the origin of the time coordinate.

▼ **Work through the solution**

The path difference for the arriving wave is $B\sin\theta \simeq B\theta$. Its contribution to the relative phase is $2\pi B\theta/\lambda$. Equal aperture amplitudes add as $a[1 + e^{i(\phi+2\pi B\theta/\lambda)}]$, up to an overall phase. Taking the squared modulus gives $2|a|^2[1 + \cos(\phi + 2\pi B\theta/\lambda)]$. An overall phase disappears from this expression, while a relative phase remains observable when the delay is scanned. The calculation assumes equal throughput and matching polarisation in the two arms.

Combine independent sources and recover their separation

Let two mutually incoherent emitters of equal brightness have angles $\pm\delta/2$. Show why intensities add between emitters, even though fields added between apertures for each emitter. Find the first visibility minimum and use a first null at $B = 20\,\mathrm{m}$, $\lambda = 600\,\mathrm{nm}$ to determine δ.

▼ **Work through the solution**

Each source is coherent with its own delayed copy for sufficiently narrow bandwidth. The relative phase between the two different emitters wanders during the exposure, removing their mutual cross term. Average the two cosine fringes. Their sum is proportional to $1 + \cos(\phi)\cos(\pi B\delta/\lambda)$. The measured visibility, defined by the extremal intensities, is the absolute value of the second cosine.

$$\mathcal{V} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = |\cos(\pi B\delta/\lambda)|,$$

$$\delta = \frac{\lambda}{2B_{\text{first}}}.$$

These data give $1.5 \times 10^{-8}\,\mathrm{rad}$, or about 3.09 milliarcseconds. A single nonzero visibility may correspond to many angular separations because the cosine repeats. Scanning outward from small baselines distinguishes the first null from subsequent ones. A baseline perpendicular to the separation measures zero projected separation, irrespective of the distance on the sky.

Infer an intensity ratio from a filled minimum

The second emitter has intensity r times that of the first with $0 < r \leq 1$. Derive the new visibility. A well calibrated instrument obtains minimum 0.25. Infer r. Decide whether unequal emitter brightness can move the first minimum in this ideal model.

▼ **Work through the solution**

The normalised complex fringe amplitude, after dropping an overall phase, is $(1 + re^{i2\pi B\delta/\lambda})/(1 + r)$. Its modulus is the visibility. At odd half-fringe offsets the phasors oppose and the minimum is $(1 - r)/(1 + r)$. Inversion gives

$$r = (1 - 0.25)/(1 + 0.25) = 0.60.$$

The phase of opposition is unchanged. Interchanging source labels leaves the measured visibility invariant, so it identifies a ratio up to that choice.

Design an observation that separates the instrument from the source

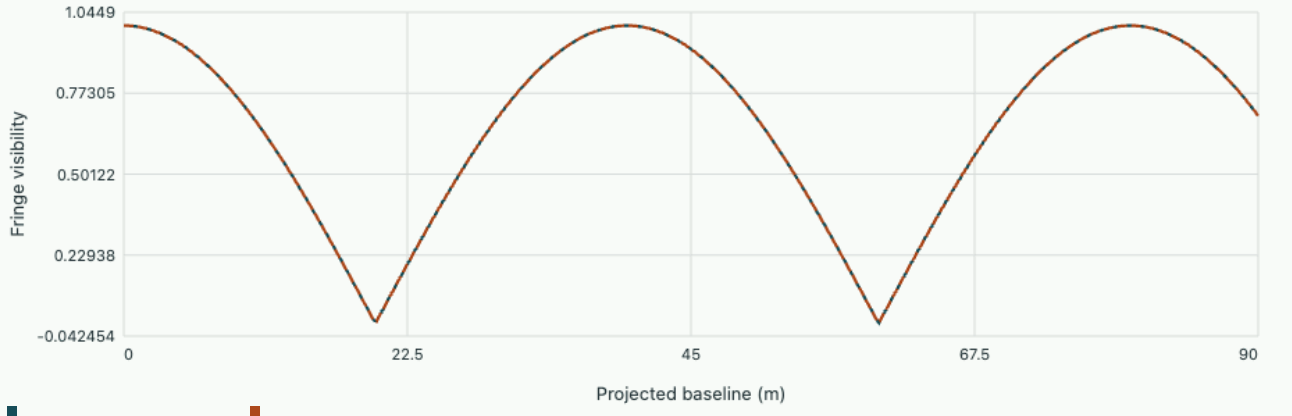
Change angular separation and intensity ratio before reading any numerical values. Compare the location of the first trough with its depth. A calibrator at effectively zero angular size yields visibility 0.80. A binary has raw minimum 0.20 in the same setup. What can be inferred if the entire contrast loss is multiplicative? Which additional observations test this assumption?

Open this experiment online

Try the calculation

Angular separation (milliarcseconds) **3.1**

Intensity ratio **1**



First minimum at 19.961 m. Its intrinsic visibility is 0.

▼ **Work through the solution**

Dividing by the calibrator visibility gives an intrinsic trough of 0.25 and the same inferred intensity ratio 0.60. The model assumes the instrument suppresses contrast by a common factor at that baseline, wavelength and time. Observing the calibrator as a function of baseline checks one part of that assumption. Narrowing the spectral band checks whether the fringe phase changes appreciably across the passband. Turning the baseline tests whether the measured spacing follows the projected binary direction. Each intervention singles out a different step in the derivation.

The imaging and spatial coherence material in TU Delft's Interactive Optics develops the continuous-source version of the same phasor sum.

Notebook practice · Why a grating makes images of itself in empty space

Optical fields, instruments and observations

Trace an optical field through a physical instrument. Calculate the signal a detector measures and test the approximations on which the interpretation rests.

1. ## What an objective must collect to resolve a specimen

   A microscope can transmit plenty of light and still erase fine detail. Treat the specimen as a weak grating, follow its spatial harmonics through the objective, and determine why changing illumination coherence changes the transfer of contrast.

2. ## From a slowly varying wave to a curved ray in a graded index

   A ray bends continuously in a glass whose index changes across its width. Derive where that ray equation comes from, predict the focusing length of a graded-index rod, and locate the points at which a smooth ray description ceases to determine an optical amplitude.

3. ## Measure rotation with two waves following the same path

   A stationary optical loop gives equal return times for opposite propagation directions in a reciprocal medium. When the loop rotates the beams meet their starting beamsplitter at different times. Derive the measurable differential phase and compare a passive interferometer with a ring resonator.

4. ## Dispersive pulse compression and a detector's view of a tilted pulse

   An ultrashort pulse can broaden because its spectral components acquire different group delays or because different parts of its beam arrive at different times. Derive the two mechanisms separately, predict when added glass can shorten an already chirped pulse, and decide which measurements distinguish them.

5. ## Which polarisation measurements identify an optical element?

   A retarder rotates the state of a coherent beam, while unresolved mixtures can reduce its measured degree of polarisation. Derive the passage from Jones fields to a Mueller matrix, reconstruct a simple instrument from measured intensities, and identify claims the available input states cannot support.

6. ## Photon counting, bunching and the time resolution of a coincidence test

   An intensity trace can be steady while individual detector clicks show correlations. Derive the observable of a two-detector Hanbury Brown–Twiss measurement, quantify how background and bin width change it, and decide whether a dip near zero delay belongs to the source or to a detector.

7. ## Build a time lens from dispersion and a quadratic temporal phase

   A temporal imaging system can stretch the time separation of two optical features without requiring them to arrive more slowly at the source. Derive the imaging condition from delay and frequency, calculate the sign of time magnification, and estimate the finite temporal aperture that limits which features can be recovered.

8. ## Recover an optical phase from one off-axis intensity frame

A camera measures intensity and discards field phase. A known tilted reference wave can shift a cross term away from the intensity background so the missing complex field becomes accessible. Derive how much camera bandwidth that costs, and convert the recovered phase into an optical-path measurement with a stated ambiguity.

9. Resolve two nearby wavelengths by watching fringes disappear

A Michelson interferometer shows very fine bright and dark fringes as one mirror moves. Two narrow spectral lines produce a much slower variation of the fringe contrast. Use the mirror travel between contrast minima to determine the separation of the lines.

10. Where a thick lens focuses and which aperture an instrument actually sees

The distance from a lens mount to a sharply focused sensor is often shorter than the focal length printed on the barrel. Two principal planes, together with the image of the physical aperture stop, reconcile these measurements and determine which rays an instrument accepts.

11. From measured wavefront error to coma, astigmatism and focus

A point source can become a comet-shaped patch or two perpendicular line foci. Describing these observations by a single blur diameter loses the distinction between aberrations and suggests the wrong correction. The gradient of a pupil wavefront predicts the ray displacement at the detector.

12. Why light can generate its second harmonic and then lose it again

A nonlinear polarisation can oscillate twice during one period of the applied field. Whether the generated wave grows through a crystal depends on how successive source layers arrive in phase. This spatial phase relation makes crystal length and dispersion as important as local nonlinearity.

13. How a material changes the speed, attenuation and direction of light

A detector behind a sample loses light both when matter absorbs energy and when light is scattered outside its acceptance. A microscopic oscillator connects refraction to absorption. A radiating dipole then shows how a transparent particle can also reduce the measured direct beam.

Optical fields, instruments and observations · 1. What an objective must collect to resolve a specimen

A microscope can transmit plenty of light and still erase fine detail. Treat the specimen as a weak grating, follow its spatial harmonics through the objective, and determine why changing illumination coherence changes the transfer of contrast.

Diffract a sinusoidal object into orders

Use a thin specimen with amplitude transmission $t(x) = 1 + \epsilon\cos(2\pi f x)$, $|\epsilon| \ll 1$, and uniform normally incident coherent light. Write the cosine as half the sum of two exponentials. Immediately after the sample the field has spatial frequencies zero and $\pm f$. A lens maps transverse spatial frequency into angle, and its finite numerical aperture $\mathrm{NA} = n\sin\theta_{\max}$ accepts $|f| \le \mathrm{NA}/\lambda_0$ for coherent normal illumination, where λ_0 is the vacuum wavelength. Beyond that limit both first orders are blocked, leaving only a constant image field and no sinusoidal contrast.

If the zero order and both sidebands pass through a clear ideal pupil, inverse Fourier transformation reproduces $1 + \epsilon\cos(2\pi f x)$. Its intensity is approximately $1 + 2\epsilon\cos(2\pi f x)$. If only one sideband and the zero order pass, their interference still produces a feature at f, but the first-order intensity contrast is halved for this object and the image field acquires a spatially varying phase. If the zero order is suppressed as well, the intensity contrast follows a different mixing law. These possibilities show why the statement that one needs two interfering orders must specify which orders and which image quantity is measured.

Oblique coherent illumination shifts the object's spatial spectrum by the incident transverse frequency f_i. A sideband initially outside the pupil can enter while one from the opposite direction exits. With illumination aperture $\mathrm{NA}_{\mathrm{illum}}$ and objective aperture $\mathrm{NA}_{\mathrm{obj}}$, the extreme accessible object frequency from angular coverage can reach $(\mathrm{NA}_{\mathrm{illum}} + \mathrm{NA}_{\mathrm{obj}})/\lambda_0$. That support limit alone does not guarantee uniform contrast or isotropic information. The illumination distribution and reconstruction determine how well any frequency inside the support can be measured.

Derive the incoherent transfer from a pupil autocorrelation

For spatially incoherent emitters, intensities from different object points add. If h is the coherent amplitude point-spread function, the incoherent intensity point-spread function is $|h|^2$, and the intensity image is its convolution with object brightness. Fourier transforming the product $|h|^2$ produces an autocorrelation of the pupil P. A clear circular pupil gives a transfer value proportional to the overlap of two equal disks displaced by a spatial-frequency-dependent distance. At zero shift the disks overlap fully; at a diameter of shift their overlap vanishes. The incoherent cutoff is therefore $f_c = 2\mathrm{NA}/\lambda_0$, twice the coherent on-axis field cutoff.

For $u = |f|/f_c \le 1$, the normalised overlap area is

$$\mathrm{MTF}(u) = \frac{2}{\pi}\left[\cos^{-1}u - u\sqrt{1-u^2}\right].$$

To derive it divide the overlap lens into two symmetric disk segments. Each segment is a circular sector of angle $2\arccos u$ with its isosceles triangle subtracted. Summing and dividing by the full area yields the expression above. It declines continuously and gives about 0.391 at half the incoherent cutoff. The coherent transfer of a clear pupil has unit field transmission inside its support; it operates on complex amplitude. These

two curves describe different imaging experiments. Replacing a coherent field by its intensity before applying a coherent pupil would use the wrong linear system.

For $\lambda_0 = 0.550\,\mu\mathrm{m}$ and $\mathrm{NA} = 0.80$, the coherent cutoff is 1.45 cycles per micrometre and the incoherent cutoff 2.91 cycles per micrometre. An incoherent sinusoidal fluorescent brightness pattern at $f = 2.0$ cycles per micrometre has $u = 0.6875$ and ideal transmitted contrast about 0.20 of its original. A perfectly transmitting objective cannot recover the missing higher frequencies merely by magnifying the image farther. Magnification can, however, change camera sampling and therefore recover detail that the objective passed but undersampled camera pixels discarded.

Separate diffraction, sampling and noise in a resolution estimate

A detector with physical pixel pitch p at magnification M samples the specimen at p/M. The Nyquist frequency is $M/(2p)$. For $p = 6.5\,\mu\mathrm{m}$, $M = 40$ this gives 3.08 cycles per micrometre, just above the illustrative incoherent cutoff. Finite pixel area also averages over a square patch and multiplies the axial-frequency contrast by a sinc factor $\sin(\pi f p/M)/(\pi f p/M)$. At 2.0 cycles per micrometre that factor is about 0.835. The optical and pixel factors multiply, leaving less measured contrast than an optical-only transfer curve predicts. A sinusoidal target in a diagonal orientation sees the product of the factors in the two camera axes.

A cutoff is a support statement. Discrimination in real noise requires a contrast or likelihood criterion, a measured background and a known photon count. Deconvolving a frequency where the transfer is 0.1 multiplies noise in that component by roughly ten when dividing by the transfer. At a true zero no such division can restore the missing datum without additional assumptions or measurements. A point-pair Rayleigh separation, the highest sinusoidal frequency transmitted, and precision in locating the centre of an isolated bright emitter answer three different quantitative questions.

Compare the pupil models directly

The live graph plots the normalised coherent field pupil and incoherent intensity modulation transfer on the same specimen-frequency axis, with a separately identified pixel-averaged incoherent curve. Change numerical aperture while holding camera sampling fixed, then change wavelength. Predict whether optical cutoff or camera Nyquist is limiting before reading the returned values. The rectangular coherent curve should be interpreted as amplitude transmission for normally incident coherent light, whereas the smoother curves act on brightness contrast of incoherent emitters.

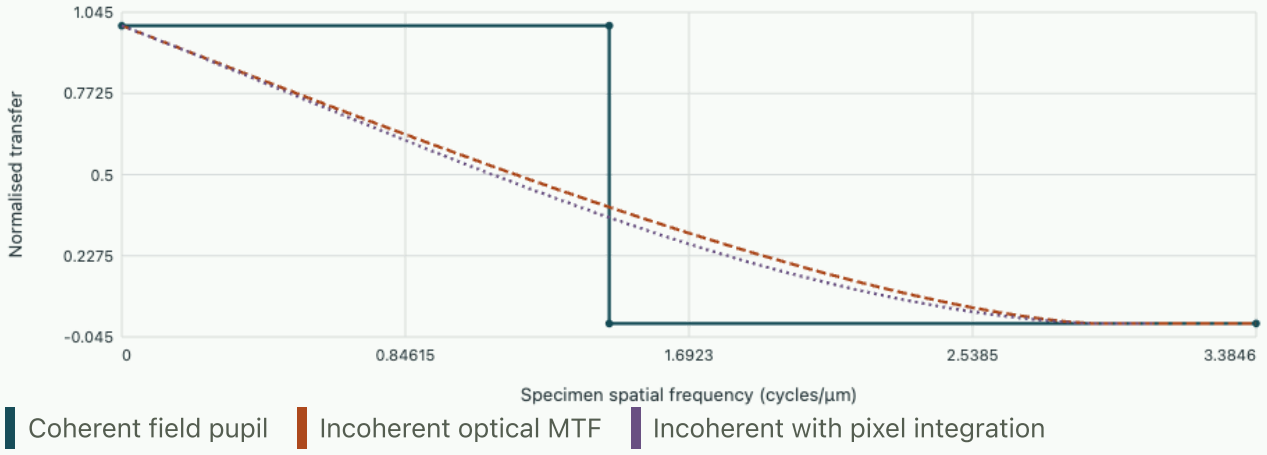
Open this experiment online

Try the calculation

Numerical aperture **0.8**

Vacuum wavelength (nm) **550**

Magnification **40**



Coherent field cutoff 1.45, incoherent optical cutoff 2.91, camera Nyquist 3.08 cycles/µm. The sampled curve ends at Nyquist.

Measure transfer with a sinusoidal calibration target

For an incoherent test pattern of mean brightness I_0 and intrinsic modulation m_o, measure the image modulation as $m_i = (I_{\max} - I_{\min})/(I_{\max} + I_{\min})$ after correcting the additive detector background. In a linear shift-invariant imaging regime the ratio m_i/m_o estimates the magnitude of the total transfer at the target frequency. A constant stray background reduces the measured modulation even if the pupil is unchanged. It increases the denominator without increasing the numerator, so subtract its independently measured intensity before comparing with a diffraction formula.

For a target of modulation 0.60 producing corrected maximum and minimum signals 550 and 450 counts, the image modulation is 0.10, so the inferred transfer is 0.167. The earlier optical and pixel factors at two cycles per micrometre predict approximately this value. A target with sharp bars contains many harmonics rather than one pure frequency. Its peak-to-trough contrast cannot generally be inserted into the same single-frequency formula without accounting for those harmonics. A sinusoidal target or a measured Fourier coefficient gives a cleaner comparison with the pupil calculation.

Compare diffraction conventions across the pupil model

When checking a spatial transfer against a diffraction calculation, keep the vacuum wavelength, refractive index and numerical aperture together: index already enters through numerical aperture. Inserting the wavelength in the medium a second time in a formula written with vacuum wavelength would count that index twice. The distinction matters most for immersion objectives and for comparisons of cutoffs reported from different beam geometries.

The open TU Delft diffraction chapter develops the wave-optical treatment of apertures and optical systems.

Tutorial on resolution claims

A normal-incidence coherent image of a grating with period $0.50\ \mu\mathrm{m}$ is uniform under the example objective. Can changing only illumination angle help?

▼ Work through the solution

The spatial frequency is 2.0 cycles per micrometre, beyond the on-axis coherent limit 1.45. An illumination shift of order 0.55 cycles per micrometre can bring one first order inside while retaining the reference order, restoring some contrast. The shift corresponds to illumination NA about 0.30 at the stated wavelength. Other orientations can respond differently.

At $20\times$ magnification with the same camera, which limit applies to an incoherent 2.0-cycle-per-micrometre pattern?

▼ Work through the solution

The specimen-plane sampling pitch is $0.325\ \mu\mathrm{m}$, giving Nyquist 1.54 cycles per micrometre. Optics can transmit the pattern, but the camera aliases it. Increasing magnification can help until sampling ceases to be limiting; it cannot raise the optical transfer at a fixed specimen frequency.

In an ideal coherent microscope a weak pure-phase grating has $t(x) \approx 1 + i\epsilon\cos(2\pi f x)$. Does passing all three orders guarantee first-order intensity contrast in focus?

▼ Work through the solution

No first-order intensity term survives because the perturbation is in quadrature with the reference. The field modulation is present but the intensity changes only at higher order in this approximation. Introduce a controlled reference phase shift, defocus, or another phase-sensitive measurement to convert the phase modulation into brightness contrast. Distinguish absence of an order from absence of interference in the detected quadrature.

TU Delft's Advanced Optical Imaging course derives coherent image formation from the pupil.

Optical fields, instruments and observations · 2. From a slowly varying wave to a curved ray in a graded index

A ray bends continuously in a glass whose index changes across its width. Derive where that ray equation comes from, predict the focusing length of a graded-index rod, and locate the points at which a smooth ray description ceases to determine an optical amplitude.

Expand the wave equation in inverse optical frequency

In a transparent isotropic nonmagnetic medium, a scalar monochromatic field can satisfy $\nabla^2 U + k_0^2 n(\mathbf{r})^2 U = 0$, with $k_0 = 2\pi/\lambda_0$. Assume n and a slowly varying envelope change over distances long compared with a local wavelength. Seek $U = A(\mathbf{r}) e^{i k_0 S(\mathbf{r})}$. Substitution gives a leading term $k_0^2 A[n^2 - |\nabla S|^2]$, a next term $i k_0 [2\nabla A \cdot \nabla S + A \nabla^2 S]$, and $\nabla^2 A$. Vanishing of the leading term yields the eikonal equation $|\nabla S| = n$. The gradient of optical path is the local normal to a phase front.

For a real leading amplitude the next term gives $\nabla \cdot (A^2 \nabla S) = 0$: flux through a thin ray tube is conserved to this order. In a one-dimensional monotone geometry with rays along z, $S' = \pm n$ and the travelling-wave amplitude scales as $n^{-1/2}$. Its field intensity parameter $|A|^2$ falls as index rises while the transported flux proportional to $n|A|^2$ stays constant. At an abrupt interface a reflected wave is generally needed to satisfy matching conditions, so the single-branch envelope law cannot determine the reflected power from a sudden index step.

The locally one-dimensional adiabatic criterion can be written $|k'|/k^2 \ll 1$ with $k = k_0 n$, accompanied by adequate smoothness of higher derivatives. Near a turning point the relevant longitudinal or transverse component of k can vanish even when n itself is finite. Then the WKB amplitude proportional to that component's inverse square root diverges. A finite-wave treatment gives a reflected and penetrating field. Thus merely comparing vacuum wavelength with the overall lens diameter cannot certify the ray approximation at every point.

Derive the graded-index ray equation and paraxial solution

A ray makes the optical path $\int n\, ds$ stationary. With longitudinal axis z and transverse coordinate $x(z)$, the integrand is $n(x, z)\sqrt{1 + x'^2}$. Its Euler–Lagrange equation is

$$d[n x'/\sqrt{1 + x'^2}]/dz = n_x \sqrt{1 + x'^2}.$$

For small slopes and a smooth index $n(x) \simeq n_0[1 - g^2 x^2/2]$, retain the leading transverse restoring term. The equation becomes $x'' + g^2 x = 0$, with g in inverse metres. The axis has the highest index, and rays curve back toward it. Solving gives

$$x(z) = x_0 \cos(gz) + \alpha_0 \sin(gz)/g,$$

where $\alpha_0 = x'(0)$ is the internal paraxial slope.

A full spatial pitch is $2\pi/g$, and a quarter-pitch length $\pi/(2g)$ transforms entrance position into angle and entrance angle into exit position. Collimated input rays with $\alpha_0 = 0$ cross the axis together at a quarter pitch. Rays launched from an axial point $x_0 = 0$ instead emerge parallel there since $x' = \alpha_0 \cos(gz)$ vanishes. These are reciprocal ray maps. At a flat end face outside the rod, refraction changes the external slope

according to the index ratio, so an internal angle-to-position scaling should not be directly applied to the angle of an external beam without that conversion.

For $g = 200\,\mathrm{m}^{-1}$ and $n_0 = 1.50$, the quarter-pitch length is $7.85\,\mathrm{mm}$. A ray launched at $x_0 = 0.40\,\mathrm{mm}$ parallel to the axis reaches it with slope $-g x_0 = -0.080$, compatible with a paraxial starting estimate. The assumed index depression at that entrance radius is

$$n_0 g^2 x_0^2/2 = 0.0048.$$

If a measured rod reaches only 0.002 depression at the same radius, either g was overstated or the parabolic model is inaccurate. An independent transverse index map and a measured pitch provide complementary checks.

Connect ray turning points to confined wave modes

For a guide uniform in z, seek $U(x,z) = u(x)e^{i\beta z}$. The scalar Helmholtz equation gives

$$u'' + [k_0^2 n^2(x) - \beta^2]u = 0.$$

The transverse local wave number $k_x = \sqrt{k_0^2 n^2 - \beta^2}$ becomes imaginary outside the classical turning points. Away from them an oscillatory WKB mode has amplitude $1/\sqrt{k_x}$ and phase $\int k_x dx$; the evanescent tail decays as $\exp[-\int |k_x| dx]$. Matching through two simple smooth turns gives the approximate condition $\int_{x_-}^{x_+} k_x dx = (m + 1/2)\pi$. The half-integer arises from the turning-point phase shifts and differs from the phase shifts at two rigid reflecting walls.

Under $n^2(x) \simeq n_0^2(1 - g^2 x^2)$, the transverse equation takes harmonic-oscillator form with quadratic confinement. Direct polynomial-Gaussian normalisability yields

$$k_0^2 n_0^2 - \beta_m^2 = (2m+1) k_0 n_0 g.$$

For small mode index relative to $k_0 n_0/g$,

$$\beta_m \simeq k_0 n_0 - (m + 1/2) g.$$

Adjacent propagation constants then differ by nearly g, so their relative phases refocus at $2\pi/g$, exactly the ray pitch. A real finite core and a departure from parabolic index spoil exact equal spacing and can broaden a refocused image or pulse.

Estimate the wave scale of a smooth turning region

Near a transverse turning point x_t suppose $k_x^2(x) \simeq -\gamma(x - x_t)$, where $\gamma > 0$ has units of inverse length cubed and the oscillatory side is $x < x_t$. The transverse equation becomes $u'' - \gamma(x - x_t)u = 0$. Rescale $X = (x - x_t)\gamma^{1/3}$; the resulting dimensionless equation is $u_{XX} - Xu = 0$, whose decaying Airy solution matches a sinusoid with a definite quarter-phase shift on the oscillatory side. Thus the width of the turning region is $l_t = \gamma^{-1/3}$. The wave does not reflect at an infinitely sharp geometric point, even for a perfectly smooth deterministic index profile.

If k_x^2 changes by $10^{10}\,\mathrm{m}^{-2}$ over $10\,\mu\mathrm{m}$, the slope scale is $10^{15}\,\mathrm{m}^{-3}$ and l_t is roughly $10\,\mu\mathrm{m}$. A detector averaging over a larger transverse spot cannot establish the detailed penetration profile from a direct image. Increasing the index gradient makes the turning region narrower only as its inverse cube root. A ray sketch may still predict the trajectory well over millimetres while missing a micrometre-scale reflected-wave phase or modal shift. Compare the length scale of the particular observable with the turning region before replacing the local wave equation by a specular bounce.

A second reflection from a nearby turning point can overlap this Airy region. The usual two independent turning phases are then no longer justified. Solve the transverse wave equation with both sides at once or compare a matched model with the actual index profile and verify normalisable boundary conditions.

Check the paraxial assumptions against a measured ray map

Measure entrance and exit positions for several collimated beams launched at progressively larger distances from the axis. The ideal paraxial solution gives a straight line of exit versus entrance position with slope $\cos(gL)$. Curvature in that plot can reveal aberration or departures from parabolic index. Comparing opposite launch offsets separates symmetric aberration from an off-centre guide or tilted coordinate origin.

Tutorial on a graded-index instrument

A guide with $g = 200\,\mathrm{m}^{-1}$ has length $5.0\,\mathrm{mm}$. A ray enters at $x_0 = 0.30\,\mathrm{mm}$ with zero internal slope. Predict its exit position and slope.

▼ **Work through the solution**

The phase advance is $gz = 1.00$. Thus

$$x = 0.30\cos 1 = 0.162\,\mathrm{mm},$$

$$x' = -200(0.00030)\sin 1 = -0.0505.$$

The ray is still on its original side of the axis and moving inward. Treating this rod as an exact quarter-pitch element would predict a premature focus.

Using $\lambda_0 = 0.60\,\mu\mathrm{m}$ and the example n_0, g, estimate the approximate Gaussian ground-mode width scale $w_0 = (k_0 n_0 g)^{-1/2}$.

▼ **Work through the solution**

Here $k_0 = 1.047 \times 10^7\,\mathrm{m}^{-1}$, so $w_0 = 17.8\,\mu\mathrm{m}$. The transverse field is proportional to $\exp[-x^2/(2w_0^2)]$ in this convention. A quoted intensity $1/e^2$ radius uses a different numerical factor. The computed width should be much smaller than the parabolic region for this ideal mode to be valid.

The ray-tube transport equation predicts unbounded amplitude at the common ray focus. Which measurement shows what the approximation omitted?

▼ Work through the solution

Measure the finite transverse focal spot while changing wavelength and aperture. Wave diffraction limits resolution even when all geometric rays meet at one point. The leading transport approximation conserves flux as ray-tube area shrinks but cannot describe interference and phase near a caustic. A wave propagation or mode superposition through the same index map should produce a finite profile and permit a direct comparison with those measurements.

TU Delft derives optical rays and geometrical imaging.

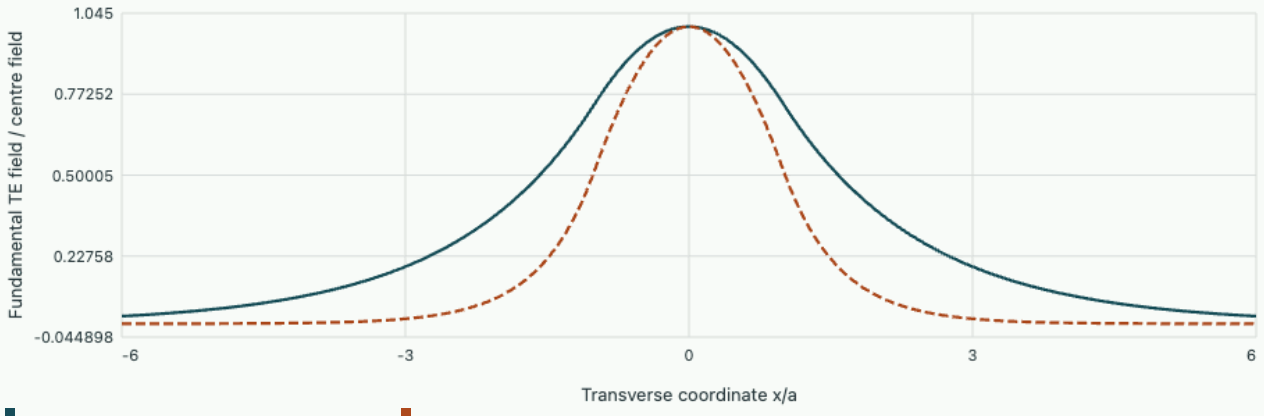
Compare smooth index turning points with modes confined at an index jump.

Open this experiment online

Try the calculation

Normalised half-width V

1



Root u = 0.7391, cladding exponent w = 0.6736. Field amplitude decays by 1/e over 1.485 half-widths outside the core.

Optical fields, instruments and observations · 3. Measure rotation with two waves following the same path

A stationary optical loop gives equal return times for opposite propagation directions in a reciprocal medium. When the loop rotates the beams meet their starting beamsplitter at different times. Derive the measurable differential phase and compare a passive interferometer with a ring resonator.

Calculate the return times in a rotating circular loop

Take a vacuum circular loop of radius R rotating at constant angular speed Ω in its plane, with a co-moving source, beamsplitter and receiver. In an inertial frame the receiver travels at $v = \Omega R$. A beam propagating in the direction of rotation must catch up with the moving receiver, while the oppositely directed beam meets it early. For an ideal constrained circular optical path, $ct_+ = 2\pi R + vt_+$ and $ct_- = 2\pi R - vt_-$. Thus $t_\pm = 2\pi R/(c \mp v)$, and their difference is $4\pi Rv/(c^2 - v^2)$. To first order in v/c this is $4\Omega A/c^2$, with $A = \pi R^2$.

The times just calculated are laboratory coordinate times. A clock on the rotating detector accumulates proper time with factor $\sqrt{1 - v^2/c^2}$; this changes the first-order result only at higher order. The local vacuum speed remains c for either beam. The effect comes from comparing their arrival at a moving part of a closed apparatus. There is no requirement to assign a different local fundamental light speed to the clockwise and anticlockwise beams.

A general co-moving loop at low rotational speed gives an extra time of flight $dt_{\mathrm{corr}} = \mathbf{v}(\mathbf{r}) \cdot d\mathbf{l}/c^2$ along one orientation. Reversing orientation doubles its contribution to the difference. With $\mathbf{v} = \boldsymbol{\Omega} \times \mathbf{r}$,

$$\Delta t = \frac{2}{c^2} \oint (\boldsymbol{\Omega} \times \mathbf{r}) \cdot d\mathbf{l} = \frac{4}{c^2} \boldsymbol{\Omega} \cdot \mathbf{A},$$

$$\mathbf{A} = \frac{1}{2} \oint \mathbf{r} \times d\mathbf{l}.$$

The oriented area handles tilted loops and folded paths. A figure-eight winding with two equal oppositely oriented lobes can enclose a large total unsigned area yet have zero net response to uniform rotation. Winding a fibre around the same area N times in the same orientation multiplies the directed area by N.

Convert the time shift into optical phase

A continuous source of vacuum wavelength λ_0 accumulates an optical phase difference $\Delta\phi = \omega_0 \Delta t = 8\pi \boldsymbol{\Omega} \cdot \mathbf{A}/(\lambda_0 c)$ at leading order. For a co-moving nondispersive dielectric loop the reciprocal static propagation time increases with index, but the standard first-order geometric Sagnac phase at a fixed external optical frequency retains this scale; applying relativistic velocity addition to propagation in the moving medium accounts for the cancellation. Flow of the medium relative to the loop, time-dependent index disturbances and nonreciprocal phase shifts can introduce additional effects.

For a hundred turns of radius $R = 0.10\,\mathrm{m}$, total directed area is $3.142\,\mathrm{m}^2$. At $\lambda_0 = 1.55\,\mu\mathrm{m}$ and $\Omega = 0.10\,\mathrm{rad/s}$ perpendicular to the loop, the time difference is only $1.40 \times 10^{-17}\,\mathrm{s}$, yet the phase is $0.0170\,\mathrm{rad}$. Direct electronic time-of-flight timing would be exceptionally demanding, while interference converts this small delay into an intensity change. Tilting the normal by 60° relative to rotation halves the signal and reversing the winding or rotation reverses its sign.

With equal returning intensities, a detector output can be written $I = I_{\mathrm{mean}}[1 + \mathcal{V}\cos(\Delta\phi + \phi_b)]$, where $\mathcal{V}$ is visibility and ϕ_b a controlled bias. Near $\phi_b = 0$ a small rotation changes intensity only quadratically and does not identify its sign. At quadrature, $\phi_b = \pi/2$, the change is approximately $-I_{\mathrm{mean}}\mathcal{V}\Delta\phi$. Alternating the applied bias or using both complementary outputs can reject a common power drift. A measured percentage intensity change only directly represents phase when the visibility and operating point are known.

Derive a ring resonator's beat instead of a passive phase

In an active vacuum ring laser or a tracked passive ring cavity, each propagation direction can have a separate resonant frequency satisfying an integer round-trip phase. In the simple nondispersive vacuum case let perimeter be P, unperturbed time $T = P/c$, and resonance frequency $\nu = m/T$. A small directional time difference Δt gives a resonance frequency splitting of magnitude $\Delta\nu \simeq \nu\Delta t/T = 4A\Omega_{\perp}/(\lambda_0 P)$. This readout measures a beat frequency. The passive interferometer measured a single-pass phase difference; the dimensions and perimeter dependence of the two formulas distinguish them.

A square vacuum ring of side $1.0\,\mathrm{m}$ has $A = 1.0\,\mathrm{m}^2$, $P = 4.0\,\mathrm{m}$. At $633\,\mathrm{nm}$ with $\Omega_{\perp} = 10^{-5}\,\mathrm{rad/s}$, the predicted directional splitting is $15.8\,\mathrm{Hz}$. Double every side at fixed wavelength and angular speed: area quadruples but perimeter doubles, so the beat doubles. In a physical active ring, backscattering can couple the counterpropagating modes and produce a dead band near low rotation. A zero observed beat does not uniquely show a stationary loop unless this coupling and other bias effects have been characterised.

Estimate a rotation sensitivity from counted photons

At a balanced quadrature readout with visibility $\mathcal{V}$, a small phase changes the mean differential count divided by total count by roughly $\mathcal{V}\Delta\phi$. For independently detected photons and a shot-noise-limited balanced measurement the standard deviation of that normalised difference is approximately $1/\sqrt{N}$, giving a phase uncertainty near $1/(\mathcal{V}\sqrt{N})$. This estimate assumes the photons used in the signal actually reached the detectors and that gain mismatch, intensity noise and nonreciprocal drift do not dominate. It is a count budget, separate from the geometric scale relating phase to rotation.

For $\mathcal{V} = 0.80$ and $N = 10^8$ detected photons the ideal phase uncertainty is $1.25 \times 10^{-4}\,\mathrm{rad}$. The hundred-turn example then has projected-rate uncertainty about $7.4 \times 10^{-4}\,\mathrm{rad/s}$ over that acquisition interval. Repeating statistically independent acquisitions improves the random part by the square root of their number, but cannot reduce a persistent reference bias in the same way. Rotating the coil normal through a known angle changes the physical signal as a cosine and can help identify additive biases. The calibration must use the actual normal orientation and total directed area, rather than assuming the apparatus remains level.

A longer fibre winding increases area at fixed turn radius while also increasing attenuation and susceptibility to distributed disturbances. The ideal phase scale grows with turn count; the number of detected photons and long-time noise can change at the same time. Compare those quantities independently when evaluating a different winding.

Tutorial on rejecting false rotations

What happens if a rigid loop translates at constant velocity without rotating in the first-order geometric calculation?

▼ **Work through the solution**

Its velocity is constant around the path, and $\oint \mathbf{v} \cdot d\mathbf{l} = \mathbf{v} \cdot \oint d\mathbf{l} = 0$. There is no corresponding Sagnac area signal. A moving segment of a nonrigid loop can have a different line integral; the stated cancellation is for common translation of the entire closed path.

A passive interferometer with visibility 0.80 reports a quadrature change $\Delta I / I_{\mathrm{mean}} = -0.0080$ for the hundred-turn geometry above. Estimate the projected rotation.

▼ **Work through the solution**

At quadrature the phase change is $0.0100\,\mathrm{rad}$. The example scale is $0.170\,\mathrm{rad}$ of phase per radian per second, giving $\Omega_{\perp} \simeq 0.0589\,\mathrm{rad/s}$. Use the actual directed area and vacuum wavelength. A visibility correction or a bias angle error propagates directly into the rotation estimate.

Can a drifting temperature distribution mimic rotation in a geometrically reciprocal fibre loop?

▼ **Work through the solution**

Two counterpropagating waves encounter a given fibre segment at different times. A static reciprocal phase shift cancels to leading order, but a changing local index can give different integrated phases on the two traversals. Compare runs with reversed physical rotation, symmetric thermal conditions and a calibrated artificial phase bias. Winding arrangement and timing of disturbances help distinguish a true oriented-area response from a transient nonreciprocal measurement error.

The Sagnac interferometer chapter hosted by Rutgers derives the phase shift of the two counterpropagating beams.

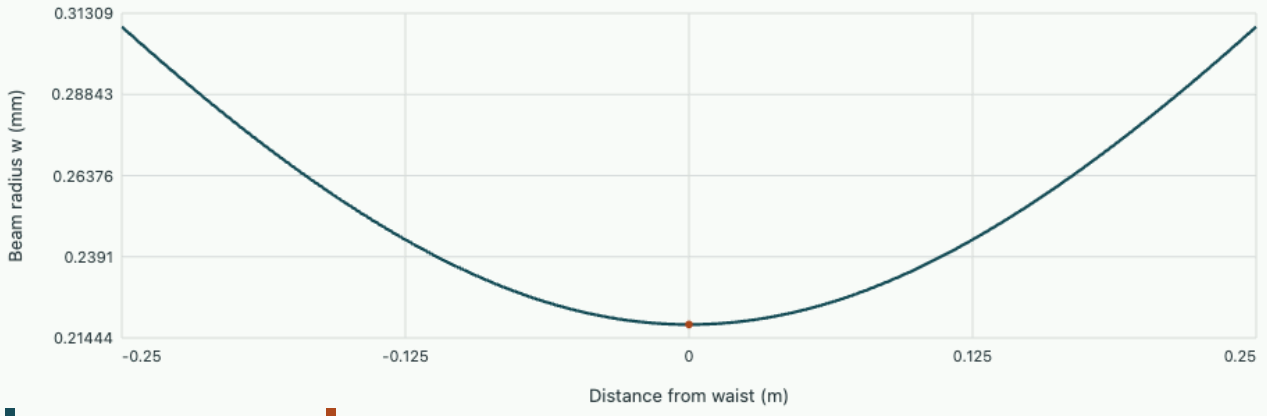
Compare a resonator's mode conditions with the directional frequency readout.

Open this experiment online

Try the calculation

Mirror separation / curvature radius **1**

Wavelength (nm) **600**



Gaussian beam radius Waist

Rayleigh length 0.25 m, waist radius 0.2185 mm, mirror spot radius 0.309 mm.

Optical fields, instruments and observations · 4. Dispersive pulse compression and a detector's view of a tilted pulse

An ultrashort pulse can broaden because its spectral components acquire different group delays or because different parts of its beam arrive at different times. Derive the two mechanisms separately, predict when added glass can shorten an already chirped pulse, and decide which measurements distinguish them.

Derive the propagation of a spectral envelope

Write the positive-frequency electric field as the real carrier with complex envelope $A(z,t)$, using the Fourier convention $A(t) \propto \int \tilde{A}(\Omega) e^{-i\Omega t} d\Omega$,

$$\Omega = \omega - \omega_0.$$

Propagation through a transparent homogeneous length L multiplies the spectrum by $e^{ik(\omega)L}$. Expand $k = k_0 + k_1\Omega + k_2\Omega^2/2 + \cdots$. The constant is an overall phase, the linear term shifts the pulse by group delay $k_1 L$, and the quadratic term gives group-delay dispersion $D = k_2 L$ in seconds squared. The group-delay difference between spectral components separated by $\Delta\Omega$ is $D\Delta\Omega$.

Take a Gaussian spectral amplitude $\tilde{A}(\Omega) = A_0 \exp[-\tau_0^2\Omega^2/2]$, and give it an initial quadratic spectral phase $\phi_2\Omega^2/2$. The effective quadratic coefficient after propagation is $\Phi = \phi_2 + D$. Fourier transforming the Gaussian at the retarded time $T = t - k_1 L$ gives an envelope proportional to $(\tau_0^2 - i\Phi)^{-1/2} \exp[-T^2/(2(\tau_0^2 - i\Phi))]$. Its intensity is Gaussian with temporal scale $\tau = \tau_0\sqrt{1 + (\Phi/\tau_0^2)^2}$, where intensity is proportional to $\exp[-T^2/\tau^2]$. The intensity full width at half maximum is $2\sqrt{\ln 2}\,\tau$; keeping this convention explicit prevents confusing amplitude widths, intensity widths and rms durations.

The intensity spectrum is unchanged by quadratic phase in a lossless element. Temporal broadening reduces the peak at fixed integrated pulse energy by τ_0/τ. Its duration depends on Φ^2, so one measured width at a single setting cannot determine the sign of chirp. Measuring the width after a known added dispersion does break that degeneracy. The full temporal phase contains a quadratic term giving instantaneous frequency that changes approximately linearly through the pulse. A transform-limited Gaussian has $\Delta t_{\mathrm{FWHM}} \Delta\nu_{\mathrm{FWHM}} = 2\ln 2/\pi \simeq 0.441$ when frequency is in cycles per second.

Calculate an imperfect compressor setting

Start with a transform-limited duration $30.0\,\mathrm{fs}$, so $\tau_0 = 18.0\,\mathrm{fs}$ and $\tau_0^2 = 325\,\mathrm{fs}^2$. A spectral phase of $+400\,\mathrm{fs}^2$ broadens this pulse to about $47.6\,\mathrm{fs}$, with peak intensity about 0.630 of its transform-limited value. Passing it through a component that contributes $-300\,\mathrm{fs}^2$ leaves $\Phi = 100\,\mathrm{fs}^2$ and duration $31.4\,\mathrm{fs}$. The same component applied to a pulse initially at $-400\,\mathrm{fs}^2$ instead gives $-700\,\mathrm{fs}^2$ and substantially more broadening. A quoted dispersion accuracy without the initial sign does not establish compression.

Experimental scans are simplest in squared width. Vary known D and fit $\Delta t^2 = \Delta t_0^2[1 + ((D + \phi_2)/\tau_0^2)^2]$. The parabola has a minimum at $D = -\phi_2$. A non-Gaussian spectrum or a cubic spectral phase can produce pedestals, asymmetric profiles or a minimum width larger than the transform limit inferred from spectrum. In an intensity autocorrelation a Gaussian pulse has a correlation width $\sqrt{2}$ times its intensity width. Apply the correct deconvolution for the assumed shape, or measure a spectrally resolved temporal diagnostic when the shape itself is unknown.

Derive pulse-front tilt and the spatially averaged width

Suppose the centre arrival time varies across one transverse coordinate as $T_c(x) = sx$. The slope s has units of time per length. A grating or misaligned dispersive optic can correlate frequency, propagation angle and position. For a local pulse of intensity $\exp[-(t - sx)^2/\tau^2]$ and beam weighting $\exp[-x^2/w^2]$, a detector integrating the whole beam sees the temporal convolution of the local profile with the distribution of arrival centres. Independent Gaussian variances add, giving measured temporal scale $\tau_\mathrm{d}^2 = \tau^2 + s^2 w^2$. A single-pixel trace at fixed x remains narrow and shifts in time. A whole-beam trace broadens even when no additional spectral phase changes along any one local ray.

With $w = 0.50\,\mathrm{mm}$ in this intensity convention and $s = 40\,\mathrm{fs/mm}$, the extra scale is $20\,\mathrm{fs}$. A locally transform-limited $30.0\,\mathrm{fs}$ pulse has $\tau_0 = 18.0\,\mathrm{fs}$, so the detector width becomes

$$2\sqrt{\ln 2}\sqrt{18.0^2 + 20^2} = 44.8\,\mathrm{fs}.$$

Changing the aperture alters the broadened trace by changing the sampled transverse weighting. A hard slit does not produce a Gaussian weighting, so the variance addition remains applicable for independent position and local pulse shape but the full width at half maximum need not follow the same Gaussian conversion.

Angular dispersion can create such a delay gradient. Near the reference propagation axis a spectral plane wave carries transverse phase $k_x(\omega)x$. In the inverse transform the term $(dk_x/d\omega)\Omega x$ shifts the local envelope time by $(dk_x/d\omega)x$, giving $s = dk_x/d\omega$ under the adopted conventions. Keeping only $k_x(\omega_0)$ accounts for a tilted carrier phase but misses the tilted pulse front. Mapping cross-correlation delay versus transverse position separates this linear slope from uniform material dispersion.

Predict a compression scan and perturb its detection area

The live temporal traces compare a local pulse and the second trace integrated across a known Gaussian beam, with the same normalised pulse energy. Predict the added dispersion that gives the narrowest local pulse, then turn up the transverse delay slope without moving the compressor. Which curve keeps its width? A larger minimum in an area-integrated scan is compatible with spatial averaging even when a local sample reaches its transform limit. Use a second spatial sampling condition before fitting that extra duration entirely to a higher-order material phase.

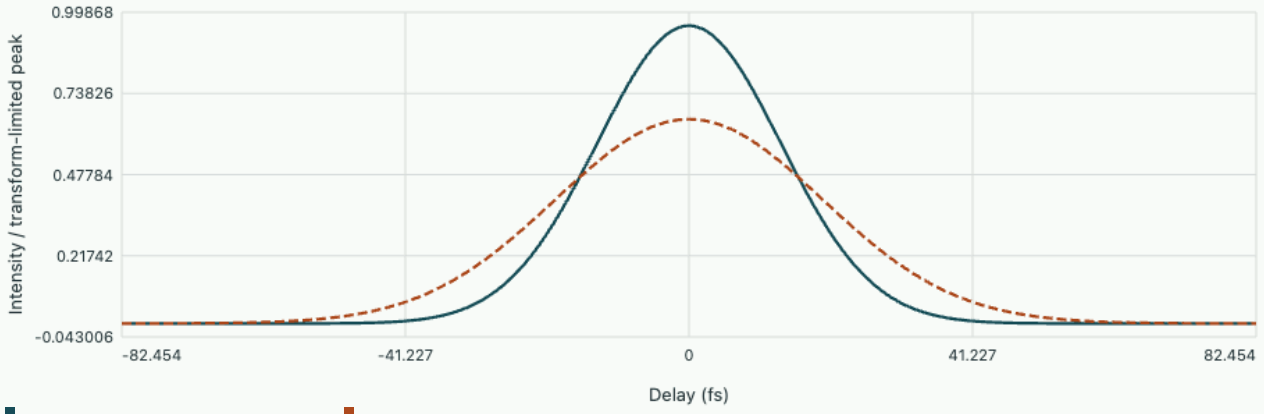
Open this experiment online

Try the calculation

Input spectral phase (fs²) **400**

Added dispersion (fs²) **-300**

Arrival slope (fs/mm) **40**



Local intensity FWHM 31.39 fs. Full-beam intensity FWHM 45.77 fs. Residual quadratic phase 100 fs².

Check the group delay with two narrow spectral slices

A pulse-width scan determines the magnitude of residual quadratic phase but can hide its sign at a single setting. Split off two narrow spectral slices centred at offsets $+\Delta\Omega/2$ and $-\Delta\Omega/2$, and compare their envelope delays against a common reference. To first order the arrival-time difference is $\Phi\Delta\Omega$. At $\Phi = +400\,\mathrm{fs}^2$ and an ordinary-frequency separation of $5.0\,\mathrm{THz}$, $\Delta\Omega = 0.0314\,\mathrm{rad/fs}$, giving a $12.6\,\mathrm{fs}$ delay difference. Under the adopted propagation convention the higher-frequency slice arrives later for positive dispersion.

The spectral filters themselves can contribute unequal group delays and broaden their selected pulses. Calibrate their differential delay using a reference path or a reference spectrum with independently characterised phase. Narrowing each slice improves its frequency selectivity while increasing its temporal duration, so a cross-correlation delay fit must retain adequate signal and temporal stability. Measuring several frequency centres permits a slope fit; visible curvature of delay against angular frequency indicates the need for a cubic or higher spectral phase. A camera position offset correlated with selected wavelength can add an apparent slope through pulse-front tilt, so compare at the same physical beam position as well as after spatial integration.

Tutorial on interpreting pulse widths

A $50\,\mathrm{fs}$ transform-limited Gaussian has a measured optical bandwidth of $5.0\,\mathrm{THz}$. Are the two measurements consistent if both are intensity full widths at half maximum?

▼ **Work through the solution**

Their product is 0.250, below the Gaussian transform-limited value 0.441. A $50\,\mathrm{fs}$ Gaussian requires about $8.82\,\mathrm{THz}$ of ordinary-frequency width. Check whether the reported bandwidth was angular frequency, an amplitude rather than intensity width, or clipped by the spectrometer. Pure spectral dispersion cannot reduce the time-bandwidth product below the transform limit.

The minimum duration occurs at $D = -250\,\mathrm{fs}^2$. What is the input quadratic phase, and what additional observation would show significant residual cubic phase?

▼ **Work through the solution**

The simple model gives

$$\phi_2 = +250\,\mathrm{fs}^2.$$

A minimum much longer than the transform limit for the measured spectrum, or asymmetric temporal satellites that persist near the minimum, motivates a higher-order spectral-phase measurement. Spatial averaging must be controlled separately before attributing every residual feature to cubic dispersion.

Two local cross-correlation peaks differ by $24\,\mathrm{fs}$ between transverse positions separated by $0.60\,\mathrm{mm}$. What extra rms time spread is expected across the $w = 0.50\,\mathrm{mm}$ Gaussian beam?

▼ **Work through the solution**

The slope is $40\,\mathrm{fs/mm}$. Since the transverse rms width is $w/\sqrt{2}$ for the stated intensity convention, the rms arrival spread is

$$40(0.50)/\sqrt{2} = 14.1\,\mathrm{fs}.$$

The corresponding Gaussian scale parameter $sw = 20\,\mathrm{fs}$ is larger by $\sqrt{2}$. Match the width convention to the formula used by the detector fit.

Kolner and Nazarathy connect dispersive evolution to temporal imaging.

Compare quadratic phase in temporal propagation with its spatial diffraction counterpart.

Optical fields, instruments and observations · 5. Which polarisation measurements identify an optical element?

A retarder rotates the state of a coherent beam, while unresolved mixtures can reduce its measured degree of polarisation. Derive the passage from Jones fields to a Mueller matrix, reconstruct a simple instrument from measured intensities, and identify claims the available input states cannot support.

Move from a field vector to a coherence matrix

For a monochromatic transverse field write the complex amplitudes as $\mathbf{e} = (E_x, E_y)^T$ with time convention $e^{-i\omega t}$. A deterministic optical element gives

$$\mathbf{e}_{\mathrm{out}} = J\mathbf{e}_{\mathrm{in}}.$$

A detector averaging over unresolved time, frequency or spatial channels instead sees a coherence matrix $C = \langle \mathbf{e}\mathbf{e}^{\dagger} \rangle$, which remains a 2×2 Hermitian positive semidefinite matrix. A fixed Jones element acts as $C_{\mathrm{out}} = JC_{\mathrm{in}}J^{\dagger}$. An incoherent mixture of distinct elements acts as $\sum_a p_a J_a C_{\mathrm{in}} J_a^{\dagger}$, generally not representable by one common Jones matrix.

Use the convention $S_0 = C_{xx} + C_{yy}$, $S_1 = C_{xx} - C_{yy}$, $S_2 = 2\Re C_{xy}$, $S_3 = 2\Im C_{xy}$. This explicitly sets the sign of the circular component independent of competing right-handedness naming conventions. The determinant is $[S_0^2 - S_1^2 - S_2^2 - S_3^2]/4$. Positive eigenvalues require $P = \sqrt{S_1^2 + S_2^2 + S_3^2}/S_0 \leq 1$ for nonzero signal. A pure Jones field gives rank one and $P = 1$; a mixture can give smaller P. A fixed lossy polariser can convert an unpolarised input into a fully polarised transmitted field, so attenuation and loss of polarisation are separate observations.

Derive the Mueller matrix of a retarder and an unresolved mixture

Write $C = \frac{1}{2}\sum_j S_j \tau_j$, with τ_0 identity, $\tau_1 = \mathrm{diag}(1, -1)$, τ_2 the real off-diagonal Pauli matrix, and τ_3 chosen with $+i$ in its upper off-diagonal entry for the stipulated convention. Since $\mathrm{tr}(\tau_i \tau_j) = 2\delta_{ij}$,

$$S_i = \mathrm{tr}(\tau_i C).$$

Inserting $JCJ^{\dagger}$ gives the real Mueller matrix $M_{ij} = \frac{1}{2}\mathrm{tr}(\tau_i J \tau_j J^{\dagger})$. A common phase on J cancels. These intensity measurements cannot determine a spatially uniform optical carrier phase, even if every Mueller entry is known.

For $J = \mathrm{diag}(1, e^{i\delta})$, C_{xy} is multiplied by $e^{-i\delta}$. Thus S_0, S_1 remain fixed while $S_2' = S_2 \cos\delta + S_3 \sin\delta$ and $S_3' = -S_2 \sin\delta + S_3 \cos\delta$. A 45° linear input has $(S_1, S_2, S_3)/S_0 = (0, 1, 0)$. A quarter-wave retardance sends it to $(0, 0, -1)$ in the stated sign convention; a half-wave retardance sends it to $(0, -1, 0)$. Measuring only horizontal and vertical powers would return equal readings for all three states. A diagonal analyser and a quarter-wave analyser are needed to distinguish them.

Now suppose half the unresolved beam spots undergo retardance $+\delta$ and half $-\delta$. Average their Mueller actions rather than averaging Jones fields. The S_2, S_3 cross terms cancel and both components acquire a factor $\cos\delta$. For $\delta = \pi/3$, the diagonal linear input becomes $P = 0.50$, while a horizontal input remains fully polarised. This is anisotropic depolarisation. Resolving the two spots would reveal two fully polarised local

states. A coherent recombination of amplitudes with a controlled phase is a different physical experiment and requires adding fields before averaging.

Reconstruct an element from independent incident states

Put four independent known incident Stokes vectors into columns of a matrix W, and the corresponding measured output vectors into Y. A linear optical response obeys $Y = MW$, so $M = YW^{-1}$ if W is nonsingular. Horizontal, vertical, diagonal-linear and a circular state with calibrated sign are one possible independent set. Four purely linear input states have $S_3 = 0$, so their columns span at most a three-dimensional subspace. No amount of repetition of only those input states reveals the fourth response column. Additional independent preparations are more informative than merely increasing nominal data volume.

For normalised inputs $H = (1, 1, 0, 0)$, $V = (1, -1, 0, 0)$, $D = (1, 0, 1, 0)$, $C_+ = (1, 0, 0, 1)$, call measured outputs y_H, y_V, y_D, y_C. The first column is $(y_H + y_V)/2$, the second $(y_H - y_V)/2$, the third $y_D - M_{:0}$, and the fourth $y_C - M_{:0}$. These differences assume known equal incident intensities or a correction to a common monitored intensity. A source that dims during one preparation can look like diattenuation. Measure an independent input monitor and repeat the preparations in a different order to distinguish drift from element properties.

Suppose the outputs are $(1, \pm 1, 0, 0)$ for H, V, $(1, 0, 0.50, 0)$ for D, and $(1, 0, 0, 0.50)$ for C_+. The reconstructed matrix is diagonal with entries $1, 1, 0.50, 0.50$, as in the unresolved equal-retardance example. The matrix establishes the action on Stokes states but does not uniquely identify that microscopic mechanism. A temporal fluctuator and an unresolved spatial mixture can give the same matrix. Change acquisition time, spectral bandwidth or spatial aperture to test which ensemble the detector is averaging.

Recover the retardance of a real plate with an imperfect analyser

Send a unit 45° linear input through the diagonal retarder above. Measuring the two independent output components gives

$$S_2/S_0 = \cos\delta,\ S_3/S_0 = -\sin\delta.$$

Thus $\delta = \operatorname{atan2}(-S_3, S_2)$ modulo 2π. A fit using only the diagonal analyser, and hence only $\cos\delta$, cannot distinguish opposite retardances and becomes insensitive near a maximum or minimum. The circular analyser supplies the missing quadrature. If the beam is partially depolarised equally in these two components, the angle from their ratio can remain well determined while their magnitude shrinks.

Suppose calibrated output readings give

$$S_2/S_0 = 0.48,\ S_3/S_0 = -0.64,$$

with negligible S_1. The inferred retardance is $\operatorname{atan2}(0.64, 0.48) = 53.1^\circ$, while the degree of polarisation is 0.80. A single ideal Jones retarder acting on a pure input at one frequency could produce the angle but would preserve unit degree; something in the measured ensemble or instrument reduced it. A pure retarder model fitted only to S_2 would instead infer $\arccos 0.48 = 61.3^\circ$, wrongly absorbing the contrast loss into its phase.

An analyser with leakage can also reduce the measured Stokes components. A differential pair of orthogonal polarisers with identical leakage fraction η scales its normalised difference by $(1-\eta)/(1+\eta)$. Calibrate each channel using known pure preparations and monitor intensity drifts before assigning all contrast loss to the specimen. Repeating with narrower spectral bandwidth or a smaller beam spot then tests the ensemble dependence of the inferred retardance and depolarisation.

Separate common optical phase from relative retardance

Inserting an isotropic plate can add the same phase to both orthogonal components of every coherent input. Its ideal Mueller matrix is unchanged because J and $e^{i\alpha}J$ act identically on the coherence matrix. An interferometer with an external reference arm can nevertheless detect the added phase. This experiment supplies precisely the reference that a standalone polarimeter lacks. The two instruments can thus disagree about whether anything changed while measuring distinct physical observables accurately.

Tutorial on polarimetric inference

An analyser reports intensities $I_H = 0.60,\ I_V = 0.40,\ I_D = 0.75,\ I_A = 0.25$ in common units. Which Stokes entries follow, and what remains unknown?

▼ **Work through the solution**

The first pair gives

$$S_0 = 1.00, S_1 = 0.20;$$

the diagonal pair gives $S_2 = 0.50$. No circular analysis was performed, so S_3 remains unknown within $|S_3| \le \sqrt{1 - 0.20^2 - 0.50^2} = 0.843$. Assuming a pure state would set its magnitude to that bound but would still not determine its sign. A partially polarised beam can have any smaller magnitude.

An ideal deterministic $J = \mathrm{diag}(\sqrt{0.8}, \sqrt{0.2})$ transmits an unpolarised input with unit incident intensity. What are the output intensity and degree of polarisation?

▼ **Work through the solution**

The input coherence matrix is half the identity. The transmitted diagonal entries are 0.40 and 0.10, so $S_0 = 0.50, S_1 = 0.30$ and $P = 0.60$. The element selectively transmits one component of the mixture. It has not generated optical energy; half the incident intensity was lost.

A fitted Mueller matrix maps one measured positive input into $P = 1.08$. How should the inference be checked?

▼ **Work through the solution**

That output violates positivity for this convention. Check dark subtraction, intensity normalisation, handedness of the circular analyser and conditioning of the preparation and analysis matrices. Collect more

independent states with uncertainties and fit within a physical response model. Clipping each output degree to unity separately need not produce one common linear physical Mueller map for the instrument.

Reconstruct a Stokes state numerically from analyser intensities.

TU Delft's optical course covers polarisation states and their measurement.

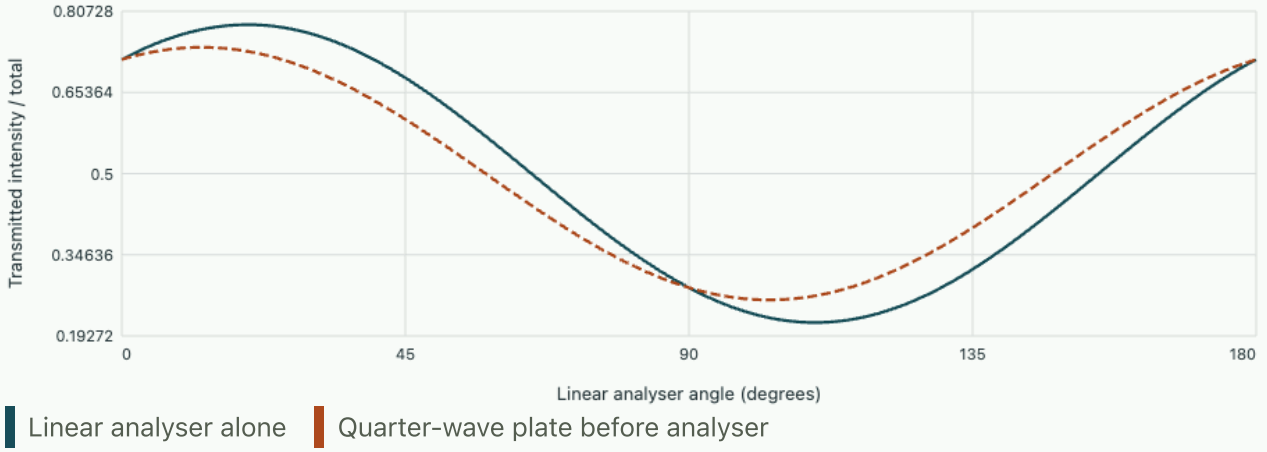
Open this experiment online

Try the calculation

Polarised fraction P **0.6**

Azimuth ψ (degrees) **20**

Ellipticity χ (degrees) **10**



Measured complementary pairs (horizontal, diagonal, circular): 0.716 / 0.284 · 0.6812 / 0.3188 · 0.6026 / 0.3974. Reconstructed $(S_1,S_2,S_3)/S_0$ = (0.4319, 0.3624, 0.2052); degree 0.6.

Optical fields, instruments and observations · 6. Photon counting, bunching and the time resolution of a coincidence test

An intensity trace can be steady while individual detector clicks show correlations. Derive the observable of a two-detector Hanbury Brown–Twiss measurement, quantify how background and bin width change it, and decide whether a dip near zero delay belongs to the source or to a detector.

Derive the coincidence baseline from two count streams

Split light on a beamsplitter and send the two outputs to separate photon-counting detectors. Write their mean count rates as R_1, R_2. For stationary independent streams observed for time T, the expected number of pairs in a delay bin of width Δt near zero is $R_1 R_2 T \Delta t$ when the bin is short relative to T. Each click in arm one defines a short opportunity of duration Δt in arm two. At long delays near the edges of a finite record, the available pair exposure is reduced; explicitly accounting for overlap time avoids a spurious triangular envelope.

The normalised second-order correlation $g^{(2)}(\tau)$ is the measured coincidence density divided by that independent baseline. In quantum notation it uses a normally ordered intensity product, so one photon cannot register as two distinct photons at the opposite outputs of an ideal splitter. For one occupied mode at equal time,

$$g^{(2)}(0) = \langle n(n-1)\rangle / \langle n\rangle^2.$$

A number state with exactly one photon has zero; a coherent field with Poisson photon statistics has one; one ideal thermal mode gives two. A low mean count rate alone implies none of these cases: attenuation lowers the count rate but independent loss ideally cancels from the normalised ratio.

For $R_1 = 40\,\mathrm{kcounts/s}$, $R_2 = 30\,\mathrm{kcounts/s}$, $T = 100\,\mathrm{s}$, and $\Delta t = 1.0\,\mathrm{ns}$, the expected uncorrelated baseline is 120 counts per bin. A measured 60 counts gives a raw correlation of 0.50. A simple Poisson estimate of the numerator uncertainty is $\sqrt{60}/120 = 0.065$ before baseline, dead-time and other uncertainties. Shrinking the bin can increase temporal contrast but reduces the counts per bin, so reporting a very small $g^{(2)}(0)$ without the actual counts or uncertainty conceals an important limitation.

Calculate the effect of uncorrelated backgrounds

In arm i, let $R_i = S_i + B_i$ with signal fraction $\rho_i = S_i/R_i$. Assume both backgrounds are Poissonian, independent of each other and of the source, and that counting response is linear. Every coincidence involving a background contributes only its independent baseline, while true signal–signal pairs carry the nontrivial source correlation. Expanding the product gives

$$g^{(2)}_{\mathrm{obs}}(\tau) = 1 + \rho_1 \rho_2 [g^{(2)}_{\mathrm{signal}}(\tau) - 1].$$

For equal signal fractions 0.80, an ideal single-emitter zero-delay dip reaches only 0.36 in an otherwise perfect instrument. The same background reduces a thermal single-mode peak from two to 1.64. A correction using ρ_i estimated at another alignment or pump power can overcorrect. Correlated background,

detector crosstalk and blinking invalidate the independent-background relation or add longer-delay structures. Measure backgrounds with the relevant source and optical settings whenever possible.

A simple stationary emitter with one exponential recovery time gives

$$g_s^{(2)}(\tau) = 1 - e^{-|\tau|/\tau_c}.$$

This shape can arise in a two-state incoherent pumping model because detecting an emitted photon resets the system to its ground state and the excited population recovers. A thermal bunching example has $g_s^{(2)} = 1 + e^{-|\tau|/\tau_c}$ under a corresponding exponential intensity-correlation assumption. The time constant used in the latter is the intensity-correlation time; it need not equal a field-amplitude coherence time. Multi-level emitters or multimode light require a different model.

Average the correlation over the time bin

A rectangular zero-centred bin reports its mean correlation, not the instantaneous value. For the exponential departure from unity the zero-bin attenuation is $A = (2\tau_c/\Delta t)[1 - e^{-\Delta t/(2\tau_c)}]$. It tends to one for a narrow bin and approximately $2\tau_c/\Delta t$ for a very wide one. An ideal anticorrelation with $\tau_c = 2\,\mathrm{ns}$, $\Delta t = 4\,\mathrm{ns}$ and signal fractions 0.80 in each arm yields

$$g_{\mathrm{bin}}(0) = 1 - 0.64(1 - e^{-1}) = 0.595.$$

Inferring two emitters solely because the observed value exceeds one half would ignore known attenuation by both background and timing.

Electronic timing jitter gives an additional convolution with the probability distribution of the difference of the two timing errors. For independent Gaussian jitters their variances add. The measured bin width and this jitter serve distinct roles: rebinning time-tagged data cannot undo detector jitter. Extremely large bins can make single-emitter and Poisson observations nearly indistinguishable even with many total counts. A calibrated timing response and a physically justified line model can constrain an underlying time scale, but extremely narrow intrinsic features can remain unresolved.

Compare photon statistics with detector statistics

A single detector with a recovery dead time after a click produces an artificial deficit of very short same-detector intervals even for coherent light. Using separate arms avoids that particular single-detector autocorrelation artifact. Each detector's saturation can still alter measured rates, and an electronic dead time shared between channels can create a false coincidence dip. Conversely optical or electronic crosstalk can create a false peak. Send a stable coherent reference through the complete acquisition chain at comparable rates to test the baseline, and vary attenuation or cable delays to separate features tied to absolute electronics timing from those tied to optical delay.

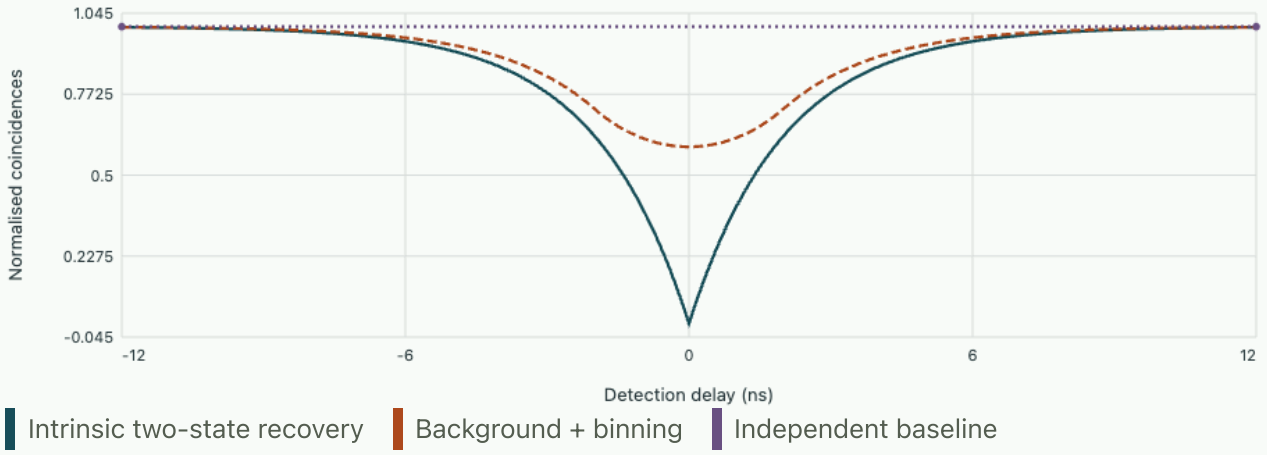
The live graph displays an ideal recovery dip, its background-attenuated and bin-averaged counterpart, and the independent baseline. Predict the zero-bin counts by multiplying the plotted normalised value by the expected independent counts for a chosen acquisition. Increase bin width and lower signal fraction

separately. Both reduce contrast, but only the finite bin averages over neighbouring delays. Measuring data with several bins from the same stored time tags tests that dependence without changing the source.

Open this experiment online

Try the calculation

Recovery time (ns)	2
Bin width (ns)	4
Signal fraction in each arm	0.8



The raw zero-bin correlation is 0.595. At 40 and 30 kcounts/s for 100 s, the expected zero-bin count is 285.8 against an independent baseline of 480.0.

Read pulsed coincidences as areas of a peak train

A pulsed source with repetition period much longer than its emission decay produces coincidence peaks separated by that period. The central peak counts two detections attributed to the same pulse; neighbouring peaks can estimate the independent probability from different pulses. Integrate comparable windows around those peaks rather than comparing single-bin heights if jitter spreads clicks. A normalised central area of 25 counts against a mean adjacent area of 100 gives raw $g^{(2)}(0) \simeq 0.25$ under identical per-pulse brightness and efficiency. The result can coexist with obvious timing spread in every individual peak.

Slow blinking or pulse-to-pulse brightness correlation makes neighbouring peaks poor independent references. Plot several side-peak areas versus separation and compare with the mean at delays longer than the brightness correlation time. A decaying envelope in those areas contains source information that one baseline value conceals. At high detection probability, detector recovery across successive pulses can also suppress one or more nearby peaks. Vary repetition period or attenuation while keeping the binning procedure fixed to distinguish that recovery from intrinsic photon statistics.

Tutorial on coincidence data

Two equally bright independent ideal single-photon emitters contribute to one unresolved beam. Ignoring background and jitter, what zero-delay correlation follows?

▼ Work through the solution

Each emitter cannot supply two simultaneous signal photons, while cross-emitter pairs remain possible. For N identical independent sources $g^{(2)}(0) = 1 - 1/N$. Two therefore give one half. If their intensities differ, the reduction is the sum of squared fractional intensities, so equal brightness is a physical assumption. Background and finite resolution can mimic a change of emitter number.

If both arms are attenuated by a factor two and the signal-to-background ratio stays fixed, how should the expected coincidence counts and normalised correlation change?

▼ Work through the solution

The independent baseline $R_1 R_2 T \Delta t$ falls by a factor four and correlated counts scale the same way for independent linear losses. Normalised $g^{(2)}$ stays the same but counting uncertainties grow. If detector dark counts stay fixed while the signal falls, signal fractions also fall and the normalised contrast decreases.

A measured zero-delay peak gives $g^{(2)} = 1.40$ for thermal light with both signal fractions 0.80 and negligible timing loss. What effective number of equally populated independent thermal modes would explain it?

▼ Work through the solution

For M independent equal modes the excess is $1/M$. Correcting background gives $g_s^{(2)} - 1 = 0.40/0.64 = 0.625$, so $M_{\mathrm{eff}} = 1.60$. The noninteger value can represent unequally weighted modes or unresolved averaging. It is an effective contrast parameter unless the mode distribution has been established independently.

Hanbury Brown and Twiss analyse intensity fluctuations and correlations.

Connect source operation above threshold with the separate measurement of photon statistics.

Optical fields, instruments and observations · 7. Build a time lens from dispersion and a quadratic temporal phase

A temporal imaging system can stretch the time separation of two optical features without requiring them to arrive more slowly at the source. Derive the imaging condition from delay and frequency, calculate the sign of time magnification, and estimate the finite temporal aperture that limits which features can be recovered.

Derive two elementary maps in delay and frequency

Describe a narrowband wave packet by its delay t relative to a reference group and angular-frequency offset $\Omega = \omega - \omega_0$. With Fourier convention $A(t) \propto \int \tilde{A}(\Omega) e^{-i\Omega t} d\Omega$, a dispersive element contributes spectral phase $D\Omega^2/2$. Its derivative is group delay $D\Omega$. A narrow local time-frequency feature obeys $t' = t + D\Omega,\ \Omega' = \Omega$. This is a shear in time-frequency space, directly analogous in algebra to propagation of a paraxial ray, but D has units of time squared and the second coordinate is angular frequency.

A time lens applies temporal phase $+Kt^2/2$ to the envelope. With this sign convention the instantaneous angular-frequency offset is minus the derivative of envelope phase, so $\Omega' = \Omega - Kt$ and $t' = t$. The coefficient K has units of inverse time squared. Producing this modulation requires a physical phase-modulating or nonlinear optical process synchronised to the signal. A passive time-invariant piece of glass cannot implement an arbitrary quadratic temporal phase on an unknown incident time profile; it acts on spectrum instead.

Derive the imaging condition and track an individual feature

Place dispersion D_1 before the time lens and D_2 after it. Start with (t_0, Ω_0). At the lens $t_\ell = t_0 + D_1\Omega_0$; just after it $\Omega_\ell = \Omega_0 - Kt_\ell$. The output time is

$$t_{\mathrm{out}} = (1 - KD_2)t_0 + (D_1 + D_2 - KD_1D_2)\Omega_0.$$

Sharp imaging requires the output time of an input feature to be independent of its initial frequency spread, so set $K = D_1^{-1} + D_2^{-1}$. Then $t_{\mathrm{out}} = Mt_0$ with magnification $M = -D_2/D_1$. This is the finite-conjugate time-lens equation. For positive pre- and post-dispersion the magnification is negative: a later input feature appears earlier relative to the image reference. This is temporal inversion about a chosen reference delay, not propagation of information backward in laboratory time. The whole apparatus has a positive baseline transit delay common to the processed waveform.

With $D_1 = 100\,\mathrm{ps}^2$, $D_2 = 400\,\mathrm{ps}^2$, the required lens curvature is $0.0125\,\mathrm{ps}^{-2}$ and $M = -4$. Two pulses initially separated by $8.0\,\mathrm{ps}$ emerge $32.0\,\mathrm{ps}$ apart in reversed relative order. A detector having a response width of $10\,\mathrm{ps}$ may now resolve the separation, provided that the optical imaging system itself supplied adequate temporal resolution. Ideal energy-conserving magnification rescales the intensity envelope by $1/|M|$ as it stretches time, apart from losses and a possible output phase. Multiplying every time separation by four without reducing its peak intensity would create optical energy.

If the curvature is 2% below nominal,

$$K = 0.01225\,\mathrm{ps}^{-2}.$$

Then the coefficient of Ω_0 is $500 - 0.01225(40000) = 10\,\mathrm{ps}^2$, and $1 - KD_2 = -3.90$. An input spectral spread $0.20\,\mathrm{rad/ps}$ contributes an output delay spread of order $2.0\,\mathrm{ps}$ before other aberrations. A calibration

pulse with a different spectral width can thus appear sharper even at the same imaging setting. Measure the output dependence on carrier detuning of a narrow feature to test the frequency coefficient directly.

Recover imaging from the wave integral and locate an aperture limit

The temporal propagation kernel of a quadratic spectral phase, up to constants and a uniform phase, is $\exp[-i(t-t')^2/(2D)]$ under the adopted transform convention. In the two-propagation integral multiply the first kernel, the lens phase $e^{iKt_\ell^2/2}$, and the second kernel. The coefficient of t_ℓ^2 is $(K - D_1^{-1} - D_2^{-1})/2$. At the imaging condition it vanishes. The remaining linear phase in t_ℓ is $t_\ell(t_0/D_1 + t_{\mathrm{out}}/D_2)$. Integrating over an unlimited lens time produces a delta condition $t_{\mathrm{out}} = -D_2 t_0/D_1$, recovering the ray-style map directly from wave interference. Quadratic input and output phase factors remain, even where the intensity envelope is correctly imaged.

A finite flat temporal aperture of total width T_a replaces the delta condition by a sinc amplitude. For a point feature its first zero in output delay is at $2\pi|D_2|/T_a$; referring it back to the input divides by $|M|$, giving $2\pi|D_1|/T_a$. With $T_a = 100\,\mathrm{ps}$, $D_1 = 100\,\mathrm{ps}^2$ this input first-zero scale is $6.28\,\mathrm{ps}$. A much larger time magnification cannot restore features the lens failed to collect over its temporal aperture. The criterion is a first zero of a rectangular-aperture field response; an intensity width or a Gaussian gate would use a different numerical factor.

Construct and calibrate an approximate phase profile

An electro-optic modulator driven sinusoidally can approximate a quadratic phase near an extremum. Expanding $\Phi_0\cos(\omega_m t)$ gives a constant, $-\Phi_0\omega_m^2 t^2/2$, then a quartic term $\Phi_0\omega_m^4 t^4/24$. Choose bias phase and sign to obtain the required K. The ratio of quartic to quadratic magnitude is approximately $(\omega_m t)^2/12$, so only part of the modulation period is a good lens. Widening the accepted temporal window improves nominal aperture resolution but can increase aberration in the wings. An independent phase calibration and a scan over input delays expose the tradeoff.

Distinguish magnification calibration from the image of a point

Launch two independently timed reference pulses flanking the centre of the field of view. If their input times are -5.0 and $+5.0\,\mathrm{ps}$, an ideal $M = -4$ instrument places their output times at $+20$ and $-20\,\mathrm{ps}$ relative to a common delay. Their $40\,\mathrm{ps}$ separation calibrates temporal magnification without requiring knowledge of an absolute cable or optical path delay. The widths of the two output features probe the point-spread response and detector resolution. A single pulse can demonstrate broadening but cannot by itself distinguish image magnification from an imperfectly focused, unmagnified pulse.

Vary K around the predicted imaging value and use two thin pulses with different carrier detunings at the same nominal time. The spread of their output positions is proportional to the frequency coefficient $D_1 + D_2 - KD_1D_2$, which changes sign through best focus. Once that coefficient vanishes, repeat with inputs at different reference delays to measure M. This sequence separates a focus error from a scale error. A medium with large cubic dispersion can make the zero coefficient depend on carrier detuning; sampling a third detuning diagnoses the failure of the purely quadratic matrix model.

The live phase-space calculation tracks several narrow spectral components across the device for selected input times. At the correct lens strength their output lines have equal output time at fixed input time. Detuning the lens gives them different output arrival times. Changing signs of D_2 distinguishes an inverted image from one that preserves time ordering.

Open this experiment online

Try the calculation

First dispersion (ps²) **100**

Second dispersion (ps²) **400**

Lens K (10^{-3} ps^{-2}) **12.5**



Temporal map coefficients A = -4.00, B = 0.00 ps². The image is focused at M = -4.00.

Tutorial on an imaging system for time

A signal traverses $D_1 = +100\,\mathrm{ps}^2$ and $D_2 = -200\,\mathrm{ps}^2$. Which lens curvature and magnification form an image?

▼ **Work through the solution**

The lens condition gives

$$K = 0.010 - 0.005 = 0.005\,\mathrm{ps}^{-2},$$

with $M = +2$. The output keeps the ordering of features while doubling their separation. The signs refer to group-delay dispersion and temporal lens phase under the chapter's Fourier convention.

The previous positive-dispersion system images two pulses well near the centre of its time window but distorts them near the edge. Does this demonstrate a wrong constant D_2?

▼ **Work through the solution**

A uniform error of the imaging condition gives residual sensitivity to input frequency throughout the window. Edge-specific distortion suggests finite gate clipping, nonquadratic temporal phase or higher-order dispersion acting on different parts of the signal. Translate the same reference pulse across the lens timing and independently sweep its carrier detuning to separate field-of-view and chromatic errors.

If the input contains a short feature at $t_0 = 0$, can its optical phase always be inferred from a measured intensity image?

▼ Work through the solution

The ideal system maps complex envelopes with known quadratic phase factors, but an intensity-only detector measures the squared magnitude. It cannot generally recover unknown input phase from that one image. A coherent reference or additional phase-diverse measurements can supply missing information. The time-lens condition guarantees a mapping of the field under ideal propagation, not complete observability of every input degree of freedom.

Kolner and Nazarathy's original paper derives temporal imaging with a time lens.

Compare selection by a finite spatial Fourier aperture with the temporal acceptance window.

Optical fields, instruments and observations · 8. Recover an optical phase from one off-axis intensity frame

A camera measures intensity and discards field phase. A known tilted reference wave can shift a cross term away from the intensity background so the missing complex field becomes accessible. Derive how much camera bandwidth that costs, and convert the recovered phase into an optical-path measurement with a stated ambiguity.

Expand the recorded intensity and identify four terms

Let $O(x)$ be the unknown coherent object field on the detector and $R(x) = A_r e^{i2\pi f_c x}$ a known reference plane wave. They must have sufficient mutual coherence at the relative path delay, and their polarisation projections on the analyser must overlap. A square-law camera records $I = |O|^2 + |A_r|^2 + O A_r^* e^{-i2\pi f_c x} + O^* A_r e^{i2\pi f_c x}$. The first two terms are a real background and the last two are complex conjugates making real interference fringes. A local phase shift in O changes the fringe position without requiring the camera itself to follow the optical frequency.

Suppose the complex object spectrum has support approximately $|f| \le B$. The spectrum of $|O|^2$ is its autocorrelation and can extend to $|f| = 2B$. The cross terms occupy intervals centred at $\pm f_c$ with half-width B. In an ideal one-dimensional separation, $f_c - B > 2B$ keeps a sideband clear of the central intensity spectrum. A camera with pitch p has Nyquist frequency $f_N = 1/(2p)$, requiring $f_c + B < f_N$. Both inequalities hold simultaneously only if $B < f_N/4$ under these conservative support assumptions. Some fields have much narrower background or sparse spectra, but a designed full-bandwidth instrument cannot assume that without evidence.

Filter the negative-frequency sideband containing $O A_r^*$, shift it to zero frequency, and inverse transform. Dividing by the known complex reference amplitude recovers the object field up to an overall calibration phase. Filtering the opposite sideband gives the conjugate field. A reference tilt calibration error leaves a linear phase ramp after reconstruction. Finite detector field of view multiplies the hologram by a window and spreads its Fourier peaks. Use a tapered window or leave guard bands at the cost of spatial resolution. The filter radius can suppress noise while also discarding high object frequencies; it is a physical resolution choice as well as a plotting parameter.

Choose a carrier that a real pixel lattice can record

Take a camera pitch $p = 6.5\,\mu\mathrm{m}$, giving $f_N = 76.9$ cycles per millimetre. If the field bandwidth is $B = 12$ cycles per millimetre, the conservative carrier interval is $36 < f_c < 64.9$ cycles per millimetre. A choice $f_c = 48$ cycles per millimetre places the selected sideband between 36 and 60, clear of the central interval extending to 24. The fringe period is $20.8\,\mu\mathrm{m}$, or 3.21 pixels. At vacuum wavelength $532\,\mathrm{nm}$ and propagation close to normal in air the tilt angle is $\arcsin(\lambda f_c) \approx 1.46^\circ$. Raising the angle indefinitely would eventually alias the sideband back into the usable spectrum.

Each camera pixel integrates across its physical sensitive area. A full-width square pixel attenuates a sinusoidal fringe amplitude by $\mathrm{sinc}(\pi f_c p)$, with $\mathrm{sinc}\, x = \sin x / x$. At the example carrier this factor is about 0.847. Frequencies on either side of the band receive different attenuation, so a very broad sideband can have a frequency-dependent instrumental response even inside Nyquist. A flat-field intensity correction fixes pixel-to-pixel gain, but does not undo this finite-area transfer. The sensor's usable dynamic range and read noise also constrain the reference amplitude.

Convert a reconstructed phase into a path change

Let the reconstruction relative to a reference frame provide $O_2O_1^* = |O_1O_2|e^{i\Delta\phi}$. The angle $\arg(O_2O_1^*)$ gives phase change modulo 2π, so $\Delta\mathrm{OPL} = \lambda\Delta\phi/(2\pi)$ modulo one vacuum wavelength. For transmission through a uniform nonabsorbing layer replacing ambient index n_a, an added thickness d gives optical-path change $(n - n_a)d$. For a normally reflecting surface displacement h the outgoing and return paths both change, giving $h = \lambda\Delta\phi/(4\pi)$ in air. Using the transmission conversion on reflection doubles inferred height.

A reflection measurement at $532\,\mathrm{nm}$ gives a local phase change $0.80\,\mathrm{rad}$, implying a height change $33.9\,\mathrm{nm}$ modulo $266\,\mathrm{nm}$ under the ideal normal-incidence geometry. In transmission through a layer of contrast $n - n_a = 0.10$, the same phase corresponds to $0.677\,\mu\mathrm{m}$ of thickness modulo $5.32\,\mu\mathrm{m}$. A single phase image cannot simultaneously determine thickness and refractive index without independent information. Phase unwrapping uses spatial or temporal continuity to choose integer cycles; at dark amplitude nodes or steep unresolved steps those choices can be ambiguous.

Background · How can a second wavelength extend the unambiguous path range?

Reconstruct the phase of the same optical path change d at two vacuum wavelengths. Assume the path change is independent of wavelength and both reconstructions share a calibrated zero. Each phase is known modulo 2π. Their difference increases more slowly with path than either phase alone.

$$\Delta\phi = 2\pi d\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right) \pmod{2\pi},$$

$$\Lambda = \frac{\lambda_1\lambda_2}{|\lambda_2 - \lambda_1|}.$$

Wavelengths of 630 and 640 nm give a synthetic path period of 40.32 µm. For normal reflection the physical surface displacement is half the round-trip path change. A 0.05 rad uncertainty in the difference phase corresponds to about 0.32 µm of path uncertainty at this synthetic period. The long period helps identify a fringe order, while a single-wavelength phase can refine the displacement once that order is known.

Measure noise and reference aberration separately

The fringe visibility for uniform signal and reference intensities is $2\sqrt{I_oI_r}/(I_o + I_r)$ when coherence and polarisation overlap are perfect. Equal intensities maximise visibility. Increasing reference power can raise heterodyne signal above camera read noise, but once reference shot noise dominates, continually increasing it does not improve sensitivity proportionally and can saturate pixels. A fixed total photon budget and a fixed object photon budget are different optimisation problems. Background subtraction should use a reference-only frame, an object-only frame when feasible, and a blank complex phase calibration acquired with the same optical geometry.

A wavefront aberration in the reference multiplies the recovered object by a spatially varying phase. Dividing the complex field by a blank-field reconstruction can remove a stable aberration, but division near very low blank amplitude amplifies noise. A phase map through masked pixels should disclose where phase was

interpolated and where the signal actually supported it. Reconstruct at several sideband filter radii and carrier angles to check whether a fine feature moves or changes shape with the processing. A real specimen feature should remain in specimen coordinates under those controlled changes within the valid resolution.

Decide whether a thin specimen can be measured without unwrapping

If surface steps in reflection are known a priori to lie within a single interval of width $\lambda/2$, the wrapped phase can be mapped uniquely into that chosen range up to a common reference height. For $532\,\mathrm{nm}$ a known $60\,\mathrm{nm}$ relief is safely smaller than the $266\,\mathrm{nm}$ height period. An isolated $300\,\mathrm{nm}$ step can instead appear as $34\,\mathrm{nm}$ if there is no continuity information across its edge. A smoothness algorithm can select precisely the wrong branch at a real discontinuity.

Two wavelengths offer a different reference scale. For nearby wavelengths λ_1, λ_2, subtract their interferometric phases for the same optical path to obtain a synthetic optical-path period $\lambda_1\lambda_2/|\lambda_1 - \lambda_2|$. Reflection height has half that period. With 532 and $540\,\mathrm{nm}$, the height period is about $18.0\,\mu\mathrm{m}$, much larger than for either colour alone. The corresponding height sensitivity to a given phase noise is worse, so the synthetic phase often chooses the coarse branch while one individual wavelength supplies fine precision. This assumes both measurements interrogate the same position and calibrated path without a significant wavelength-dependent sample response.

Tutorial on carrier and phase measurements

The same $6.5\,\mu\mathrm{m}$ camera is asked to recover $B = 22$ cycles per millimetre with full separation from an object intensity spectrum extending to $2B$. Is there any carrier satisfying both conditions?

▼ **Work through the solution**

The carrier would have to exceed 66 and be below $76.9 - 22 = 54.9$ cycles per millimetre. No such carrier exists. Change magnification to reduce object bandwidth on the detector, use smaller pixels, or exploit multiple phase-shifted frames under an appropriate measurement model. Merely tuning the tilt cannot create detector bandwidth.

A reconstruction of a flat mirror shows an apparent phase ramp of 2π across $5\,\mathrm{mm}$ after demodulation. What carrier correction would remove a pure calibration ramp?

▼ **Work through the solution**

A phase change of one cycle over $5\,\mathrm{mm}$ corresponds to 0.20 cycles per millimetre. Adjust the assumed carrier by that magnitude with sign selected from the observed phase slope, then repeat on the reference mirror. Correcting a sample ramp without a blank could accidentally remove a real tilt of the sample.

A wrapped phase map appears very smooth even though the object amplitude almost vanishes on one side of an image. Does smoothness certify accuracy?

▼ Work through the solution

Near zero complex amplitude the phase responds very strongly to noise. Smoothing or interpolation can produce a visually regular map without adding phase information. Examine the sideband signal-to-noise ratio, repeat at independent exposures, and state an amplitude-based confidence mask. An optical-path conversion is quantitative only in the region that supplied a reliably measured complex field.

Cuche, Marquet and Depeursinge demonstrate Fourier separation of off-axis holographic orders.

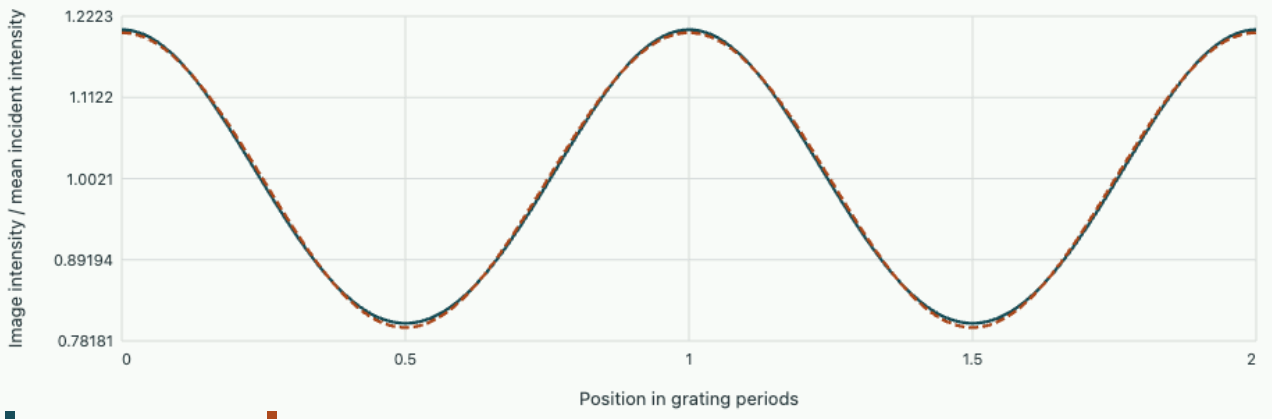
Vary a Fourier aperture and observe the resolution cost of filtering.

Open this experiment online

Try the calculation

Specimen phase amplitude ε (rad) **0.1**

Central order shift ϕ (degrees) **90**



Zero-order amplitude 0.9975. Peak-to-trough intensity difference 0.3983; linear prediction 0.4.

Optical fields, instruments and observations · 9. Resolve two nearby wavelengths by watching fringes disappear

A Michelson interferometer shows very fine bright and dark fringes as one mirror moves. Two narrow spectral lines produce a much slower variation of the fringe contrast. Use the mirror travel between contrast minima to determine the separation of the lines.

Combine two independent optical frequencies

Let d be mirror displacement from equal paths. The optical path difference of the returning beams is $2d$ and the phase at vacuum wavenumber $\sigma = 1/\lambda$ is $4\pi\sigma d$. Assume equal arm intensities for each individual line and an ideal detector that averages quickly oscillating optical fields and the beat between the distinct source frequencies. The measured intensities from the two spectral components add.

If the second line carries a fraction r of the first line's intensity, the modulated part of the signal divided by the total mean becomes

$$s = [\cos(4\pi\sigma_1 d) + r\cos(4\pi\sigma_2 d)]/(1+r).$$

Writing the two cosines as real parts separates a rapidly rotating phase from the complex envelope. Taking the magnitude produces the normalised visibility,

$$V(d) = \frac{\sqrt{1 + r^2 + 2r\cos(4\pi\Delta\sigma d)}}{1+r},$$

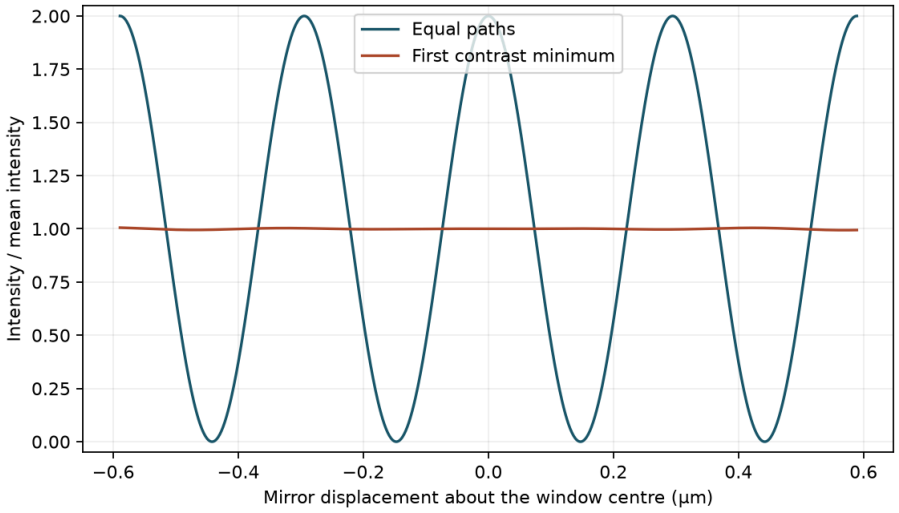
$$\Delta\sigma = |\sigma_1 - \sigma_2|.$$

This is the visibility obtained by scanning enough fine fringes to estimate local maxima and minima whilst the envelope changes negligibly. For equally strong lines it simplifies to $|\cos(2\pi\Delta\sigma d)|$. The first loss of contrast occurs at $d = 1/(4\Delta\sigma)$ and successive minima are spaced by $1/(2\Delta\sigma)$. The factor of two relative to optical path difference comes directly from reflection at the moving mirror.

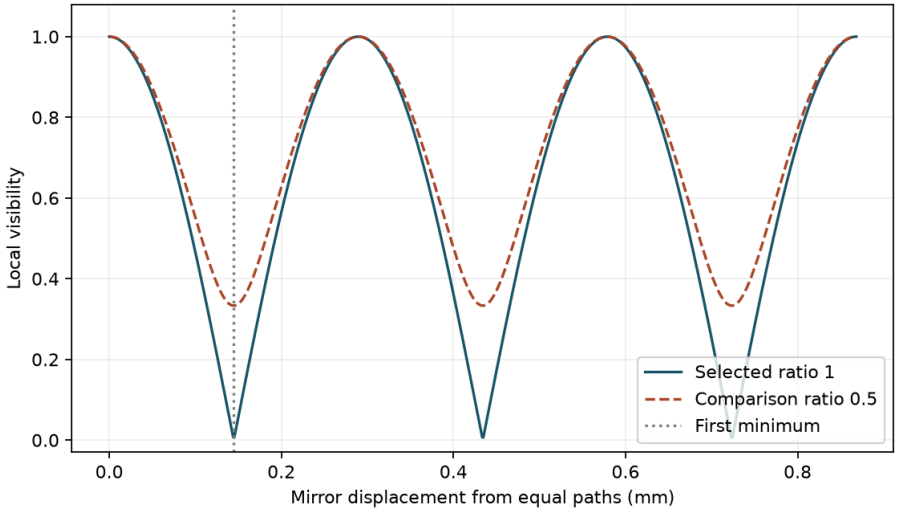
Compare mirror travel with a single optical fringe

Choose vacuum wavelengths 589.0 and 589.6 nm. Their wavenumber separation is $\Delta\sigma = 0.6/(589.0 \times 589.6)$ nm⁻¹, about 1.73×10^3 m⁻¹. The first visibility minimum lies at mirror displacement 0.145 mm and contrast revives at 0.289 mm. An individual mirror fringe spans only about 0.295 µm. Thus hundreds of carrier fringes fit inside a single envelope lobe. A coarse plot of raw intensity over a millimetre can alias those fast oscillations.

Suppose the second line carries half the intensity of the first. At the same minimum the two modulation vectors point in opposite directions. The surviving magnitude relative to their summed intensities is $|1 - 0.5|/(1 + 0.5) = 1/3$. An incomplete minimum can therefore arise even when both components are individually sharp. A finite width of either line or imperfect overlap of the arms in a real interferometer can also reduce the observed contrast. Independent measurements near equal path help identify such instrumental loss.



Each curve shows a short scan around the position indicated in its legend. Only the window near equal paths gives full contrast when the lines have equal intensity.



The separation between minima determines wavenumber splitting even when unequal line intensities prevent complete cancellation.

Exercise · infer a line separation from successive minima

A mirror moves 0.400 mm between two consecutive contrast minima. Estimate the wavelength separation for lines centred near 600 nm.

▼ **Work through the solution**

The wavenumber difference is $\Delta\sigma = 1/(2 \times 0.400\,\mathrm{mm}) = 1250$ m⁻¹. The narrow-doublet approximation $|\Delta\lambda| \simeq \lambda^2 \Delta\sigma$ gives 0.450 nm. Using the optical path difference as though it were mirror displacement would introduce a factor of two.

Exercise · infer the weaker line fraction

An ideal instrument observes maximum visibility one and minimum visibility 0.20. How strong is the weaker line relative to the stronger one? Can visibility label which line has the longer wavelength?

▼ **Work through the solution**

Calling the weaker-to-stronger ratio $r \leq 1$ gives $(1-r)/(1+r) = 0.20$, hence $r = 2/3$. The visibility magnitude is symmetric when the two lines are exchanged. Assigning the stronger intensity to a particular wavelength requires additional spectral information.

Try it · Doublet

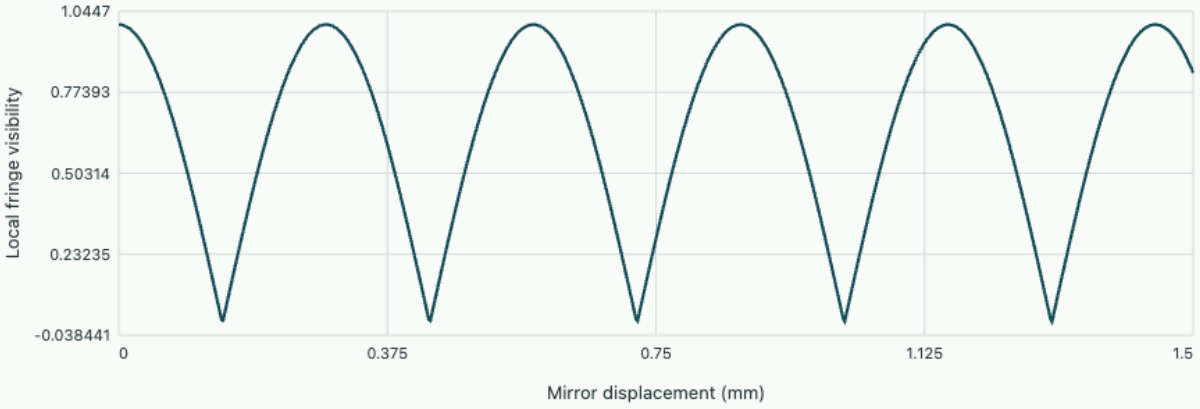
Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Wavelength splitting (nm) **0.6**

Second / first line intensity **1**



Incoherent spectral doublet

First contrast minimum 0.1447 mm; minimum visibility 0. Wavelength of first line is 589 nm.

Download the corresponding editable notebook. The browser recalculates the reduced equations directly; the notebook exposes the full calculation and its checks.

Open editable notebook Download .ipynb

Optical fields, instruments and observations · 10. Where a thick lens focuses and which aperture an instrument actually sees

The distance from a lens mount to a sharply focused sensor is often shorter than the focal length printed on the barrel. Two principal planes, together with the image of the physical aperture stop, reconcile these measurements and determine which rays an instrument accepts.

Compose refraction and translation with an explicit sign convention

Light travels left to right. Write a paraxial ray as (y, u) with $u = n\theta$. A positive radius has its centre of curvature to the right of the surface. Small-angle Snell refraction at a spherical surface gives $u_2 = u_1 - \Phi_i y$ with $\Phi_i = (n_2 - n_1)/R_i$. Translation by d in material of index n gives $y_2 = y_1 + (d/n)u_1$. Heights and slopes can therefore be propagated in the same two-component vector even when refractive index changes.

Multiply refraction at surface 1, internal translation through thickness d, and refraction at surface 2, writing the last transformation on the left. Let $\ell = d/n$. The elements of the vertex-to-vertex matrix in air are

$$M = \begin{pmatrix} 1 & 0 \\ -\Phi_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & \ell \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\Phi_1 & 1 \end{pmatrix} = \begin{pmatrix} 1 - \ell\Phi_1 & \ell \\ -\Phi & 1 - \ell\Phi_2 \end{pmatrix}, \qquad \Phi = \Phi_1 + \Phi_2 - \ell\Phi_1\Phi_2.$$

Its determinant is one. In the thin limit the surface powers add. Treating widely separated surfaces as one zero-thickness refractor discards the product term and generally changes both the power and where distances must be measured.

Factor the matrix and locate the cardinal points

For any air-to-air matrix with entries A, B, C, D and nonzero $\Phi = -C$, factor it into an air translation l_1 to the first principal plane, a thin lens of power Φ, and an air translation l_2 from the second principal plane to the back vertex. Matrix multiplication fixes $l_1 = (1 - D)/\Phi$ and $l_2 = (1 - A)/\Phi$. The effective focal length is $f = 1/\Phi$. Hence the back focal distance from the back vertex is $f - l_2 = A/\Phi$.

For a symmetric biconvex lens with $R_1 = +50$ mm, $R_2 = -50$ mm, $n = 1.50$ and $d = 10$ mm, each surface contributes 10 m⁻¹ and $\Phi = 19.333$ m⁻¹. The effective focal length is 51.72 mm. Each principal plane is 3.45 mm inside its adjacent vertex and the back focal distance is 48.28 mm. An object 200 mm left of the first principal plane gives an image 69.77 mm right of the second plane with lateral magnification -0.349.

Recall · What is fixed when the medium changes?

The expressions $f = 1/\Phi$ and unit angular magnification at nodal points here use equal unit indices in the incident and image media. With unequal external indices the front and back focal distances differ. An afocal system has $\Phi = 0$ and is handled directly by the ray matrix.

Identify the entrance and exit pupils

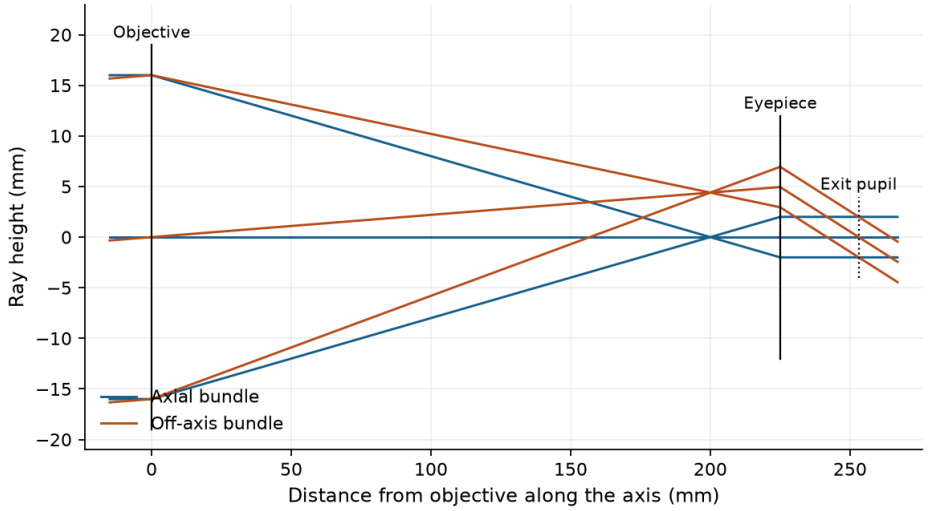
An axial point emits a bundle of possible rays. The element that first rejects an increasing cone angle in object space is the aperture stop. The entrance pupil is its image formed by the optics before it, viewed from the object. The exit pupil is its image formed by later optics, viewed from the image. A field stop constrains which off-axis object points are visible. It can be physically different from the aperture stop.

Put a 6 mm diameter iris 25 mm behind an air thin lens of focal length 50 mm. To find what an observer in front sees, trace rays backward from the iris. The thin-lens equation with propagation now to the left gives image distance -50 mm relative to the lens in that reversed convention. The stop therefore appears as a virtual 12 mm diameter opening 50 mm behind the lens. An axial object 150 mm in front of the lens sees a half-angle near $6/200 = 0.030$ rad. A sufficiently small lens rim would become the actual stop instead, invalidating this calculation.

Place the eye and relate field to throughput

An afocal two-positive-lens telescope separates its objective and eyepiece by $f_o + f_e$. A distant angle θ makes an intermediate height $f_o\theta$. The eyepiece then sends that point into an approximately parallel bundle at angle $-f_o\theta/f_e$. If the objective rim is the stop, its image through the eyepiece is the exit pupil. The eyepiece images an object at $f_o + f_e$ to distance $f_e(f_o + f_e)/f_o$ behind it. The exit-pupil diameter is $D_o f_e/f_o$.

With $f_o = 200$ mm, $f_e = 25$ mm and a 32 mm objective, angular magnification is -8, the exit pupil is 4 mm across and lies 28.1 mm behind the eyepiece. Moving a camera away from that plane can vignette off-axis bundles even with a sharp central image. A field diaphragm at the intermediate image separately clips the angular field.



Axial and off-axis paraxial bundles through the example telescope. Both fill the same 4 mm exit pupil at the indicated plane, while the eyepiece redirects the two bundles at different angles.

Try it · Telescope

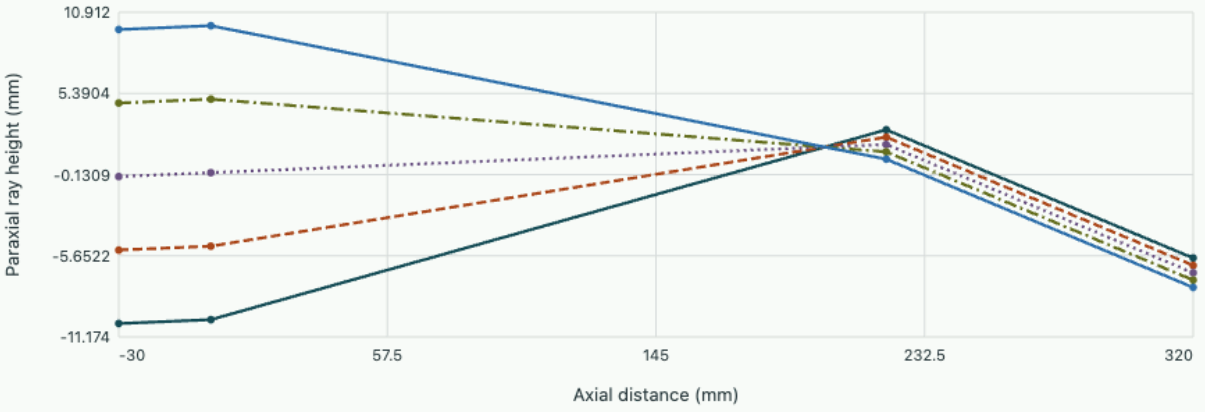
Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Eyepiece focal length (mm) 20

Field angle (degrees) 0.5



Objective height -10 mm | Objective height -5 mm | Objective height 0 mm | Objective height 5 mm | Objective height 10 mm

Afocal spacing 220 mm; angular magnification –10. A 20 mm objective gives an exit pupil 2 mm in diameter, 22 mm after the eyepiece.

Tutorial on lens-card and aperture measurements

1 · Focus a sensor from the vertex

▼ Work through the solution

In the thick-lens example the image for the 200 mm object is 69.77 mm from the second principal plane, hence 66.32 mm beyond the back lens vertex. A ruler referenced to the vertex must use the second value.

2 · Test whether the lens rim could be the stop

▼ Work through the solution

In the iris example an axial object at 150 mm and half-angle 0.030 rad uses a beam radius of about 4.5 mm at the lens. A 7 mm clear lens diameter clips it before the 6 mm iris can define the full bundle. Repeat the ray check for each potentially limiting aperture.

---

3 · Specify a telescope field stop

▼ **Work through the solution**

An intermediate focal-plane field stop of 8 mm diameter with the 200 mm objective gives an approximate full angular field $2(4/200) = 0.040$ rad or 2.29°. Halving its diameter halves this field. In a system without vignetting it leaves the axial exit-pupil diameter unchanged.

For further study see Hecht, Optics, 5th edition, chapters 5–6 on geometrical optics, cardinal planes and instruments.

Optical fields, instruments and observations · 11. From measured wavefront error to coma, astigmatism and focus

A point source can become a comet-shaped patch or two perpendicular line foci. Describing these observations by a single blur diameter loses the distinction between aberrations and suggests the wrong correction. The gradient of a pupil wavefront predicts the ray displacement at the detector.

Derive the image displacement associated with a pupil error

In a circular pupil of radius a write the normalised Cartesian coordinates $u = x/a$, $v = y/a$. A phase plate or optical error contributes a path difference $W(u, v)$ in units of length. Under monochromatic, paraxial conditions a pupil wave has the extra phase kW. Its local wave-vector tilt is $k^{-1}\nabla_{\perp}(kW) = \nabla_{\perp}W$. A focal-length f lens converts a small tilt into a displacement in its focal plane. With phase convention e^{ikW},

$$(\delta X, \delta Y) = \frac{f}{a}(\partial_u W, \partial_v W).$$

A constant piston gives no displacement. A linear tilt gives the same displacement to every ray. A quadratic rotationally symmetric term gives a radial displacement proportional to radius and can be cancelled by shifting the detector. The relation gives a geometrical spot construction. Near the diffraction limit, integrate the complex pupil amplitude to obtain the point-spread function.

Distinguish removable defocus from a fourth-order pupil error

On axis the pupil error of a centred rotationally symmetric lens has even powers of $\rho = \sqrt{u^2 + v^2}$. After choosing a focus the next independent term is $W = S\rho^4$. Its radial slope is $4S\rho^3$. Moving the detector can add a slope proportional to ρ and sharpen some zones while leaving others out of focus. Closing the aperture reduces the outer-radius contribution quickly, but increases the diffraction width on the sensor.

A lens whose full physical pupil is trimmed to half the radius still has the same physical curvature errors. For a fourth-power physical-height term, the edge value of W drops by 2^4 and its edge slope by 2^3. These scalings are measurements that can separate an aberrated lens from a diffraction-limited one.

Derive two off-axis spot signatures

Let the field point lie along the horizontal direction and let h be a dimensionless field coordinate. The leading coma term allowed by symmetry has the form $W_C = Ch\,u(u^2 + v^2)$. Differentiating gives

$$\delta X_C = (f/a)Ch(3u^2 + v^2),$$
$$\delta Y_C = (f/a)Ch(2uv).$$

At opposite edges of the horizontal diameter the horizontal error has the same sign. A round pupil therefore maps to an asymmetric fan with a brighter concentration of rays on one side. Reversing the field direction reverses the fan. With $f = 100$ mm, $a = 5$ mm and $Ch = 0.20$ µm, marginal horizontal rays shift by 12 µm.

After subtraction of a field-dependent common focus error an astigmatic term can be written $W_A = Ah^2(u^2 - v^2)$. Horizontal slopes vary as $+2Ah^2u$, vertical slopes as $-2Ah^2v$. Add a small detector defocus with radial slope of adjustable sign. It can bring either meridional family to its own line focus, but

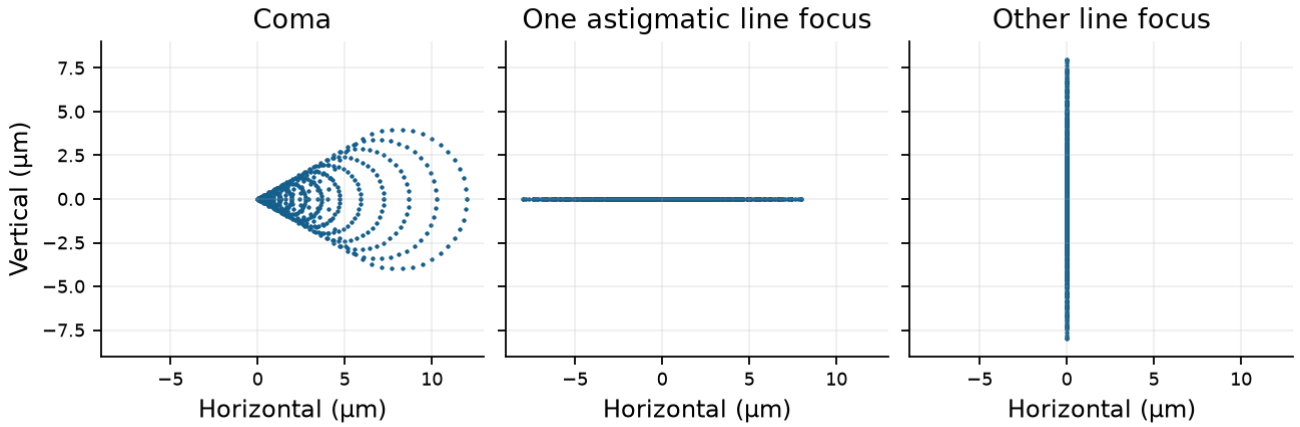
cannot make both vanish together. A longitudinal detector shift Δz changes a marginal ray position by about $a\Delta z/f$. Equating its magnitude to $2f|A|h^2/a$ gives separation of the two line foci

$$|\Delta z_{\text{pair}}| \simeq 4f^2|A|h^2/a^2.$$

For $|A|h^2 = 0.10$ µm and the same geometry this gives 0.160 mm. Translating the detector between the two foci rotates the long axis of the spot by roughly 90°. Reversing field leaves this term even, giving an independent distinction from coma.

Recall · Separate field curvature and distortion

A common quadratic pupil term varying with the square of field shifts the best image surface. A field-dependent pupil tilt shifts the image position without blurring a single field point in this approximation. Measure multiple field positions and focus settings before attributing all deviations to coma.



Ray-intercept diagrams from the pupil gradients in the worked example. Coma produces an asymmetric fan. Astigmatism produces two perpendicular line foci at distinct detector positions. Diffraction would broaden each geometrical spot.

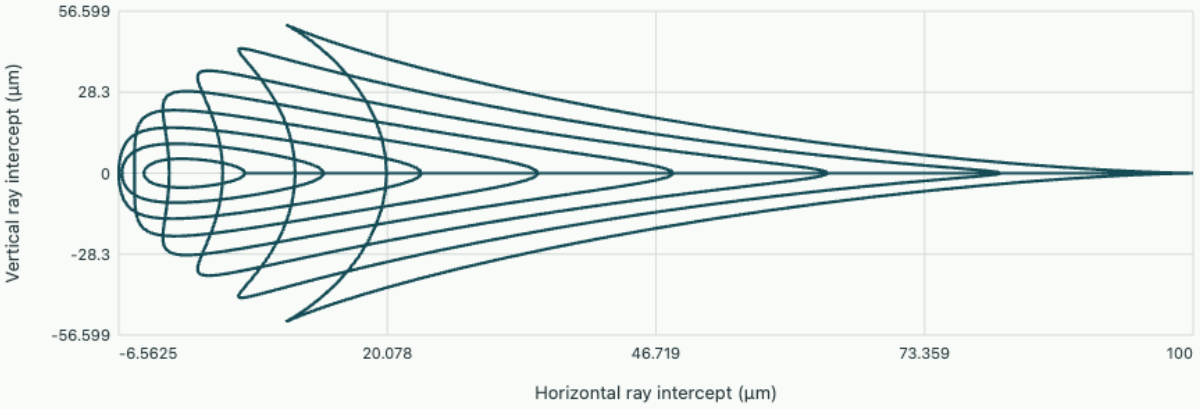
Try it · Aberration

Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Coma coefficient (µm)	1
Astigmatism coefficient (µm)	1
Defocus coefficient (µm)	0



Pupil rings mapped by wavefront gradient

W = C u(u²+v²) + A(u²−v²) + D(u²+v²), with f/a=20. D=±A gives the two astigmatic line foci when C=0. This geometrical intercept diagram omits diffraction.

Tutorial on recovering an aberration from data

1 · Diagnose an aperture sweep

▼ Work through the solution

A measured edge contribution to blur falls approximately eightfold when diameter is halved while the centre stays fixed. In geometrical optics that agrees with a cubic ray slope from primary spherical aberration. For coma the marginal physical-height dependence is quadratic at fixed field. At small apertures the Airy contribution eventually grows.

2 · Extract an astigmatic path error

▼ **Work through the solution**

A line-focus separation of 0.08 mm at $f = 100$ mm and $a = 5$ mm implies $|A|h^2 = \Delta z\, a^2/(4f^2) = 0.050$ µm. A single image at best focus would not distinguish the signs of the two principal wavefront curvatures.

---

3 · Test a tilted detector as an alternative explanation

▼ **Work through the solution**

Record a grid of pinhole images on both sides of the axis and scan focus. A detector tilt creates a systematic best-focus gradient across the field. Coma changes sign with reversed field, and astigmatism produces two line foci at the same field position. These measurements separate errors that can look similar in a single frame.

For further study see Hecht, Optics, 5th edition, chapters 5–6 on instruments, ray tracing and aberrations.

Optical fields, instruments and observations · 12. Why light can generate its second harmonic and then lose it again

A nonlinear polarisation can oscillate twice during one period of the applied field. Whether the generated wave grows through a crystal depends on how successive source layers arrive in phase. This spatial phase relation makes crystal length and dispersion as important as local nonlinearity.

Generate a new frequency by expanding the material response

In a transparent, nonmagnetic dielectric take one relevant polarisation and a specified tensor projection. For moderate fields write $P = \epsilon_0(\chi^{(1)}E + \chi^{(2)}E^2 + \cdots)$. For $E = E_0\cos(k_1 z - \omega t)$ the quadratic term contains both a steady component and an oscillatory one:

$$P^{(2)} = \tfrac{1}{2}\epsilon_0\chi^{(2)}E_0^2\{1 + \cos(2k_1 z - 2\omega t)\}.$$

In a centrosymmetric bulk medium the electric-dipole response must change sign when the electric field changes sign. This symmetry forces the bulk electric-dipole $\chi^{(2)}$ to vanish. Surfaces, suitable noncentrosymmetric crystals or other coupling mechanisms can supply a nonzero response.

Add the radiation from successive slices of crystal

At 2ω the wave equation has the driven form $(\partial_z^2 + k_2^2)E_{2\omega} = -\mu_0(2\omega)^2 P_{2\omega}^{(2)}$ in the frequency domain. The linear optical response is already included in k_2. Set $E_{2\omega} = A_2(z)e^{ik_2 z}$ and neglect A_2'' compared with $2ik_2A_2'$. For a plane, monochromatic undepleted pump the source envelope is constant. Apart from an immaterial constant phase, the derivative is proportional to $e^{i\Delta k z}$, with $\Delta k = 2k_1 - k_2$.

$$A_2(L) \propto A_1^2\int_0^L e^{i\Delta k z}\,dz = A_1^2 L\, e^{i\Delta k L/2}\,\mathrm{sinc}(\Delta k L/2).$$

The intensity thus scales as $I_1^2L^2\,\mathrm{sinc}^2(\Delta kL/2)$ under these assumptions. Define $\mathrm{sinc}(x) = \sin x/x$. The coherence length $\pi/|\Delta k|$ gives the first maximum for fixed nonzero mismatch. At twice that length equal contributions from opposite phase regions cancel. A crystal growing longer can therefore reduce the harmonic output.

Predict the coherence length and use the index tensor

For vacuum pump wavelength λ, $k_1 = 2\pi n_1/\lambda$ and $k_2 = 4\pi n_2/\lambda$. In this collinear geometry $L_c = \lambda/[4|n_1 - n_2|]$. Taking $\lambda = 1.0$ µm and an index difference of 0.050 gives $L_c = 5.0$ µm. Birefringent phase matching selects propagation direction and polarisation so the indices felt by the two fields can coincide. The tensor component driving the generated polarisation must also be nonzero for those directions.

Another route periodically reverses the sign of $\chi^{(2)}$ every coherence length. When the source would begin to cancel the previous contribution, reversal restores its sign. For a square-wave reversal of period Λ, the fundamental reciprocal vector $2\pi/\Lambda$ can cancel the mismatch. First-order quasi-phase matching needs $\Lambda = 2L_c$ and gives an ideal first-harmonic coupling magnitude $2/\pi$ times the corresponding uniformly phase-matched value. Finite bandwidth, walk-off and absorption further limit the interaction length.

Recall · Distinguish frequency conversion from spatial harmonics

The second temporal harmonic carries frequency 2ω. A linear diffraction grating can contain twice a spatial frequency while scattering light at the original optical frequency. A spectrometer separates these two mechanisms even if intensity images alone look similar.

Include the pump energy balance

At high conversion the assumed constant source term overestimates the pump. Scale amplitudes so $|a|^2$ and $|b|^2$ are the fractions of incident power at ω and 2ω. Let z now denote distance in nonlinear coupling lengths and δ the correspondingly scaled mismatch. A lossless two-wave model after a phase choice is

$$\frac{da}{dz} = -a^* b\, e^{-i\delta z},$$

$$\frac{db}{dz} = a^2 e^{i\delta z}.$$

Multiply the first equation by a^* and the second by b^*, add complex conjugates and observe the cancellation. Total power is conserved. At $\delta = 0$, $a(0) = 1$, $b(0) = 0$ give real solutions $a = \operatorname{sech} z$ and $b = \tanh z$. This supplies an independent check for integrations with nonzero mismatch, for which power can return to the fundamental.

OneSimulation experiment · Pump depletion and phase mismatch in a lossless crystal

Try phase mismatch, depletion and time-step error

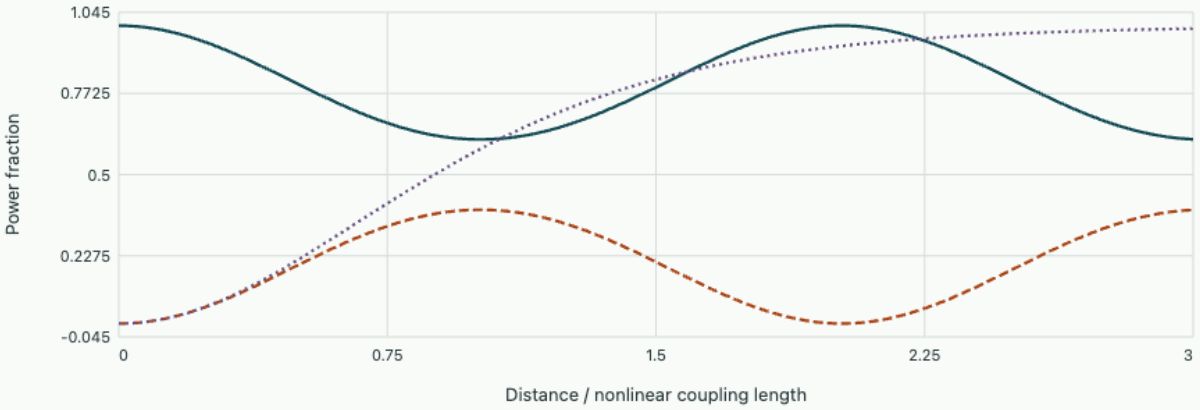
Change the phase mismatch and crystal length to locate a conversion maximum. At zero mismatch the computed second harmonic should follow tanh²(z/Lnl). At nonzero mismatch conversion can reverse; increasing length need not increase the output. Reduce the RK4 step while checking the separate power-balance residual.

The browser integrates the same four real envelope equations as the OneSimulation notebook below. Length and step are measured in nonlinear coupling lengths; mismatch is ΔkLnl. The lossless plane-wave model omits diffraction, absorption and temporal walk-off. Match `mismatch`, `step` and `count` in the notebook's `propagate` call to repeat a selected run.

Open this experiment online

Try the calculation

Phase mismatch Δk·Lnl	2
Crystal length / Lnl	3
RK4 step / Lnl	0.01



Fundamental power · RK4 | Second harmonic · RK4 | Exact second harmonic at Δk = 0

Exit harmonic fraction = 0.381279. Largest |Pω + P2ω − 1| = 3.68e-10. Actual step = 0.01. Normalised, lossless plane-wave model with pump depletion.

Reproduce with OneSimulation

Change the mismatch and step size. Inspect harmonic output and the conserved power separately. At exact phase matching compare the entire returned curve with the analytic hyperbolic-tangent solution.

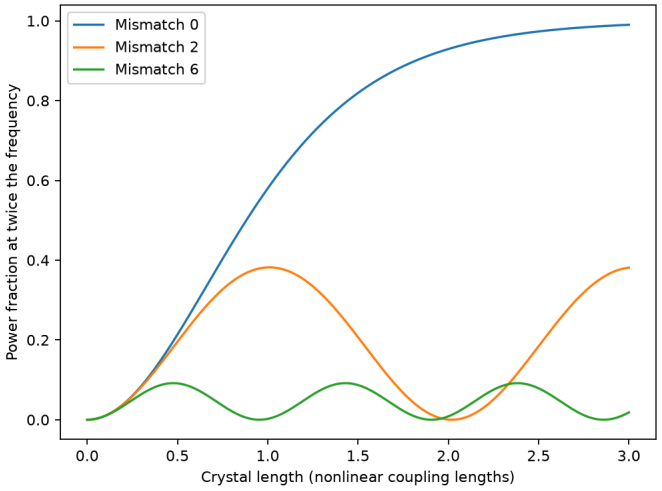
Run the downloaded notebook with OneSimulation, NumPy and Matplotlib installed. The output below was recorded with the displayed calculation.

```
"""Normalised fundamental and second-harmonic envelopes in a lossless crystal."""
import numpy as np
import matplotlib.pyplot as plt
from onesim.numerics.time import ClassicalRungeKutta4TimeIntegrator

def propagate(mismatch, step, count):
    def rates(state, z):
        a, b = complex(*state[:2]), complex(*state[2:])
        phase = np.exp(1j*mismatch*z)
        da = -a.conjugate()*b/phase
        db = a*a*phase
        return da.real, da.imag, db.real, db.imag
    method = ClassicalRungeKutta4TimeIntegrator()
    state = (1., 0., 0., 0.)
    trajectory = [state]
    for i in range(count):
        state = method.advance(state, rates, step, time=i*step)
        trajectory.append(state)
    ar, ai, br, bi = np.array(trajectory).T
    return ar**2+ai**2, br**2+bi**2

z = np.linspace(0, 3, 301)
for mismatch in (0., 2., 6.):
    pump, harmonic = propagate(mismatch, .01, 300)
```

```
    print(f"Mismatch {mismatch:g}: harmonic fraction at z=3 = {harmonic[-1]:.5f}")
    print(f"Max energy residual = {max(abs(pump+harmonic-1)):.2e}")
    plt.plot(z, harmonic, label=f"Mismatch {mismatch:g}")
    if mismatch == 0:
        assert np.max(abs(harmonic-np.tanh(z)**2)) < 1e-7
plt.xlabel("Crystal length (nonlinear coupling lengths)")
plt.ylabel("Power fraction at twice the frequency")
plt.legend()
plt.tight_layout()
plt.savefig("physics-harmonic-generation.png", dpi=160)
```

Computed output

```
Mismatch 0: harmonic fraction at z=3 = 0.99013
Max energy residual = 5.36e-11
Mismatch 2: harmonic fraction at z=3 = 0.38128
Max energy residual = 3.68e-10
Mismatch 6: harmonic fraction at z=3 = 0.01833
Max energy residual = 6.10e-10
```



Pump depletion and phase mismatch in a lossless crystal.

Download editable notebook · Download Python script

Tutorial on coherence length and power

1 · Infer a dispersion difference

▼ **Work through the solution**

A 1.064 µm pump has its first weak-conversion maximum at length 20 µm. The index difference is $\lambda/(4L_c) = 0.0133$. The next zero is near 40 µm if mismatch and pump intensity remain fixed.

2 · Separate weak conversion and depletion

▼ **Work through the solution**

Doubling length under perfect phase matching multiplies weak harmonic power by four. For a depleted pump at $z = 1$ and 2 the fractions are $\tanh^2(1) = 0.580$ and $\tanh^2(2) = 0.929$. Multiplying 0.580 by four would exceed the available power.

---

3 · Tune a grating with a measured mismatch

▼ **Work through the solution**

For $|\Delta k| = 6.28 \times 10^5$ m⁻¹ first-order quasi-phase matching needs period approximately 10.0 µm and sign reversal every 5.0 µm. A residual mismatch of 10^3 m⁻¹ in a 4 mm sample yields a weak-conversion intensity factor $\operatorname{sinc}^2(2) = 0.207$ relative to exact matching.

For further study see Hecht, Optics, 5th edition, chapter 13 on nonlinear optics and optical frequency conversion.

Optical fields, instruments and observations · 13. How a material changes the speed, attenuation and direction of light

A detector behind a sample loses light both when matter absorbs energy and when light is scattered outside its acceptance. A microscopic oscillator connects refraction to absorption. A radiating dipole then shows how a transparent particle can also reduce the measured direct beam.

Start with the electromagnetic wave equation in a medium

For a nonmagnetic, neutral, homogeneous dielectric with no free currents, Maxwell's curl equations give $\nabla \times \nabla \times \mathbf{E} = -\mu_0 \partial_t^2 (\epsilon_0 \mathbf{E} + \mathbf{P})$. A transverse plane wave in an isotropic material has $\mathbf{k} \cdot \mathbf{E} = 0$. Substitution for time dependence $e^{-i\omega t}$ gives $k^2 = (\omega^2/c^2)(1 + \chi)$. The local linear response is $\mathbf{P} = \epsilon_0 \chi(\omega) \mathbf{E}$. If $k = (\omega/c)(n + i\kappa)$, the real index controls phase accumulation while positive κ attenuates intensity as

$$I(z) = I(0) e^{-\alpha z},$$

$$\alpha = 2\omega\kappa/c.$$

Two samples of the same material with identical polished interfaces can separate bulk attenuation from a nearly constant surface loss. If increasing thickness by 5.0 mm reduces transmitted intensity by a factor of 0.80, the apparent bulk coefficient is $-\ln(0.80)/0.005 = 44.6$ m⁻¹. Multiple internal reflections, stray light and different surface treatments would spoil this cancellation.

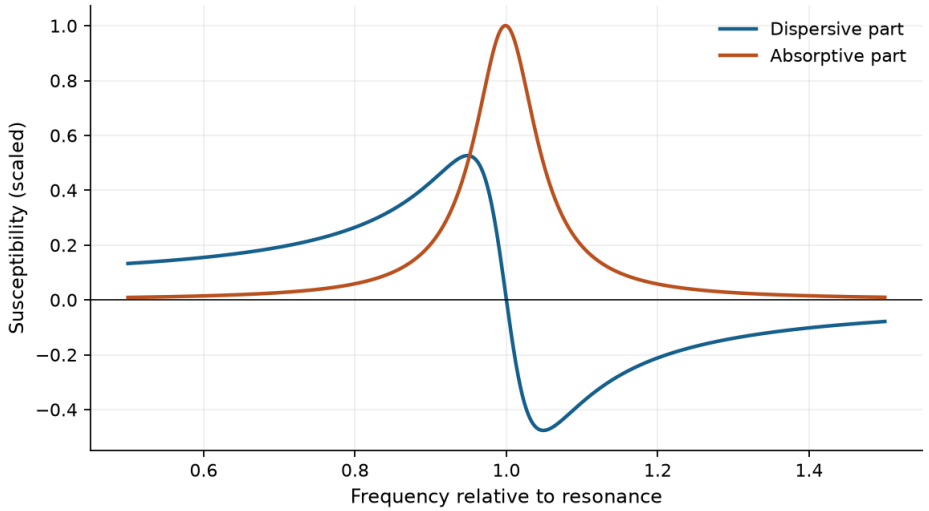
Derive the dispersive and absorptive parts of the response

Let a bound charge of magnitude e and mass m have a restoring frequency ω_0 and damping rate Γ. A charge $-e$ obeys $m(\ddot{x} + \Gamma \dot{x} + \omega_0^2 x) = -eE$. The polarisation is the number density N times the dipole $-ex$. Solving the driven equation gives

$$\chi(\omega) = \frac{Ne^2/(\epsilon_0 m)}{\omega_0^2 - \omega^2 - i\Gamma\omega}.$$

Below resonance the induced dipole tends to follow the applied field. Near resonance the imaginary part peaks and the real part evolves rapidly through a sign change. In a transparent region a narrow pulse has group index $n_g = n + \omega\, dn/d\omega$ from differentiating $k(\omega)$. Near strong absorption the pulse reshapes. A time-of-peak measurement alone may then give an unreliable travel time. The model links the two pieces of material response through a causal equation of motion.

> **Recall · Sign convention for a lossy material**
>
> With $e^{i(kz - \omega t)}$ a positive imaginary part of the wave number makes the amplitude decay. A time convention with the opposite sign reverses how the imaginary susceptibility is written while the positive measured attenuation stays the same.



A single passive resonance with damping rate one tenth of its undamped angular frequency. The dispersive phase response reverses sign near the frequency where absorption peaks.

Try it · Em dispersion

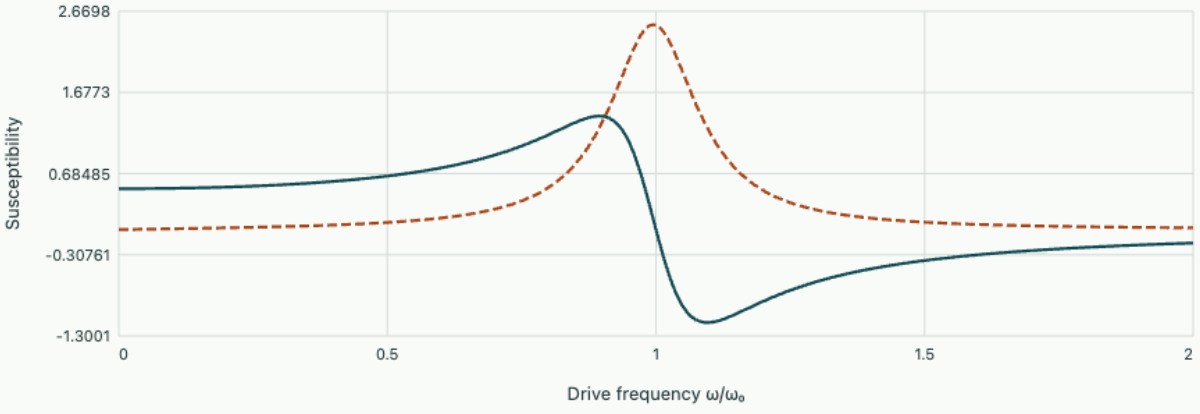
Explore the same physical relation with adjustable inputs. This independent parameter study has its own displayed reference settings; compare its assumptions and units before matching the static example.

Open this live calculation

Try the calculation

Damping γ/ω_0 **0.2**

Oscillator strength ω_p^2/ω_0^2 **0.5**



Re χ Im χ

At $\omega = \omega_0$ the susceptibility is purely imaginary with Im $\chi = 2.5$. At zero frequency the real susceptibility is 0.5.

Give a lossless particle a radiative channel

For a particle much smaller than the wavelength in vacuum use its electrostatic dipole amplitude $p = \epsilon_0 \alpha_V E$, with α_V in cubic metres. For a sphere of radius a and relative permittivity ϵ_r the electrostatic boundary conditions give $\alpha_V = 4\pi a^3 (\epsilon_r - 1)/(\epsilon_r + 2)$. Its transverse far field is proportional to $k^2 p \sin\theta/(4\pi\epsilon_0 r)$. Squaring and integrating $\sin^2\theta$ over solid angle gives

$$\sigma_{\mathrm{sca}} = \frac{k^4}{6\pi} |\alpha_V|^2 = \frac{8\pi}{3} k^4 a^6 \left| \frac{\epsilon_r - 1}{\epsilon_r + 2} \right|^2 .$$

Here the scattering angle is measured from the driven dipole direction. For a dilute independent suspension the collimated intensity approximately obeys $I = I_0 e^{-N\sigma_{\mathrm{ext}} z}$ with $\sigma_{\mathrm{ext}} = \sigma_{\mathrm{abs}} + \sigma_{\mathrm{sca}}$. The static polarizability gives the leading scattering power. For a quantitatively consistent forward amplitude include the small radiation-reaction term in the dipole response. If ka ceases to be small, higher multipoles contribute and the Rayleigh scaling ceases to hold.

For equally sized dilute scatterers with weakly dispersive contrast the cross-section ratio at 450 and 650 nm is $(650/450)^4 = 4.35$. Doubling radius at fixed particle number gives a factor of 64 when Rayleigh conditions hold at both radii. At fixed total particle volume, the number falls by eight on doubling the radius, so the attenuation instead increases by a factor of eight.

Tutorial on measuring optical constants

1 · Recover the imaginary index

▼ **Work through the solution**

At 600 nm the example attenuation 44.6 m⁻¹ corresponds to $\kappa = \alpha\lambda_0/(4\pi) = 2.13 \times 10^{-6}$ if attenuation is purely absorptive and scattering is negligible. Angularly resolved or integrating-sphere measurements test that last condition.

---

2 · Separate phase and pulse travel times

▼ **Work through the solution**

A transparent 2 cm slab has $n = 1.50$ and $n_g = 1.56$. The phase and group times are 100 ps and 104 ps respectively using $c = 3.0 \times 10^8$ m/s. A narrowband phase measurement alone does not determine the slope of the index.

---

3 · Test whether attenuation is molecular absorption

▼ **Work through the solution**

Measure transmitted collimated power and total scattered power with matched geometries. A wavelength-dependent loss proportional to λ^{-4} and increasing strongly with particle size supports Rayleigh scattering. Material absorption may also vary with wavelength, so independently test size and angular dependence.

For further study see Hecht, Optics, 5th edition, chapters 2–4 on electromagnetic waves and the propagation of light.

A wave in space, matter and a guide

Rebuild a propagating mode from the field equations. Track vector polarisation and flux through an interface, then separate material from geometric dispersion.

1. From a local field balance to travelling waves and standing modes
2. What an analyser measures about polarisation and coherence
3. Oblique interfaces and films whose reflections cancel
4. Causal material response and the propagation of a finite pulse
5. A guided mode, a cavity line and the energy they carry
6. When a periodic dielectric has no propagating Bloch wave
7. Why rotating a layered dielectric changes a capacitance measurement
8. How a magnetic field rotates the polarisation of transmitted light

From a local field balance to travelling waves and standing modes

Waves, radiation and relativity

A voltage probe moving along a line finds regularly spaced zeros. A second probe at the same point measures a magnetic field even when the electric reading vanishes. Derive both readings from Maxwell's equations, calculate the energy stored between the zeros, and decide what additional measurement would reveal the direction in which energy travels.

Specify a medium, a region and a sign convention

Take a homogeneous, isotropic, transparent medium at rest. In the frequency range of interest its permittivity ϵ and permeability μ are positive, real constants. Neglect free charge and current in the region where we will propagate a field. Sources can still exist outside the region and set its incident wave. The electric and magnetic fields are functions of position and time. Maxwell's equations impose two divergence conditions in addition to the two evolution equations. In particular, $\nabla \cdot \mathbf{E} = 0$ here because ϵ is uniform. At a material interface the normal condition concerns $\mathbf{D}$, so one must reconsider this simplification before using the same wave equation across a boundary.

Consider first a plane field depending only on z and t, with electric polarization along x and magnetic field along y. Using a right-handed coordinate system, the y component of the curl of the electric field is $+\partial_z E_x$. The x component of the curl of the magnetic field is $-\partial_z H_y$. Faraday's law and Ampère–Maxwell's law then give

$$\partial_z E_x = -\mu\, \partial_t H_y,$$
$$\partial_z H_y = -\epsilon\, \partial_t E_x.$$

Imagine a localized positive electric bump travelling to the right. At a point just ahead of its centre the field grows in time. The second equation therefore requires a negative magnetic-field slope there. This checks the relative signs before making a harmonic assumption. With this convention a positive $E_x H_y$ describes energy flow towards increasing z.

Find what travels and what remains constrained

Differentiate the first curl equation with respect to z and substitute the second. Constant ϵ, μ allow derivatives to commute with the coefficients. The electric field obeys $E_{zz} = \epsilon\mu E_{tt}$. The same procedure gives the magnetic wave equation. More generally, taking the curl of Faraday's law produces

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu\epsilon \mathbf{E}_{tt}.$$

The divergence term vanishes under the stated assumptions. A formal solution of three unrelated scalar wave equations can still violate the divergence or original curl conditions. These conditions connect the measurable polarizations and reduce the independent initial data.

$$v = (\epsilon\mu)^{-1/2},$$

$$Z = \sqrt{\mu/\epsilon},$$

$$\partial_t (E_x + Z H_y) + v \partial_z (E_x + Z H_y) = 0.$$

The corresponding combination with a minus sign travels to the left. To see the first equation, expand it and replace $E_t = -H_z/\epsilon$, $H_t = -E_z/\mu$. The coefficients cancel because $vZ = 1/\epsilon$ and $Z/\mu = v$. Hence there exist arbitrary waveform functions F and G such that

$$E_x(z,t) = F(z - vt) + G(z + vt),$$

$$H_y(z,t) = [F(z - vt) - G(z + vt)]/Z.$$

The medium determines two different quantities. Delay measurements determine the product $\epsilon\mu$; measuring electric-to-magnetic amplitude in a single travelling wave determines their ratio. Impedance has units volts per ampere. It relates fields in volts per metre and amperes per metre at one location. A port voltage divided by current additionally depends on how a particular line geometry integrates those fields over a cross-section.

If only the initial electric field $E_x(z,0)$ is known, its decomposition into left- and right-moving components is ambiguous. A simultaneous magnetic measurement supplies the missing combination. For example, suppose a local measurement gives $E_x = 3.0$ V/m and $H_y = 4.0$ mA/m in a medium with $Z = 250\ \Omega$. There $ZH_y = 1.0$ V/m, so the right-moving field equals 2.0 V/m and the left-moving field equals 1.0 V/m. These numbers describe local instantaneous amplitudes. Separating time-averaged directional powers requires squared waveform records or a harmonic representation.

Recover the energy current and average it correctly

Multiply the first curl equation by H, the second by E, and add. The spatial terms form $\partial_z(EH)$, while the time terms form minus the derivative of $\epsilon E^2/2 + \mu H^2/2$. Thus in the absence of work on current the field obeys a local conservation law. The two travelling pieces give an especially useful representation of both sides:

$$u = \frac{\epsilon E^2 + \mu H^2}{2} = \epsilon (F^2 + G^2),$$

$$S_z = EH = \frac{F^2 - G^2}{Z},$$

$$u_t + \partial_z S_z = 0.$$

The energy densities add while the currents subtract. A field carrying equal left- and right-going powers can have substantial stored energy and zero mean transport. A cancellation in E alone says little about stored energy because it often accompanies a maximum in H. For the two local amplitudes of the preceding example the instantaneous net flux is

$$(4 - 1)/250 = 0.012\,\mathrm{W/m^2}.$$

Multiplying the original readings gives $3.0 \times 0.004 = 0.012$ W/m², an independent check.

For harmonic calculations use real physical fields equal to $\mathrm{Re}[\widetilde{\mathbf{E}}e^{-i\omega t}]$, so the phasors denote peak amplitudes. A product of real sinusoids averages to one half the real part of one phasor times the conjugate of the other. Therefore $\langle \mathbf{S} \rangle = \mathrm{Re}(\widetilde{\mathbf{E}} \times \widetilde{\mathbf{H}}^{*})/2$. A single forward sinusoid with electric peak 6.0 V/m in vacuum carries $36/(2Z_0) = 0.0478$ W/m², using $Z_0 = 376.73$ Ω. If an instrument already reports RMS electric amplitude, the factor of one half is absorbed in that reading. Mistaking an RMS calibration for a peak calibration changes an inferred power by a factor of two.

Enforce a conductor condition on the two waves

Place a perfect conducting plane at z=0 and illuminate it from negative z. The tangential electric field must vanish at the surface at all times. For peak incident phasor $E_0 e^{ikz}$, with $k = \omega/v$, the reflected field must be $-E_0 e^{-ikz}$; the accompanying magnetic fields add at the surface. Choosing real E_0, the total phasors are

$$\widetilde{E}_x = 2iE_0 \sin(kz),$$

$$\widetilde{H}_y = \frac{2E_0}{Z} \cos(kz).$$

Electric nodes are separated by $\pi/k = \lambda/2$. The magnetic antinodes sit at those same locations. At a fixed interior point the two fields are a quarter cycle out of phase. Their mean product is zero but their instantaneous product alternates sign. Examine a quarter wavelength of the standing wave: energy crossing an internal plane feeds local changes between electric and magnetic storage, while the conductor absorbs no net power in the idealization.

In a nonmagnetic dielectric with $\epsilon_r = 4$, the phase speed is $c/2$ and $Z = Z_0/2$. At 500 MHz the wavelength is 0.2998 m and successive electric zeros lie about 150 mm apart. A 4 V/m incident peak gives an 8 V/m electric antinode and magnetic amplitude $8/(188.4) = 0.0425$ A/m at the conductor. A tiny electric probe positioned at the conductor would read essentially zero. A loop sensing magnetic flux can respond strongly there. The response of a real loop also depends on its orientation, area and the frequency-dependent conversion of flux to voltage.

Build a complete family between two walls

Put conductors at $z = 0, L$. Tangential E now vanishes at both endpoints. Try a separated solution $E = X(z)T(t)$. Dividing the wave equation by XT gives $X''/X = T''/(v^2 T) = -k^2$ for an oscillatory solution. The first condition eliminates the cosine term in X. The second enforces $k_n = n\pi/L$, $n = 1, 2, \ldots$. Frequencies are

$$\omega_n = v k_n.$$

There is no nonzero constant tangential electric mode consistent with two zero boundary values. A static magnetic bias can be independently specified through initial flux; it does not belong to this oscillatory electric-mode series.

$$E(z,t) = \sum_{n=1}^{\infty} \sin(k_n z)[A_n \cos(\omega_n t) + B_n \sin(\omega_n t)].$$

To find coefficients, multiply the initial field by $\sin(k_m z)$ and integrate over the interval. Product-to-sum identities give zero for $n \neq m$, and $L/2$ for $n = m$. Thus $A_m = (2/L)\int_0^L E(z,0)\sin(k_m z)\,dz$. Applying the same projection to the initial time derivative gives

$$B_m = (2/(\omega_m L))\int_0^L E_t(z,0)\sin(k_m z)\,dz.$$

The curl equation supplies E_t from the spatial derivative of the initial H field. These formulas convert measured initial data into mode coordinates. A point reading alone supplies one linear combination of all coordinates.

One can also derive orthogonality without already knowing the sines. Multiply the eigen-equations for X_n and X_m by the opposite functions, subtract, and integrate. The boundary term $[X_n' X_m - X_m' X_n]_0^L$ is zero, so

$$(k_n^2 - k_m^2)\int X_n X_m dz = 0.$$

Other homogeneous boundary conditions can preserve this cancellation. Spatially varying material parameters generally change the weight in an orthogonality relation. Knowing the inner product matters when estimating an amplitude from measured fields in unlike dielectrics.

Predict which modes a localized preparation actually excites

Imagine an idealized initial tangential electric field with a symmetric top-hat amplitude E_a on $L/3 < z < 2L/3$, zero elsewhere, and zero initial H. Then $B_n = 0$ and integration gives

$$A_n = \frac{2E_a}{n\pi}\left[\cos(n\pi/3) - \cos(2n\pi/3)\right].$$

Symmetry around the middle eliminates even n because their sine eigenfunctions change sign about that point. For $n = 1, 3, 5$ the coefficients are $2E_a/\pi,\ -4E_a/(3\pi),\ 2E_a/(5\pi)$, respectively. Negative A_3 is a relative phase; its contribution to stored energy is positive. Orthogonality gives initial total electric energy per transverse area $\epsilon L\sum A_n^2/4 = \epsilon L E_a^2/6$. The first mode carries $6/\pi^2 \approx 60.8\%$ of that total, while the third carries $8/(3\pi^2) \approx 27.0\%$. A three-mode truncation reproduces a large share of the energy and still rounds the sharp initial edges. Measured peak error and integrated energy error answer different questions about that approximation.

Take $L = 0.60$ m filled with the dielectric above. The first frequency is about 124.9 MHz. A receiver at midspan sees modes 1 and 3 with opposite spatial signs. A second receiver at $L/3$ has a node of mode 3 and will miss it. Detecting only one large peak after changing receiver position is therefore insufficient evidence that the source was single mode. Predict the source overlap, propagation loss and receiving overlap separately before interpreting a spectrum.

Use an open wave measurement as an independent power check

The linked calculation treats a propagating field and derives an energy-flux observable. Read its amplitude convention and the material properties beside the inputs. Predict which quantity changes quadratically when

you vary only field amplitude. If a lab measures a total power, convert the calculated flux using the illuminated area and the angle between the flux vector and the detector normal. For a plane wave in a uniform medium, changing area at fixed field changes total intercepted power without changing any local field. In a narrow cavity beam the local field itself varies across the aperture, so the area product must be replaced by a surface integral.

Compare fields, energy density and transported power in the chapter experiment

An investigation of wave speed on a space–time mesh

A pulse simulation offers a second way to test the continuum mode calculation. Instead of assuming a single-frequency source, initialise a packet in a uniform one-dimensional region and record the times when it crosses two stations. First predict how the numerical method can change this transit measurement. A code that respects the differential equations as its mesh tends to zero can have a noticeably incorrect propagation speed at a finite resolution.

Work with an electric component E_y on integer mesh points and a magnetic component H_z on the intervening half points. Store the magnetic component half a time step later than the electric component. Faraday's law compares the electric field on the two ends of one spatial interval. Ampère's law compares the two magnetic components bounding the electric location. With $d = \Delta x$ and $s = \Delta t$, their integral balances become

$$H_{j+1/2}^{n+1/2} = H_{j+1/2}^{n-1/2} - \frac{s}{\mu d}\left(E_{j+1}^{n} - E_{j}^{n}\right),$$

$$E_{j}^{n+1} = E_{j}^{n} - \frac{s}{\epsilon d}\left(H_{j+1/2}^{n+1/2} - H_{j-1/2}^{n+1/2}\right).$$

For a forward wave the indicated components have the same sign. A change in either sign convention requires changing both updates consistently. Placing the variables on staggered meshes also matters when sampling power. A product of an electric field and the nearest stored magnetic number is not automatically their product at a common event. Interpolate both in time and space for a quantitative flux comparison, or use a discrete energy identity appropriate to the scheme.

Eliminate the magnetic values by taking a second time difference of the electric update. The resulting equation equates the centred second time difference to C^2 times the centred second space difference, where $C = us/d$ and $u = 1/\sqrt{\mu\epsilon}$. For the discrete mode $E_j^n = E_0 e^{i(kjd - \omega n s)}$, the difference of the forward and backward phases becomes a sine. Take the branch connected to the continuum wave at small positive k.

$$\sin(\omega s/2) = C \sin(kd/2),$$

$$\frac{v_{\mathrm{phase}}}{u} = \frac{\arcsin\left[C \sin(kd/2)\right]}{C kd/2},$$

$$\frac{v_{\mathrm{group}}}{u} = \frac{\cos(kd/2)}{\sqrt{1 - C^2 \sin^2(kd/2)}}.$$

The group speed follows by differentiating the frequency with respect to wavenumber before choosing a particular packet. The right-hand side of the sine equation must remain within the range of real sines for every mesh mode. This gives $C \le 1$ in this one-dimensional uniform problem. If a short-wavelength mode

violates the condition, a growing discrete solution accompanies the decaying one. A seemingly smooth start may seed very little of either, but roundoff or a sharp boundary can excite the growing solution. A stable choice still permits frequency-dependent speed.

For ten samples per wavelength and $C = 0.50$, the phase and group speed ratios are approximately 0.988 and 0.963. Over twenty wavelengths a phase crest is about a quarter of a cycle late, and the envelope peak has an even larger relative lag. With twenty samples per wavelength the errors shrink roughly fourfold. Reducing only the time step at fixed space resolution need not improve the result: the spatial difference retains its own dispersion error, and the temporal error can partially compensate it. At $C = 1$ in this idealised one-dimensional uniform case the phase relation on the resolved branch is exact. A slanted multidimensional wave, a discontinuous medium or a different mesh does not inherit that special cancellation.

A worked time-of-flight measurement with uncertainty

To construct an initial forward pulse, give the electric field a broad Gaussian envelope multiplying its carrier and assign the associated magnetic field $H = E/Z$ at its proper staggered locations and times. Setting the magnetic field to zero instead launches a superposition of left- and right-moving packets of comparable amplitude. A reflecting wall can return one packet through the sensor after the other. Before fitting a speed, identify a time window that excludes every return from the domain boundaries. Check this by increasing the distance to the nearest wall without changing either station.

Suppose the expected continuum transit between stations is 100.0 ns, each threshold crossing is timed to 0.3 ns and the station spacing is known to 0.2 percent. Independent timing errors contribute $\sqrt{2}(0.3)/100 = 0.00424$ relative uncertainty to the difference. Combined in quadrature with the length uncertainty this gives 0.47 percent. The roughly 3.7 percent group-speed bias at ten points per wavelength should be distinguishable, while the smaller phase error calls for careful identification of the same crest at both stations. A common clock offset cancels from the difference, whereas a common timebase scale error does not.

A threshold on the leading edge tracks a changing fraction of peak amplitude if the model disperses the pulse. Use a complex narrow-band signal and fit the slope of the unwrapped phase difference against frequency as a second estimate of group delay. For a single frequency only the phase modulo 2π is observed. An apparent speed that jumps as the station separation changes by a small fraction of a wavelength can be an unwrapping error, even though the electric time series looks smooth. Begin with stations close enough to choose a branch, then follow the branch as their separation increases.

An additional stringent check uses two packets with different carrier wavelengths but equal initial envelope width in space. If the narrow-band condition holds for both, their group delays should match the two predictions of the mesh dispersion relation. Measured delays that approach the same continuum value with refinement while their difference obeys the predicted wavelength dependence support the diagnosis of numerical dispersion. A nonvanishing delay difference in the refined limit calls for revisiting the assumed physical material law, the timing observable or the boundaries. A single apparent agreement at one wavelength cannot distinguish these possibilities.

Separate flux from standing-wave energy

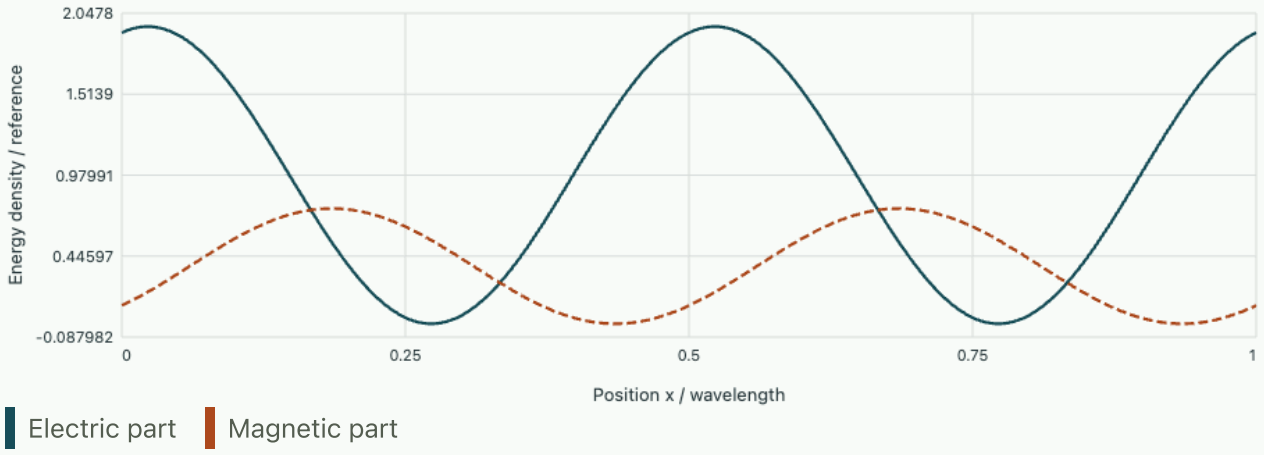
Change the balance of waves travelling in opposite directions. Identify a point where the instantaneous field is large while net power transfer is small.

Open this experiment online

Try the calculation

Reflected electric amplitude / incident **0.6**

Phase of incident wave (degrees) **30**



The cycle-averaged net power is 64.0% of the incident power. Reversing the sign of reflection swaps the pattern at the boundary.

Worked inverse problem · estimate the energy of an invisible mode

A conducting one-dimensional cavity contains its first and third electric standing modes. At a common electric maximum in time let $E(x) = a\sin(\pi x/L) + b\sin(3\pi x/L)$. A small signed probe at $L/2$ reads 2.0 V/m, and a probe at $L/6$ reads 4.0 V/m. Determine the two amplitudes and the fraction of energy in the third mode. Compare this result with what could be concluded from two intensity probes alone.

▼ Solve the signed and unsigned versions

The centre gives $a - b = 2.0$. The second station gives $a/2 + b = 4.0$. Summing the equations yields $a = 4.0$ V/m and hence $b = 2.0$ V/m. At this instant the compatible free modes have zero magnetic amplitude. Spatial orthogonality makes integrated electric energy proportional to $a^2 + b^2$, even though there are local interference terms. The third mode therefore carries 20 percent of total energy. This fraction stays fixed during free evolution in an ideal lossless cavity. The instantaneous electric share of each mode alone oscillates in time.

Two intensity probes would give only magnitudes of the two readings. Choose the centre reading positive by a global sign convention. The second could then be -4.0 instead. Solving $a - b = 2.0$, $a/2 + b = -4.0$ gives $a = -4/3$ and $b = -10/3$ V/m, for a third-mode fraction $100/116 = 0.862$. Both field preparations

produce the same two intensities. A third carefully chosen position, a phase-sensitive receiver or a frequency-resolved trace distinguishes them. A probe at a location that simply duplicates either existing linear combination adds precision but cannot supply the missing sign.

An eigenmode decomposition is therefore both a prediction tool and an experimental design tool. Before choosing sensors, form the matrix of sampled mode shapes and check whether its columns remain distinguishable within noise. If two stations approach a shared node, the inferred coefficients can grow unstable while the predicted signal at those stations remains an excellent fit. A held-out measurement away from the common node exposes the ambiguity.

Why a sharp preparation resists uniform modal convergence

A rectangular initial displacement has a jump. Any finite sum of the smooth sine modes is continuous and therefore cannot converge uniformly to a function with different one-sided limits. Near the edge a partial sum oscillates. Increasing the cutoff makes the oscillatory region narrower, but its largest overshoot does not tend to zero relative to the jump. Pointwise comparisons a fixed positive distance from the jump and spatially integrated energy tell a different story from the maximum error.

This is physically relevant to preparing an electromagnetic cavity. No source creates a perfectly discontinuous field with unbounded bandwidth. Replace the top hat by its convolution with a smooth narrow preparation profile. The modal coefficients then acquire the Fourier transform of that profile as a high-frequency envelope. Doubling the number of retained modes beyond the preparation bandwidth should change both the integrated energy and the measured field little. A finite-width detector performs its own additional averaging, even when the preparation is sharp.

An informative validation uses two limits in a stated order. Fix the physical source and detector widths while increasing the number of modes until the detector prediction converges. Then vary the physical widths to investigate the sharp-edge limit. Shrinking the model widths each time the numerical resolution increases changes the problem being solved and may keep the apparent peak error from decreasing.

Tutorial · reconstruct a partial reflection from three observations

A coherent line scan in a nonmagnetic medium of known $Z = 200\ \Omega$ finds successive electric maxima 80 mm apart at 1.0 GHz. Maximum and minimum peak electric amplitudes are 5.0 and 1.0 V/m. Determine phase speed, incident and reflected peak fields and mean net flux. Explain exactly what fixes the phase of the reflection coefficient and what remains ambiguous in these observations.

▼ **Derive the two directional amplitudes**

For nonnegative amplitude magnitudes a, b, the maxima and minima of the sum of two opposite travelling waves are $a + b$ and $|a - b|$. If the source is known to lie to the left and net transport is rightward, $a > b$, so $a = 3.0,\ b = 2.0$ V/m. Wavelength is twice the 80 mm maximum spacing and $v = f\lambda = 1.60 \times 10^{8}$ m/s. The reflection magnitude is $2/3$, and the average net flux is

$$(9 - 4)/(2 \cdot 200) = 0.0125\,\mathrm{W/m^2}.$$

A measurement of node spacing constrains the relative phase change over distance. Determining reflection phase at a named load additionally requires the position of a maximum or minimum relative to that reference plane. Without an independent source direction, swapping a and b produces the same electric envelope but reverses the net flux.

Tutorial · test whether a receiver can reconstruct two stored modes

At an instant when both cosine time factors are one, only the first and third electric modes contribute. At midspan the field is 3.0 V/m; at one third span it is 1.732 V/m. Determine A_1, A_3 and the stored electric energy per unit area for $L = 0.60$ m,

$$\epsilon = 4\epsilon_0.$$

Would a second sensor at two thirds span provide independent information if it replaced the midspan sensor?

▼ **Solve the observation system and identify its rank**

At one third span $\sin(\pi/3) = \sqrt{3}/2$ and $\sin\pi = 0$, so $A_1 \simeq 2.00$ V/m. At midspan the combination is $A_1 - A_3 = 3.0$, hence $A_3 = -1.0$ V/m. The energy per area is $\epsilon L(4+1)/4 = 2.66 \times 10^{-11}$ J/m². At two thirds both sinusoids have the same values as at one third for these odd modes: the first is $\sqrt{3}/2$ and the third is zero. Those two rows would be identical. More digits in either measurement cannot resolve A_3 from that pair. An offset from a node helps, but near a node a small measured voltage also amplifies calibration noise in the reconstruction.

David Staelin, MIT 6.013 · electromagnetic waves, power and Maxwell boundary conditions

What an analyser measures about polarisation and coherence

Waves, radiation and relativity

Two red beams produce the same reading through every rotating linear polarizer. One might be circularly polarized and the other an incoherent mixture. Build the measurements that distinguish them, derive the visibility of their interference fringes, and decide which information is lost on a slow detector.

Begin with the measured quantity and the averaging time

Fix a beam travelling in positive z and choose perpendicular laboratory axes x and y in the detection plane. Represent a narrowband field by $\mathrm{Re}[\mathbf{e}(t)e^{-i\omega_0 t}]$. The complex envelope $\mathbf{e} = (e_x, e_y)$ varies slowly compared with an optical cycle. A photodiode collecting many cycles measures a time average of energy flux, possibly over many independent changes of the envelope. In a transparent isotropic medium a convenient intensity convention is $I = \langle |e_x|^2 + |e_y|^2 \rangle / (2Z)$. All comparisons in this unit assume equal transmissive losses and detector responsivity unless a calibration factor is stated.

A deterministic monochromatic field has a Jones vector. Its overall complex phase disappears from all single-beam intensities, while a relative phase between components can change the polarization ellipse. For example, use the normalized vectors $(1,0),\ (1,1)/\sqrt{2},\ (1,i)/\sqrt{2}$. The first is x-linear, the second linear at +45°, and the last circular. At a fixed point choose the time origin so the physical electric vector for the last case begins on the positive x axis and then moves towards positive y as t increases. This explicit convention avoids ambiguity in the words left and right circular, which depend on where an observer stands.

A fast field sampler would reconstruct both real components directly. Optical intensity measurements normally come with a response time much longer than a cycle. The data naturally contain squares and correlations of fields. A random global phase changing between shots destroys the inferred absolute phase but need not depolarize a single beam. If $e_y = ie_x$ in every shot, their relative phase stays fixed however rapidly the common phase wanders. A detector can measure that relationship with suitable polarizing optics.

Derive a quadratic form for any analyser setting

An ideal analyser transmitting normalized Jones vector $\mathbf{a}$ first projects the incident field to complex amplitude $\mathbf{a}^\dagger \mathbf{e}$. Squaring and averaging gives a quadratic form in $\mathbf{a}$. Define the coherency matrix in intensity units by $J_{ij} = \langle e_i e_j^* \rangle / (2Z)$. It is Hermitian, its diagonal entries are nonnegative, and its trace is total intensity. For any vector b,

$$\mathbf{b}^\dagger J \mathbf{b} = \langle |\mathbf{b}^\dagger \mathbf{e}|^2 \rangle / (2Z) \geq 0.$$

This positivity will later provide an immediate check on measured data.

$$I_{\mathbf{a}} = \mathbf{a}^\dagger J \mathbf{a},$$

$$S_0 = J_{xx} + J_{yy},$$

$$S_1 = J_{xx} - J_{yy},$$

$$S_2 = 2\Re J_{xy},$$

$$S_3 = -2\Im J_{xy}.$$

With this sign convention the circular vector $(1, i)/\sqrt{2}$ has positive S_3. Invert the four definitions to reconstruct $J = \frac{1}{2}\begin{pmatrix} S_0+S_1 & S_2-iS_3 \\ S_2+iS_3 & S_0-S_1 \end{pmatrix}$. A linear polarizer with real $\mathbf{a} = (\cos\alpha, \sin\alpha)$ gives

$$I(\alpha) = \tfrac{1}{2}\left[S_0 + S_1\cos(2\alpha) + S_2\sin(2\alpha)\right].$$

This follows by expanding the two squared direction cosines and the cross term. The imaginary part of the correlation drops out for every real analyser. Scanning a linear analyser with arbitrarily fine angular steps cannot supply that missing observable. Measuring the two circular analyser states supplies $I_+ - I_- = S_3$; their sum, like any pair of orthogonal ideal outputs, equals S_0. In a dual-beam measurement it is useful to compare that sum with $I_x + I_y$ as a consistency check on losses.

Build a circular analyser from a retarder

An ideal lossless plate with fast and slow axes x and y introduces a relative phase δ. Factoring out an unobservable common phase, its Jones matrix in those axes is $D = \mathrm{diag}(1, e^{i\delta})$ for the chosen sign of retardance. If the axes are rotated by an angle α, express the incoming vector in the rotated basis, apply D, then transform back. With $R(\alpha)$ the ordinary real two-dimensional rotation, the laboratory matrix is $R(\alpha)DR(-\alpha)$. Reversing the sign of δ interchanges retardation conventions, so calibrate a handedness label using a known input.

For a quarter-wave plate with $\delta = \pi/2$ followed by a +45° linear polarizer, the transmitted amplitude is $(e_x + ie_y)/\sqrt{2}$. The analyzer equivalent to this sequence is the vector $(1, -i)/\sqrt{2}$. Hence it reads I_- in the preceding convention. Interchange the downstream polarizer between +45° and −45° to realize the two orthogonal circular measurements. For a half-wave plate, a linear input at angle ψ relative to its fast axis emerges linearly polarized at angle $-\psi$ relative to that axis, modulo π. Rotating a half-wave plate by 1° rotates the linear output by 2° when the incident direction is held fixed.

Suppose a plate designed at 600 nm has physical birefringence $\Delta n = 0.010$. If the relative phase magnitude is $2\pi\Delta n\, d/\lambda_0$, a first-order quarter wave needs $d = \lambda_0/(4\Delta n) = 15$ µm. At 650 nm, holding Δn fixed, the same plate delivers $0.923(\pi/2)$, about 83.1°. With a +45° linear input, equal component amplitudes are preserved but a residual $S_2/S_0 = \cos\delta \approx 0.121$ remains. A linear analyser scan that varies slightly with angle can therefore reveal retardance error even when the output looks nearly circular.

Reconstruct the degree and axes of polarization

Hermitian J has real nonnegative eigenvalues. Calculating the trace and determinant yields eigenvalues $\lambda_\pm = (S_0 \pm s)/2$, where $s = (S_1^2 + S_2^2 + S_3^2)^{1/2}$. The degree of polarization is $P = s/S_0$ for positive intensity. Positivity requires $P \le 1$. A single deterministic Jones vector gives a rank-one outer product and P=1. A completely unpolarized field gives $J = (S_0/2)I$ and P=0. An isotropic unpolarized contribution adds equally to both eigenvalues. The excess of the larger eigenvalue over the smaller gives the uniquely polarized contribution in this decomposition.

$$2\psi = \mathrm{atan}\,2(S_2, S_1),$$

$$\sin(2\chi) = S_3/s.$$

Here ψ is the ellipse orientation modulo π; χ is an ellipticity angle between $-\pi/4$ and $+\pi/4$. The ratio of minor to major field axes is $|\tan\chi|$. At circular polarization the orientation is undefined, and near this limit even small noise on S_1, S_2 can rotate a fitted orientation substantially. At P=0 neither angle describes a measurable polarized component.

Consider readings of 7 and 3 mW for horizontal/vertical, 6 and 4 mW for ±45°, and 8 and 2 mW for the positive/negative circular analysers. Every pair sums to 10 mW. The Stokes vector is $(10, 4, 2, 6)$ mW, $s = \sqrt{56} = 7.48$ mW, $P = 0.748$. The polarized-component angle is 13.28°, and $\chi = \frac{1}{2}\arcsin(6/\sqrt{56}) = 26.6^\circ$. A linear analyser alone would have suggested a modulation depth $\sqrt{20}/10 = 0.447$. The recovered eigenvalues, about 8.74 and 1.26 mW, are the largest and smallest transmitted powers over all ideal elliptical analysers. The largest intensity available to a linear analyser is only $(10 + \sqrt{20})/2 = 7.24$ mW.

Derive the two-arm visibility and its ambiguities

Now split a quasimonochromatic scalar field between two interferometer arms with delay τ and relative phase φ. At one output combine complex fields $e_1(t)$ and $e_2(t-\tau)e^{i\varphi}$, absorbing beam-splitter factors into arm intensities I_1, I_2. Expanding the squared sum gives

$$I(\varphi) = I_1 + I_2 + 2\sqrt{I_1 I_2}\,\Re[\gamma_{12}(\tau)e^{i\varphi}],$$

$$\mathcal{V} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}|\gamma_{12}(\tau)|.$$

Define $\gamma_{12} = \langle e_1^*(t)e_2(t-\tau)\rangle / \sqrt{\langle|e_1|^2\rangle\langle|e_2|^2\rangle}$, provided both intensities are nonzero. Cauchy–Schwarz enforces $|\gamma| \le 1$. For vector beams with constant normalized polarizations $\mathbf{a}_1, \mathbf{a}_2$, an additional overlap $\mathbf{a}_1^\dagger \mathbf{a}_2$ appears. Orthogonal polarizations can erase fringes even when relative phase is perfectly stable. A common output analyser projects both beams into the same polarization and can restore modulation at a reduced transmitted flux.

An interferometer gives $I_1 = 4.0$ mW, $I_2 = 1.0$ mW and observed visibility 0.60 with matched polarization. The intensity imbalance alone limits visibility to $2\sqrt{4}/5 = 0.80$. Thus $|\gamma| = 0.75$. If the polarizations are instead linear at 30° to each other, the polarization overlap is 0.866 and the inferred scalar degree of coherence is

$$0.60/(0.80 \cdot 0.866) = 0.866.$$

Reporting visibility as a direct measure of source coherence requires both arm powers and polarization overlap. Detector backgrounds additionally increase the denominator without increasing fringe modulation.

Translate a spectrum into a coherence envelope

For a stationary scalar field, different frequency components of a long acquisition add in average intensity. A relative delay changes the phase of each by its own frequency times τ. Consequently the mutual coherence of two copies of the same field is proportional to the Fourier transform of the optical power spectrum. If the normalized spectrum is Gaussian in ordinary frequency, with centre ν_0 and standard deviation σ_ν, completing the square in the Gaussian Fourier integral gives

$$|\gamma(\tau)| = \exp(-2\pi^2\sigma_\nu^2\tau^2),$$

$$\tau_{1/e} = \frac{1}{\sqrt{2}\pi\sigma_\nu}.$$

The rapid carrier phase varies at ν_0, while finite bandwidth sets the envelope. For $\sigma_\nu = 1.0$ THz the amplitude coherence drops to $1/e$ at 0.225 ps. This is 67.5 µm of vacuum optical path difference. In a Michelson geometry the corresponding mirror travel is half that, since light traverses the additional gap twice. An instrument quoting a coherence length should specify which envelope level, which bandwidth convention and which path geometry it uses. Replacing a standard deviation by a full width at half maximum without conversion changes the answer substantially.

Predict three analyser readings before changing the source

Use the polarimetry calculation with a polarized fraction below one. Set the ellipse orientation to 0°, choose positive ellipticity, and compute the horizontal, diagonal and circular contrasts using the definitions above. A change in polarized fraction scales the three contrast components together. Changing orientation alone rotates the linear contrasts through twice the physical angle while leaving the circular contrast fixed. A separate crystal model follows energy flow when two eigenpolarizations propagate in a birefringent plate. For comparison, separate the output phase retardance of a plate from any spatial displacement that could reduce the overlap of two exiting beams on an interferometer detector.

Change ellipticity and reconstruct an unknown polarization in the Stokes experiment

Compare eigenpolarizations and displacement in a crystal

An investigation with a photon-counting polarimeter

A polarimeter reports a count for each analyser, not an exact Stokes vector. A useful second reconstruction starts with a statistical model for those counts. Let all six exposures have the same duration and gain after calibration. For a pair of complementary analyser settings labelled $+$ and $-$, the mean counts at one exposure time are $N(1+s)/2$ and $N(1-s)/2$, with $s = S_i/S_0$. Here N is the expected pair total after background removal. Independent Poisson counting models shot noise from a steady source. If the source changes between exposures this independence assumption no longer describes the common intensity drift.

Condition on the observed sum $n = n_+ + n_-$. The number in the positive channel then has a binomial distribution with success probability $(1+s)/2$. The natural estimator and its conditional variance follow without assuming equal noise in bright and dark channels:

$$\hat{s} = \frac{n_+ - n_-}{n_+ + n_-},$$

$$\operatorname{var}(\hat{s} \mid n) = \frac{1 - s^2}{n}.$$

The uncertainty becomes small near a pure eigenpolarisation in the matching analyser, while a transverse component of the same beam has $s = 0$ and variance about $1/n$. Exact intervals are preferable near the boundary, where a symmetric error bar would extend into unphysical values. Conditioning is appropriate for a

single pair. If one uses the same intensity measurement to normalise several pairs, their reported components become correlated.

Consider the six counts $(800, 200), (620, 380), (550, 450)$ in complementary linear-horizontal, diagonal and circular bases. Their normalised components are 0.60, 0.24 and 0.10. Raw degree of polarisation is

$$\sqrt{0.6^2 + 0.24^2 + 0.10^2} = 0.654.$$

Individual shot-noise estimates are 0.025, 0.031 and 0.031. To first order, uncertainty in the degree is obtained by projecting the component variance along the measured unit vector. For independent pairs the result is approximately 0.026. Fitting an ellipse to this partially polarised beam recovers the polarised component's ellipse; the unpolarised intensity contributes to every positive and negative analyser reading.

For an unpolarised source, taking a square root of three squared noisy estimates always returns a nonnegative number. In fact the expected squared reported degree is approximately $3/n$ when all pair totals equal n. At $n = 1000$ the root-mean-square value is 0.055, even with a perfect instrument. A positive fitted degree around 0.04 in a single exposure therefore cannot establish weak polarisation. Repeat the acquisition at four times the counts. A statistical excess falls in scale by about two, whereas stable true components persist with their signs. A forced projection onto the physical Stokes sphere cannot remove calibration drift or substitute for this uncertainty calculation.

Which-path information encoded in a polarisation state

A balanced interferometer supplies a different quantitative probe. Suppose its path lengths are equal, the beam in arm one has unit horizontal polarisation $\mathbf{e}_1$, and a rotator in arm two produces $\mathbf{e}_2 = \cos\alpha\,\mathbf{e}_1 + \sin\alpha\,\mathbf{e}_y$. The detector receives equal-amplitude contributions with relative phase δ. Squaring the sum in the two output ports of an ideal lossless beamsplitter gives complementary fringes.

$$I_{\pm} = \frac{I_{\mathrm{tot}}}{2}(1 \pm \cos\alpha\cos\delta).$$

For $\alpha = 60^\circ$, sweeping the phase from constructive to destructive moves one port between $0.75 I_{\mathrm{tot}}$ and $0.25 I_{\mathrm{tot}}$; the visibility is 0.50. Their sum should remain constant throughout the sweep. If both measured ports fade together, look for path loss, a changing source or detector gain. Polarisation distinguishability moves fringe energy between the two outputs in this ideal result. It does not remove energy from both.

At $\alpha = 90^\circ$, complete temporal coherence does not suffice for an unpolarised intensity detector to see fringes. The two field vectors are orthogonal and their scalar product vanishes. Insert an analyser at 45° in front of one detector. It projects both contributions onto the same unit vector with equal positive amplitude $1/\sqrt{2}$, so the maximum visibility returns while the phase-averaged brightness of that port halves. An analyser at -45° gives a fringe displaced by half a period. Summing those two orthogonal analyser intensities restores the original featureless signal. The recovered fringes do not require a change in either source clock or path length.

For unequal arm intensities and a time delay, the interference coefficient additionally contains $2\sqrt{I_1 I_2}/(I_1 + I_2)$ and the magnitude of the normalised temporal coherence evaluated at that delay. This factorisation assumes the temporal waveform and polarisation can be separated in each arm. A broadband retarder with appreciable frequency-dependent axes couples the two. Then compute the spectral integral of

the polarisation-resolved field products, preserving the complex phase at each frequency before integration. Multiplying three averaged magnitudes can give the wrong result even if every factor lies between zero and one.

A practical discrimination uses three sweeps at fixed geometry. First measure both single-arm intensities by blocking one arm at a time. Second sweep delay near zero with the analyser removed, and third repeat with the two orthogonal diagonal analysers. If the two diagonal traces regain equal but opposite fringes and sum to the original low-visibility result, distinguishability is a plausible explanation. If none shows variation after path intensities have been balanced, check whether the coherence envelope is narrower than the available delay step. Finite detector bandwidth during a fast phase sweep also suppresses every trace. Repeating the sweep more slowly and reversing its direction distinguishes that response delay from a static polarisation transformation.

Determine what a rotating analyser can detect

Compare projections of a polarised field. Identify which polarisation information is missing from a scan with only a linear analyser.

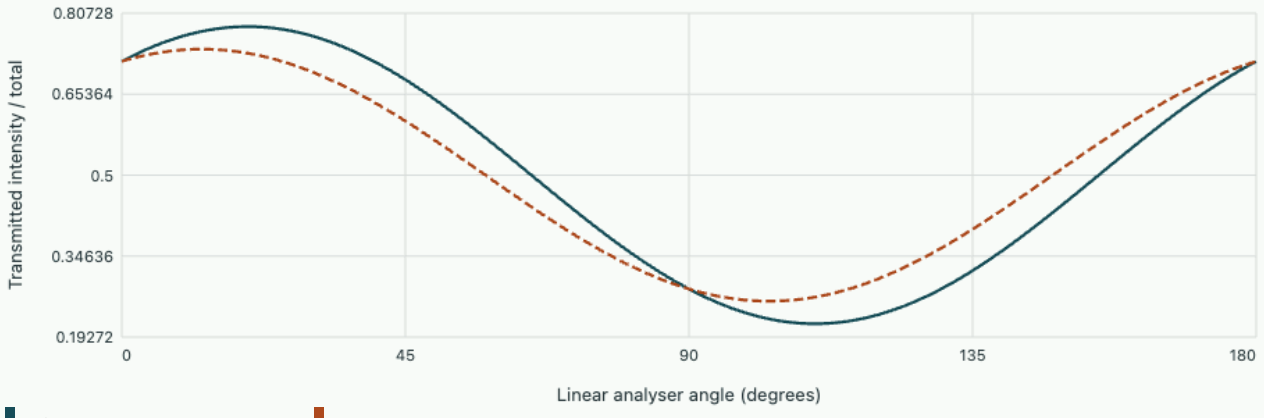
Open this experiment online

Try the calculation

Polarised fraction P **0.6**

Azimuth ψ (degrees) **20**

Ellipticity χ (degrees) **10**



Measured complementary pairs (horizontal, diagonal, circular): 0.716 / 0.284 · 0.6812 / 0.3188 · 0.6026 / 0.3974. Reconstructed $(S_1,S_2,S_3)/S_0$ = (0.4319, 0.3624, 0.2052); degree 0.6.

Worked inverse problem · distinguish a retarder error from source ellipticity

A linear input is sent through a retarder whose fast axis defines the horizontal direction. The input makes 45° to that axis. A diagonal analyser records a normalised Stokes component $s_2 = 0.100$; a circular analyser gives $|s_3| = 0.990$. Determine possible retardance values, then design a measurement separating imperfect retardance from a slightly elliptical input. Assume no diattenuation for the first inference.

▼ **Track the phase and design an independent preparation**

The equal input components acquire a relative phase δ while keeping equal intensities. The outgoing Stokes components are $s_1 = 0$, $s_2 = \cos\delta$, with $|s_3| = |\sin\delta|$. The first measurement gives $\delta = 84.26^\circ$ or 275.74° modulo one full turn. Both predict $|s_3| = 0.995$, close to the reported 0.990. A signed, calibrated circular analyser selects the sense of the quarter-wave action, provided the basis and time convention are stated. Measuring two intensities more precisely cannot tell a one-quarter from a five-quarter retardance at a single frequency.

If the input already possessed relative phase δ_i, the observation measures $\delta + \delta_i$. Bypass the retarder and reconstruct the input. Then rotate the retarder by 90° to exchange its two eigenaxes while holding the input and detector fixed. The retarder-induced phase changes sign; the source phase does not. Two signed phase measurements can recover the sum and difference, modulo ambiguities eliminated by a nearby wavelength or a known approximate plate order.

Record the calibrated horizontal and vertical transmissions as well. Differential loss makes the two output amplitudes unequal and reduces the accessible circular Stokes component even at perfect quarter-wave phase. A measured degree less than one can result from averaging several slightly different retardances across a finite spectrum. Changing the spectral filter tests that hypothesis. These checks illustrate how independent preparations constrain a device and source separately, where one very precise polarisation measurement of their combination cannot.

A rotating source exposes the detector's averaging time

Imagine an otherwise perfectly linearly polarised beam whose axis rotates uniformly during an exposure. At every instant its degree of polarisation is one. A horizontal analyser transmits an intensity proportional to the squared cosine of the axis angle, and a diagonal one samples the same function displaced by forty-five degrees. Over a complete rotation their average intensities are equal. A circular analyser also receives half the total power. Time-integrated tomography would therefore assign zero net polarisation to the exposure.

For a shorter exposure covering an angular sweep Δ about a central angle, average the complex combination $s_1 + i s_2 = e^{2i\theta}$ uniformly over the sweep. Its magnitude becomes $|\sin\Delta/\Delta|$. A ninety-degree sweep retains a degree of approximately 0.637, whereas a full half-turn yields zero because an unoriented linear polarisation axis repeats every half-turn. The reported degree can thus identify neither instantaneous purity nor temporal stability without an exposure model.

Repeat the measurement with shorter synchronised gates and accumulate many exposures starting at the same rotation phase. Stable rotation should recover the predicted phase-resolved Stokes trajectory. A genuinely random uncorrelated orientation cannot be recovered by simply synchronising to an unrelated

clock. If the analyser states are acquired sequentially, synchronise their gates too; otherwise different components may sample different portions of the trajectory and combine into a physically impossible vector.

Tutorial · recover a missing analyser setting and audit positivity

A pair of equal-throughput linear analysers reports $I_x = 4,\ I_y = 2$ and $I_{45} = 4.5$ in the same power units. Find the predicted −45° reading and bound a possible circular contrast. A circular analyser then reports $I_+ = 6.0$. Can these measurements describe a single stationary ensemble?

▼ **Reconstruct only what the measured subspace determines**

The orthogonal sum fixes $S_0 = 6$, hence $I_{-45} = 1.5$, $S_1 = 2$, and $S_2 = 3$. Remaining positivity requires $S_3^2 \le 6^2 - 2^2 - 3^2 = 23$, so $|S_3| \le 4.796$. A positive circular reading of 6 would imply $I_- = 0$ and $S_3 = 6$. That would make the reconstructed coherency matrix have a negative smaller eigenvalue. Merely normalizing all six readings by their common sum would retain the inconsistency. Check gain, background, plate retardance and whether the source changed during sequential measurements. If data have known statistical error, fitting J subject to positive semidefiniteness can enforce physical consistency while exposing the residuals.

Tutorial · distinguish delay decorrelation from an unresolved doublet

Two equal narrow spectral lines are separated by 80 GHz, the arms have equal powers, and polarizations match. Obtain the fringe-visibility envelope. Find the first null and first revival in optical path difference. A single visibility reading at the null is compatible with several physical explanations. Propose a scan that tests the doublet hypothesis.

▼ **Add the complex correlations before taking the magnitude**

Each narrow line contributes a phase at its optical frequency. Averaging the two phasors gives $e^{i2\pi\nu_0\tau}\cos(\pi\Delta\nu\,\tau)$, so visibility is $|\cos(\pi\Delta\nu\,\tau)|$. The first null occurs at $1/(2\Delta\nu) = 6.25$ ps, or 1.87 mm path difference. The first revival is at 12.5 ps or 3.75 mm. In a Michelson instrument those mirror travels are 0.937 and 1.87 mm. For real spectral lines of finite width the revival sits under their own coherence envelope. Scan beyond the first null, independently match arm powers and verify polarization. A clean revival distinguishes a narrow doublet from a single broad Gaussian spectrum reproducing the same initial contrast loss.

MIT 2.71 Optics · polarisation, interference and spatial coherence

Oblique interfaces and films whose reflections cancel

Waves, radiation and relativity

A film only one quarter of a wavelength thick can suppress reflection from a transparent surface. Follow the boundary conditions all the way from an oblique incident wave to the input admittance of a stack, then calculate how much reflection returns when the wavelength, angle or layer order changes.

Tangential phase is an observable condition at an extended interface

Let the plane z=0 separate two homogeneous, lossless, isotropic dielectrics. Incident light lies in the xz plane and travels towards positive z. Use phasors $\exp[i(k_x x + k_z z) - i\omega t]$; choose square roots for transmitted waves so energy flows away from the interface, or evanescent fields decay into the second medium. Linear time-independent boundary conditions hold at every point x and time t. If their Fourier factors had unequal k_x or ω, cancellation of the tangential boundary fields at every x,t would be impossible. Thus incident, reflected and transmitted waves share both quantities. The reflected z component reverses sign. The transmitted normal component follows from its own medium's dispersion relation.

$$k_x = k_0 n_1 \sin\theta_1 = k_0 n_2 \sin\theta_2,$$

$$k_{zj} = \sqrt{k_0^2 n_j^2 - k_x^2},$$

$$k_0 = \omega / c.$$

The angles are measured from the normal. Equality of tangential phase gives Snell's law independently of a ray argument, and remains useful when the normal wavenumber becomes imaginary and a real ray angle no longer exists. For example, vacuum light at 600 nm falling at 45° onto a dielectric of index 1.50 has $k_x = 7.40 \times 10^6$ m⁻¹ and $\theta_2 = 28.1^\circ$. Relative phase across a pair of boundary probes at fixed separation along x must be the same whether they detect the incident or transmitted wave. Applying a different frequency-dependent normal propagation constant in each material does not violate that observation.

Reduce each polarization to one tangential electric amplitude

Define s polarization with E along y and p polarization with H along y. For s let $e = E_y,\ h = -H_x$; for p let $e = E_x,\ h = H_y$. Both choices make normal average flux $\mathrm{Re}(eh^*)/2$. Maxwell's curl relation $\mathbf{H} = (\mathbf{k} \times \mathbf{E})/(\mu\omega)$, together with $\mathbf{k} \cdot \mathbf{E} = 0$, gives for a forward wave $h = Ye$ and for a reverse wave $h = -Ye$, where

$$Y_s = \frac{k_z}{\mu\omega} = \frac{\cos\theta}{Z},$$

$$Y_p = \frac{\epsilon\omega}{k_z} = \frac{1}{Z\cos\theta}.$$

For p polarization e is the tangential projection of the total E vector. Many optics tables use a reflected p basis whose tangential component flips sign on reflection. The resulting p reflection coefficient differs in sign from this tangential-e coefficient; predicted powers and a consistent full electric field agree. Here all layers use the same laboratory tangential axes, which removes unnecessary sign conversions between interfaces.

In the absence of free surface currents the two tangential boundary conditions give

$$e_i + e_r = e_t,\; Y_1(e_i - e_r) = Y_2 e_t.$$

Solving for reflected and transmitted tangential amplitudes yields

$$r = \frac{Y_1 - Y_2}{Y_1 + Y_2},$$

$$t = 1 + r = \frac{2Y_1}{Y_1 + Y_2},$$

$$R = |r|^2,$$

$$T = \frac{\Re Y_2}{\Re Y_1}|t|^2.$$

For lossless propagating waves Y is real and $R + T = 1$ follows by expanding the two squared binomials. The amplitude transmission can exceed one when transmitted admittance is smaller, because energy is proportional to admittance times squared tangential field. Omitting the admittance factor leads to failed power balance even if the two field equations were solved perfectly. At normal incidence into index 1.5 from air, r=−0.20, t=0.80, R=0.04, and T=$1.5(0.80)^2 = 0.96$.

Distinguish Brewster cancellation from total internal reflection

For positive nonmagnetic indices, p reflection vanishes when $n_1/\cos\theta_1 = n_2/\cos\theta_2$. Combined with Snell's law this yields $\tan\theta_B = n_2/n_1$ at a nontrivial interface. At that angle p admittances match even while phase speed and ray direction differ across the boundary. At air-to-glass incidence with index 1.5 the Brewster angle is 56.31°. The s reflection at that same angle is substantial because s admittance varies with cosine in the opposite direction. A measurement of an intensity minimum only approximately zero can reflect polarization impurity, absorption or a surface film; a minimum angle alone cannot establish all three properties.

If $n_1 > n_2$, transmission at incidence above $\theta_c = \arcsin(n_2/n_1)$ gives

$$k_{z2} = i\alpha.$$

The decaying exponential satisfies the same tangential phase matching. For a single semi-infinite lossless second medium its admittance is imaginary, so T vanishes and $|r| = 1$. A finite evanescent field remains on the low-index side and carries energy parallel to the surface. For glass index 1.50 facing air at 50° incidence and 600 nm vacuum wavelength, $\alpha = k_0\sqrt{(1.5\sin 50^\circ)^2 - 1} \approx 5.93 \times 10^6$ m⁻¹. The electric amplitude decays to $1/e$ in about 169 nm; intensity in about 84.5 nm. A second high-index surface within that region changes the boundary condition and can admit transmitted power.

Locate the Brewster zero and total-reflection threshold for two dielectric indices

Propagate a forward and a backward field through one film

For a film of thickness d the entrance phasors are $e_0 = A + B,\; h_0 = Y(A - B)$. At the exit they are $e_d = Ae^{i\delta} + Be^{-i\delta},\; h_d = Y(Ae^{i\delta} - Be^{-i\delta})$, with $\delta = k_z d$. Eliminate A and B to propagate the exit fields back to the entrance:

$$\begin{bmatrix} e_0 \\ h_0 \end{bmatrix} = \begin{bmatrix} \cos\delta & -i\sin\delta/Y \\ -iY\sin\delta & \cos\delta \end{bmatrix} \begin{bmatrix} e_d \\ h_d \end{bmatrix}.$$

A one-line formula conceals useful checks. At zero thickness the matrix is the identity. Its determinant is one because $\cos^2\delta + \sin^2\delta = 1$. At a half-wave thickness the matrix is minus the identity; both amplitudes change sign and their ratio is unchanged. For a stack, form the product in physical order from entrance to exit because each matrix here maps its own exit backward. Reversing two layers can change the response. For very thick evanescent layers transfer matrices contain exponentially large and small entries; a scattering matrix or scaled recursion can avoid catastrophic loss of precision.

If the substrate is semi-infinite and admits only an outgoing wave, its boundary admittance is $Y_b = h_d/e_d$. Divide the two matrix rows to obtain

$$Y_{\mathrm{in}} = \frac{Y_b\cos\delta - iY\sin\delta}{\cos\delta - i(Y_b/Y)\sin\delta}.$$

Reflection from the whole film at its entrance now follows from the same single-interface equation with $Y_2 = Y_{\mathrm{in}}$. At a quarter wave, $\delta = \pi/2$, this simplifies to $Y_{\mathrm{in}} = Y^2/Y_b$. From the incident medium the film and substrate look perfectly matched if $Y = \sqrt{Y_0 Y_b}$. The field in the film still contains both travelling directions; zero reflected field at the entrance fixes a coherent sum of all returns. Adding reflected intensities interface by interface would destroy the interference responsible for the matching.

Design a coating and predict an out-of-band reading

At normal incidence choose $n_0 = 1,\ n_b = 2.25$, with no magnetic response. The required coating index is 1.50. At $\lambda_0 = 600$ nm choose $d = \lambda_0/(4n) = 100$ nm. The uncoated substrate reflects $R = [(1-2.25)/(1+2.25)]^2 = 0.148$. The ideal coated substrate has R=0 at design wavelength. A half-wave optical thickness instead returns the original admittance and restores the 14.8% reflection. These two limits can detect a factor of two in a code that confuses a one-way phase with a round-trip phase.

For a single film the same result follows by summing multiple reflections. Let $r_{ij} = (Y_i - Y_j)/(Y_i + Y_j)$. Each additional round trip contributes $r_{10}r_{12}e^{2i\delta}$. Summing the geometric series and using $r_{10} = -r_{01}$ yields

$$r = (r_{01} + r_{12}e^{2i\delta})/(1 + r_{01}r_{12}e^{2i\delta}).$$

In the quarter-wave example both interface coefficients are –0.20. If wavelength shifts to 500 nm at fixed indices, $\delta = 0.6\pi$. The reflected power is $0.04[2 + 2\cos(1.2\pi)]/[1 + 0.04^2 + 0.08\cos(1.2\pi)] \approx 0.0163$. The coating still reduces reflection relative to the bare substrate, but a strictly zero field requires the design phase and amplitudes to match.

At oblique incidence the phase uses the film's normal wavenumber, giving $\delta = 2\pi n d\cos\theta_{\mathrm{film}}/\lambda_0$. Increasing incidence angle shortens the design wavelength that achieves the same phase. At the same time the matching admittances change differently for s and p, so keeping only the phase shift is incomplete. For illumination by a focused beam a range of tangential wavenumbers reaches the substrate. A single normal-incidence calculation cannot predict its whole cone's reflection without an angular/polarization weighting.

Build a finite dielectric mirror from admittance transformations

Suppose the incident side sees an alternating high-index H and low-index L layer, each quarter-wave thick at normal incidence. Starting at the substrate admittance Y_b, the last low layer gives Y_L^2/Y_b; the preceding high layer transforms that to $Y_b(Y_H/Y_L)^2$. Repeating N pairs yields

$$Y_{\text{in}} = Y_b(Y_H/Y_L)^{2N}.$$

The growth is geometric in pair count despite each individual interface reflecting only a fraction of the wave. This is a coherent reflection measurement and depends on all pair phases staying near their design values.

Take $n_H = 2,\ n_L = 1.5,\ n_0 = n_b = 1.5$. With three HL pairs the normalized input admittance is $(4/3)^6 = 5.62$ times its external value, giving R about 0.487. With six pairs the ratio is $(4/3)^{12} = 31.6$, giving R about 0.881. The same total number of random-thickness layers need not achieve those values. Swapping HL for LH in the symmetric exterior configuration inverts the input admittance ratio. It reverses the reflected field phase while leaving the centre-wavelength reflected intensity unchanged. Phase-sensitive reflection can therefore distinguish two stacks with identical intensity at one wavelength.

A stack on a substrate different from the input medium removes that simple intensity degeneracy. The layer order can then alter centre-wavelength reflected intensity as well. Fabrication tolerances introduce additional phase shifts and can split spectral extrema. To identify a real film from measured data, compare phase or several wavelengths and polarization states, and account for instrument spectral width. One scalar normal-incidence reflectance generally admits many index-thickness pairs.

From a stack of interfaces to a spectral ruler

A pair of parallel partially reflecting surfaces makes more than a mirror. A short incident pulse emerges as a train of diminishing replicas, and a continuous wave samples their coherent sum. Work first at normal incidence in a lossless gap of refractive index n and thickness L. Identical ideal mirrors have intensity reflectance R. An extra round trip therefore multiplies a transmitted field by $Re^{i\delta}$, where $\delta = 2n\omega L/c$ plus any frequency-independent mirror phase. At a resonance this phase is an integer multiple of 2π. The directly transmitted field magnitude contains the product of the two mirror transmissions, $1 - R$ in the symmetric lossless case. Sum the geometric series and take the magnitude squared.

$$\frac{I_{\text{out}}}{I_{\text{in}}} = \frac{(1-R)^2}{|1 - Re^{i\delta}|^2} = \frac{1}{1 + F\sin^2(\delta/2)},$$

$$F = \frac{4R}{(1-R)^2}.$$

The unit transmission at resonance follows from interference and symmetry. It does not imply that the mirrors transmit the whole incident wave on the first encounter. At the front surface a leaked internal wave cancels the directly reflected wave. During switch-on energy accumulates in the gap, so imposing the steady resonance formula on the leading edge of a finite pulse would violate the assumed steady state. If the two mirrors have different transmissions or the gap absorbs, the two reflected amplitudes at the entrance no longer necessarily cancel at resonance.

Adjacent maxima differ in ordinary frequency by $c/(2nL)$ if dispersion and mirror phase slopes are negligible. If the material phase $n\omega L/c$ is dispersive, differentiate the round-trip phase. The local mode spacing

becomes the inverse of the round-trip group delay, including the frequency derivative of reflective phase. Replacing n by a phase index measured far from that frequency can then bias an inferred length. At an oblique angle, decide whether the scan holds external angle or tangential wavevector fixed before differentiating the phase.

For $L = 10.0$ mm, $n = 1.50$ and $R = 0.81$, the free spectral range is 9.99 GHz. The half-power points satisfy

$$\sin^2(\delta/2) = 1/F.$$

Their full frequency separation divided by one free spectral range is $2\arcsin(1/\sqrt{F})/\pi$. Here $F = 89.75$ and the linewidth is about 0.673 GHz. The finesse, spacing divided by linewidth, is 14.9. The familiar high-reflectance approximation $\pi\sqrt{R}/(1-R)$ also gives 14.9. For a very weak mirror there need not be half-maximum points at all because the minimum remains above half the peak. The exact formula tells you when a linewidth extracted by that convention is undefined.

Verify the same length in time and find a small hidden loss

The energy of a freely decaying intracavity pulse is multiplied by R^2 each round trip for identical mirrors. With round-trip time $t_r = 2nL/c$, the continuous exponential through the discrete samples has energy lifetime $\tau_E = -t_r/[2\ln R]$. For the above gap $t_r = 100.1$ ps and $\tau_E = 238$ ps. Its reciprocal $1/(2\pi\tau_E)$ is 0.670 GHz, close to the direct linewidth. This equality is an approximation to the isolated narrow resonance. A short enough pulse can resolve the individual replicas and reveal directly that the idealised trace is a train, not a continuously measured exponential at every instant.

Now introduce a weak internal loss. Let a be the intensity fraction remaining after one pass through the material. An additional round trip multiplies the circulating field by Ra, while the direct transmitted intensity is multiplied by a. Thus the resonant transmission and freely decaying energy factor are

$$T_{\mathrm{peak}} = \frac{(1-R)^2 a}{(1-Ra)^2},$$

$$g_E = R^2 a^2.$$

If the measured ratio of successive isolated replicas is 0.630, assuming the mirrors still have $R = 0.81$ gives

$$a = \sqrt{0.630}/0.81 = 0.980.$$

The predicted peak transmission is about 0.83, substantially below one despite the small two-percent single-pass loss. Multiple passes amplify the effect of absorption. Energy reflected in the mode, absorbed in the gap and transmitted are separate channels. A detector collecting only the central transmitted lobe can also show a low peak when the incident beam couples poorly to the cavity, even with very low material absorption.

An inference procedure uses an intensity-calibrated bypass beam to determine the incident power, a direct time spacing to estimate $n_g L$ and ringdown ratios to estimate Ra. If R is unknown, ringdown alone constrains only the product. A resonant transmission measurement supplies a second equation in the ideal symmetric setup, but beam mismatch and mirror imbalance add unknowns. Block one mirror or measure it separately to obtain an independent reflectance. Small thermal changes of gap length shift the resonance

between repeated readings. Sweep the frequency and fit the full peak instead of observing its height at a single nominally resonant wavelength.

Finally examine a common interpretation trap with broadband illumination. A spectrometer pixel integrating many evenly spaced narrow resonances reports their average transmission, while a narrow-line source scanned sufficiently slowly recovers their maxima and minima. The difference is predicted by averaging the intensity spectrum with the instrument response, not by summing the fields emitted from mutually incoherent source frequencies. Conversely the repeated passes of a single resolved frequency are mutually coherent and must be summed as fields. The same device contains both averaging operations, in a definite order. This order follows from the detector integration time and the source correlations.

Apply the interface conditions to a measurable power ratio

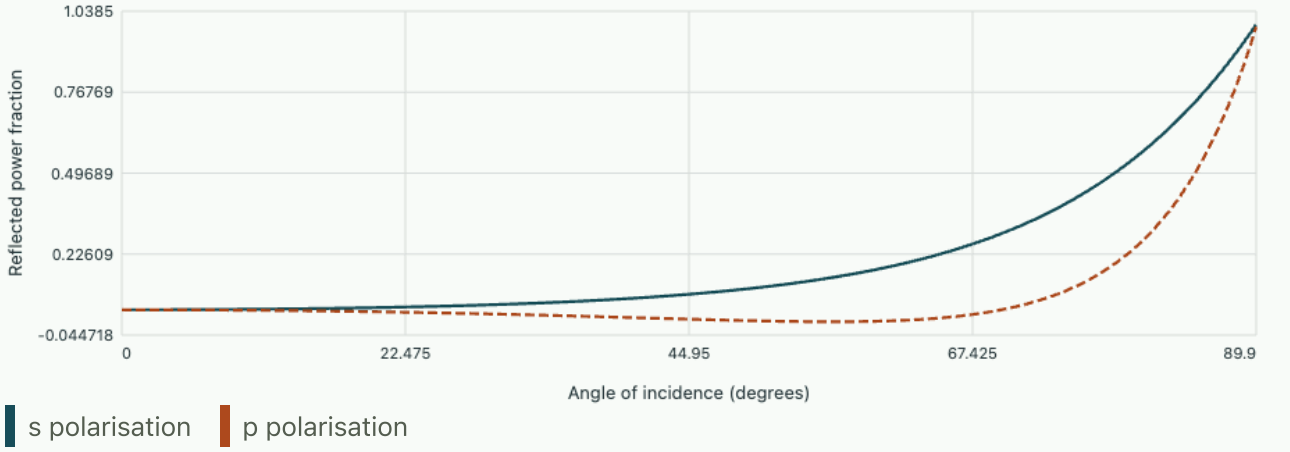
Change incidence angle and the receiving medium. Compare amplitude reflection with the power that actually crosses the interface.

Open this experiment online

Try the calculation

Incident refractive index **1**

Transmitted refractive index **1.5**



The Brewster minimum occurs at 56.3°. At 45°, reflected power is 9.2% (s) and 0.847% (p).

Worked boundary problem · where does the energy of a totally reflected wave go?

A lossless glass prism with refractive index 1.50 is illuminated internally at 50°, adjacent to air. Vacuum wavelength is 600 nm. Calculate the phase propagation along the interface, the normal field decay length in air and the associated one-sided intensity decay length. Can a detector in the second medium remove power without changing the reflection problem?

▼ **Evaluate the complex wavevector and its physical flux**

Continuity of tangential phase gives

$$k_x/k_0 = 1.50 \sin 50^\circ = 1.149.$$

The normal air wavenumber has imaginary part $\kappa = k_0\sqrt{1.149^2 - 1} = 0.566\,k_0$, giving amplitude decay length $1/\kappa = 169$ nm. Squared field decays over about 84.5 nm. The tangential phase period is

$$\lambda_0/1.149 = 522\,\mathrm{nm}.$$

Treating this as an air plane wave with a real propagation direction would imply an impossible sine and lose the normal decay that satisfies the boundary condition.

For s polarisation take $E_y \propto e^{ik_x x - \kappa z}$. The wave equation gives

$$H_x = -i\kappa E_y/(\mu\omega),$$
$$H_z = k_x E_y/(\mu\omega).$$

In $\mathrm{Re}(\mathbf{E} \times \mathbf{H}^*)/2$, the normal component vanishes because the needed product is purely imaginary, but the tangential component is positive. A finite evanescent field thus accompanies unit reflected power at this ideal isolated interface. Zero time-averaged normal flux does not require a zero field in the adjacent material.

Introducing a second prism or a genuinely absorbing near-field probe changes the boundary condition. A field returning from the second boundary permits interference that can carry normal energy. A perfectly calibrated detector nevertheless perturbs the solution if it extracts appreciable power. Increase the air gap and compare the recovery of unit reflectance as well as the detected signal. A decay in detected intensity alone could also reflect geometric collection changes, while agreement with the complementary reflected-power change constrains the power path.

A wedge turns phase dispersion into a spatial measurement

Two plates forming a small air wedge provide a complementary way to measure a layer without scanning the source frequency. Let the gap rise linearly with transverse position at small angle α. In reflection at nearly normal incidence successive equivalent fringes correspond to a full additional round-trip phase, hence to an added gap of half a wavelength in the intervening medium. For a 633 nm source and a measured fringe spacing of 0.80 mm in air, the inferred wedge slope is about 0.396 milliradians. The phase reversals on the plates choose whether the contact end is bright or dark but do not change this spacing.

If the refractive index is unknown, a single spacing constrains only the product of gap slope and index. An independent surface-profile measurement provides the geometric slope. Alternatively fill the same wedge with a transparent fluid and compare its spacing against air, keeping geometry and temperature fixed. The ratio of air spacing to fluid spacing measures the ratio of indices within the local parallel-film approximation.

Several identifiable departures can occur. A finite camera pixel averages neighbouring fringes, a broad source loses contrast with increasing path difference and a curved plate produces a changing fringe spacing. The

first two principally reduce visibility; the third changes local phase gradient. Fit fringe positions and their visibility as separate observables before attributing either effect to absorption inside the film.

Tutorial · measure a Brewster index and bound an unwanted polarization

A transparent nonmagnetic sample in air has an observed p minimum at 58.0°. Assume for this estimation that it is a single dielectric interface. Infer its index, calculate the s reflectance there, and estimate what fraction of incident power could be s-polarized if the measured minimum reflectance is 0.0040 and all other backgrounds are negligible.

▼ **Derive and use the second polarization as a calibration**

The inferred index is

$$\tan 58^\circ = 1.600.$$

At the Brewster angle the transmitted angle is 32°. Relative s admittance is $n\cos 32^\circ/\cos 58^\circ = n^2$ for air incidence. Thus $R_s = [(1-n^2)/(1+n^2)]^2 \approx 0.192$. An incident s fraction f contributes fR_s to reflected power while the ideal p component contributes zero. Matching 0.0040 gives $f = 0.0208$, about 2.1%. This is an upper bound if another nonnegative reflected background is present. Absorption or a coating can also displace the minimum, invalidating the simple index fit itself. Measuring the complementary s curve and checking transmitted power provide independent tests before interpreting the impurity.

Tutorial · separate a thickness error from an index mismatch

A nominal 600 nm quarter-wave coating with design index 1.5 on an index 2.25 substrate is measured to be 102 nm thick. First hold all indices fixed. Predict the wavelength of its exact null, then estimate R at 600 nm to leading order. If its index is actually 1.55, can changing only the thickness recover an exact null?

▼ **Expand the coherent reflection and apply the admittance condition**

The quarter-wave wavelength is $4nd = 612$ nm. At 600 nm the phase deviates from $\pi/2$ by $\Delta = \pi(0.02)/2 = 0.0314$. With $r_{01} = r_{12} = r_s = -0.20$, the small-amplitude reflection is $r \simeq -2ir_s\Delta/(1-r_s^2)$. Thus $R \simeq 4r_s^2\Delta^2/(1-r_s^2)^2 = 1.71\times10^{-4}$. An index of 1.55 gives unequal magnitudes at the two interfaces. Varying the phase can put their return fields in opposition but cannot equalize their magnitudes. At a quarter wave the best input admittance is proportional to $1.55^2/2.25 = 1.0678$, giving R about 0.00107. A wavelength scan can locate the minimizing optical thickness while the depth of that minimum constrains the impedance mismatch.

MIT 6.013 · boundary conditions, oblique waves, impedance transformations and layered systems

Causal material response and the propagation of a finite pulse

Waves, radiation and relativity

An optical pulse entering a transparent slab is delayed, broadened and sometimes reshaped. Derive these effects from the motion of bound charge, and learn what a measured group delay actually says about a pulse carrying information.

Turn the idea of delayed response into an equation

At a point in a stationary isotropic dielectric, a short electric impulse displaces bound charge. The displacement can continue after the applied field disappears. In a linear time-translation-invariant material write the polarization as a weighted sum over earlier values of the local electric field:

$$P(t) = \epsilon_0 \int_0^\infty h(\tau) E(t-\tau)\, d\tau,$$

$$\chi(\omega) = \int_0^\infty h(\tau) e^{i\omega\tau}\, d\tau.$$

The lower limit encodes causality. Switching to steady harmonic fields yields $\widetilde{P} = \epsilon_0 \chi(\omega) \widetilde{E}$ under the $e^{-i\omega t}$ convention. An instantaneous contribution could also be represented by a delta function at zero delay. The response kernel h has dimensions of inverse time when susceptibility is dimensionless. A short memory compared with the period can sometimes be approximated by a constant susceptibility over a limited band. Extending a constant complex response to all frequencies would generally violate either causality or the requirement of a real time-domain polarization.

For real h, $\chi(-\omega) = \chi(\omega)^*$, so its real part is even and its imaginary part odd. For a decaying kernel, extending ω into the upper half of the complex plane introduces $e^{-\operatorname{Im}\omega\,\tau}$, ensuring convergence. Analyticity there links dispersion and absorption. Conversely, a fitted pole in the upper half-plane in this convention would imply growth, an incompatible time convention, or an unstable model. Report the phasor convention beside an empirical complex permittivity to assign the sign of loss correctly.

Obtain a susceptibility with a recognizable resonance

Represent a bound charge of magnitude q and mass m by a displacement x, restored with frequency ω_0 and damped at rate γ. Choose x along the direction of the force on a positive effective charge; the resulting susceptibility depends on q^2. For a density N of independent oscillators the polarization is Nqx. The equation of motion and its phasor solution give

$$m(\ddot{x} + \gamma\dot{x} + \omega_0^2 x) = qE,$$

$$\chi = \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega},$$

$$\omega_p^2 = \frac{Nq^2}{m\epsilon_0}.$$

Writing $D = (\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2$ gives

$$\chi' = \omega_p^2(\omega_0^2 - \omega^2)/D,$$
$$\chi'' = \omega_p^2\gamma\omega/D.$$

A positive damping rate makes $\chi'' > 0$ at positive frequency. At low frequency the displacement follows the drive almost in phase, while far above resonance inertia produces the opposite displacement. At resonance the displacement lags the drive by a quarter cycle, making the velocity align with force and maximizing the ability of damping to absorb average work for a given displacement amplitude.

The energy balance follows directly by multiplying the equation by $N\dot{x}$. The field supplies $E\dot{P}$, equal to the rate of change of oscillator kinetic plus restoring energy and the dissipated power $Nm\gamma\dot{x}^2$. Averaging a harmonic cycle eliminates the change of stored mechanical energy. With peak phasors,

$$\langle E\dot{P}\rangle = \tfrac{1}{2}\omega\epsilon_0\chi''|\widetilde{E}|^2 = \tfrac{1}{2}Nm\gamma\omega^2|\widetilde{x}|^2.$$

This supplies a practical sign check on both the material fit and the assumed time dependence. The imaginary part that attenuates a passive forward wave must also give nonnegative cycle-averaged heating.

Reconstruct the dispersive shift from its loss spectrum

For a response that vanishes sufficiently fast at high frequency, analyticity in the upper half-plane permits a contour integral of $\chi(z)/(z-\omega)$. Close the contour above and indent around the pole on the real axis. The large arc vanishes under the decay assumption. Taking real and imaginary parts of the resulting principal-value relation, then using the symmetry between positive and negative frequencies, gives

$$\chi'(\omega) = \frac{2}{\pi}\,\mathcal{P}\int_0^\infty \frac{\Omega\chi''(\Omega)}{\Omega^2-\omega^2}\,d\Omega.$$

The principal value means combine contributions on both sides of the singular frequency before taking the limit. At zero frequency there is no interior singularity and $\chi'(0) = (2/\pi)\int_0^\infty \chi''(\Omega)/\Omega\,d\Omega$. A passive absorption band produces a finite static polarization contribution even when it lies far above the optical frequencies being measured. If instantaneous high-frequency polarization is included in a model, subtract that constant background before using this vanishing-at-infinity form.

As a simple check, let a measured narrow absorption band have area $A = \int \chi''\,d\Omega = 3.0 \times 10^{14}$ s⁻¹ centered at $\Omega_0 = 2.0 \times 10^{15}$ s⁻¹. Across the band $1/\Omega$ varies little, giving an estimated contribution $2A/(\pi\Omega_0) = 0.0955$ to $\chi'(0)$. The same numerical area translated to twice the frequency contributes only half as much. A spectrometer limited to one window cannot recover an absolute static susceptibility without assumptions about absorption outside its window. Tail models and independent low-frequency measurements constrain this extrapolation.

Distinguish the carrier phase from the envelope delay

In a nonmagnetic bulk medium define $k(\omega) = \omega n(\omega)/c$, choosing the branch of n that attenuates a wave travelling in positive z. A component accumulates $\exp[ik(\omega)L]$, so its field amplitude decays as

$\exp[-\operatorname{Im} k\,L]$ and its intensity twice as rapidly in the exponent. In a transparent band take k effectively real and expand around carrier ω_c:

$$k(\omega_c+\Omega) = k_c + k_1\Omega + \tfrac{1}{2}k_2\Omega^2 + \cdots,$$

$$k_1 = \frac{1}{v_g} = \frac{n + \omega_c\, dn/d\omega}{c}.$$

The constant $k_c L$ rotates the carrier phase. The term $k_1 L\,\Omega$ multiplies the envelope spectrum by a linear phase, corresponding to translation in time by $k_1 L$. The second derivative produces different delays for different spectral components and generally changes the pulse shape. These three operations can be measured separately with an interferometer, a time-of-flight experiment and a pulse-duration measurement. A continuous-wave phase measurement at one frequency does not independently determine k_1 or k_2.

At 800 nm suppose an index measurement gives $n = 1.50$ and $dn/d\lambda_0 = -0.050$ per µm, where λ_0 is vacuum wavelength. By differentiating $\omega = 2\pi c/\lambda_0$,

$$n_g = n - \lambda_0\, dn/d\lambda_0 = 1.54.$$

A 100 mm slab has carrier phase travel time $nL/c = 0.5003$ ns and narrowband envelope time $n_g L/c = 0.5137$ ns. Compared with a vacuum path of equal length the envelope delay is 0.1801 ns. The 13.3 ps difference between phase and group times is small relative to a nanosecond but enormous compared with a 50 fs pulse width.

Derive the width of a dispersed Gaussian by completing a square

Use an input envelope $A(0,t) = A_0\exp[-t^2/(2T_0^2)]$, so its intensity is proportional to $\exp[-t^2/T_0^2]$. The amplitude spectrum is proportional to $\exp[-T_0^2\Omega^2/2]$. In a frame at retarded time $u = t - k_1 L$, multiplication by the quadratic propagation phase gives a Gaussian integral with complex coefficient $T_0^2 - ik_2 L$. Evaluate that integral with its physical square-root branch:

$$\frac{A(L,u)}{A_0} = \frac{1}{\sqrt{1 - iD}}\exp\left[-\frac{u^2}{2T_0^2(1-iD)}\right],$$

$$D = \frac{k_2 L}{T_0^2}.$$

Hence $I(L,u) = I_0(1+D^2)^{-1/2}\exp[-u^2/(T_0^2(1+D^2))]$. The intensity width broadens by $\sqrt{1+D^2}$ and the peak falls by the reciprocal factor. Integrating over u recovers the original pulse energy per area, as expected for lossless propagation. The phase inside the exponential is quadratic in u, so the instantaneous frequency varies across the envelope. Measuring only output intensity cannot tell the sign of k_2, since the width depends on its square. A phase-sensitive measurement or a known initial chirp can reveal that sign.

Take $T_0 = 20$ fs and integrated dispersion $k_2 L = 400$ fs². Then D=1. The incident intensity FWHM is $2\sqrt{\ln 2}\,T_0 = 33.3$ fs, and the emerging FWHM is 47.1 fs. The peak intensity falls to 70.7% while the energy remains fixed. In a second slab of the same length with $k_2 L = -400$ fs², the quadratic phases cancel, provided loss and higher-order dispersion remain negligible and the beams couple without appreciable filtering. Propagating with group velocity alone would predict the arrival of the centre but miss these measurable width changes.

When a measured peak time no longer tracks an unaltered envelope

Near an absorption line both real and imaginary parts of k can vary strongly over a pulse spectrum. A simple propagation experiment then applies a complex filter $H(\omega) = t_{\text{interfaces}}(\omega) e^{ik(\omega)L}$, where interface reflections may also matter. A severely attenuated output maximum may move forward because the trailing part of the incoming waveform is preferentially removed. A negative slope of the transmission phase can therefore coexist with a causal impulse response. The arrival of a newly imposed nonanalytic wavefront requires the high-frequency behaviour of the full causal transfer function. A band-limited Gaussian already has tails at every finite time and has no sharp front.

There is also a distinction between phase delay and energy deposition in an opaque medium. If $n = n' + in'' = 1.5 + 0.02i$ at 600 nm, $\operatorname{Im} k = 2\pi(0.02)/\lambda_0 = 2.09 \times 10^5$ m⁻¹. Ignoring reflections, intensity falls to $1/e$ in $1/[2 \operatorname{Im} k] = 2.39$ µm. After 5.0 µm the intensity fraction is

$$e^{-2.094} = 0.123.$$

A pulse delay inferred from the remaining photons is a conditional timing measurement on transmitted energy; it does not describe energy that was deposited at shallower depths.

Separate a resonance shift, a loss peak and a spectral weight

In the oscillator calculation, hold oscillator strength fixed and compare two damping rates. Predict the imaginary susceptibility at resonance from $\omega_p^2/(\gamma\omega_0)$; halving damping should double this height. The static susceptibility ω_p^2/ω_0^2 stays fixed. A narrower taller peak can preserve its weighted spectral contribution. Record real susceptibility on both sides of the resonance, where its sign changes. Then change the oscillator strength and explain why this change scales absorption and dispersion together. Use the separate half-space calculation to check that the absorbed power integrated over depth accounts for the part of incoming flux that passes the surface.

Change damping in the bound-charge susceptibility experiment

Check how integrated heating closes an energy balance

A causal relaxation whose spectrum can be integrated exactly

As a second material model, consider permanent molecular dipoles that reorient toward a weak applied field. When rotational inertia relaxes much faster than the times of interest, a single exponential approach to equilibrium is useful. Let the additional static susceptibility of these dipoles be $\Delta\chi$, and let τ be the relaxation time. For the extra polarisation write $\tau\dot{P} + P = \epsilon_0 \Delta\chi E$. This is a different material approximation from a weakly damped oscillator. It still has a retarded response and provides an independent test of the dispersion–loss relation.

Solve the inhomogeneous differential equation by multiplying it by $e^{t/\tau}$, integrating from an unpolarised remote past and dividing by the same factor. A unit step turns on exponentially. A harmonic drive gives a complex susceptibility with one pole on the lower imaginary frequency axis for the chosen $e^{-i\omega t}$ sign convention.

$$P(t) = \frac{\epsilon_0 \Delta\chi}{\tau} \int_{-\infty}^{t} e^{-(t-t')/\tau} E(t')\, dt',$$

$$\chi' = \frac{\Delta\chi}{1+\omega^2\tau^2},$$

$$\chi'' = \frac{\Delta\chi\, \omega\tau}{1+\omega^2\tau^2}.$$

At $\omega\tau = 1$ the in-phase and lagging components are equal to $\Delta\chi/2$, and the loss component reaches its maximum. A material with $\tau = 8.0$ ns therefore peaks near 19.9 MHz. A step reaches 63.2 percent of its final additional polarisation after 8.0 ns and 95.0 percent after 24.0 ns. If the chosen observable is current rather than polarisation, the relaxation current is proportional to $\dot{P}$ after the step and decays, while P itself rises. An instantaneous background polarisation can deliver a narrow current spike in addition to this resolved tail.

Substitute χ'' into the zero-frequency dispersion relation. With $x = \omega\tau$, the integral becomes

$$(2\Delta\chi/\pi) \int_0^\infty dx/(1+x^2) = \Delta\chi.$$

The high-frequency background, if present, must be added separately to obtain the total static response. This exact integral is also a test for loss-spectrum simulations on a logarithmic grid. The integration measure is $d\omega/\omega = d\log\omega$. Equal weights per ordinary frequency or a missing conversion between natural and decimal logarithms gives an incorrect static increment even if the plotted peak is perfect.

A spectrometer covering $0.1 \le \omega\tau \le 10$ captures $2[\arctan 10 - \arctan(0.1)]/\pi = 0.873$ of the static increment. Two full decades centred on the peak miss almost thirteen percent. The missing low and high frequency contributions are equal in this particular symmetric logarithmic loss model. A superposition of slower relaxations or a second unrelated absorption feature destroys this equality. Report a static measurement or bounds on those extra contributions before interpreting a small discrepancy as failure of causality.

Drive the relaxation twice and distinguish memory from propagation delay

Apply a rectangular field E_0 from time zero to T, then switch it off. Continuity of polarisation at both switching instants gives

$$P(T) = \epsilon_0 \Delta\chi E_0 (1 - e^{-T/\tau}).$$

After switch-off the value decays as $P(T)e^{-(t-T)/\tau}$. For $T = \tau$ the maximum is 0.632 of the static value and at 2τ the remainder is 0.233. A pure delay with the same time of maximum would instead translate the rectangular waveform with unchanged amplitude and abrupt edges. One can test this distinction from the complete waveform even if a slow instrument makes identifying the first nonzero response difficult.

Follow the first pulse by an opposite pulse of equal height starting after a zero-field gap g. At the second start the memory is $P_s = \epsilon_0 \Delta\chi E_0 (1 - e^{-T/\tau}) e^{-g/\tau}$. The second pulse drives the polarisation toward the opposite equilibrium. It crosses zero during the pulse after a further time $\tau \ln[1 + P_s/(\epsilon_0 \Delta\chi E_0)]$. For $T = \tau$

and $g = \tau$ the result is 0.209τ, or 1.67 ns with the above time constant. This cancellation depends on the gap. A nonlinear saturation or a detector recovery effect can also produce interactions between pulses, so repeat with smaller field and with a reference cell.

Because the material equation is linear, scaling both pulses by the same factor leaves the zero-crossing time unchanged. The delay-based measurement rejects a constant gain error, but an additive baseline shifts the crossing. An experimentally estimated timing uncertainty is the noise level divided by the local slope magnitude there. At zero polarisation within the negative pulse the equation gives $\dot{P} = -\epsilon_0 \Delta\chi E_0/\tau$, making the conversion from amplitude noise especially simple. This also explains why decreasing the field too far increases the timing noise even though it improves the validity of a linear material response.

To separate propagation from memory, take data through two cell thicknesses using the same cables and electronics. A phase accumulating in the material scales with thickness, while a common cable or detector lag cancels in their phase difference. Reflections give additional echoes whose times scale with thickness too, but follow round-trip distances and can interfere with the direct signal. Gate a sufficiently isolated first arrival or include interface scattering in the transfer function. Taking a ratio of Fourier spectra near a zero of the reference pulse magnifies noise; discarding or regularising those frequencies changes the reconstructed temporal resolution and should be stated with the result.

Compare absorption and phase near a resonance

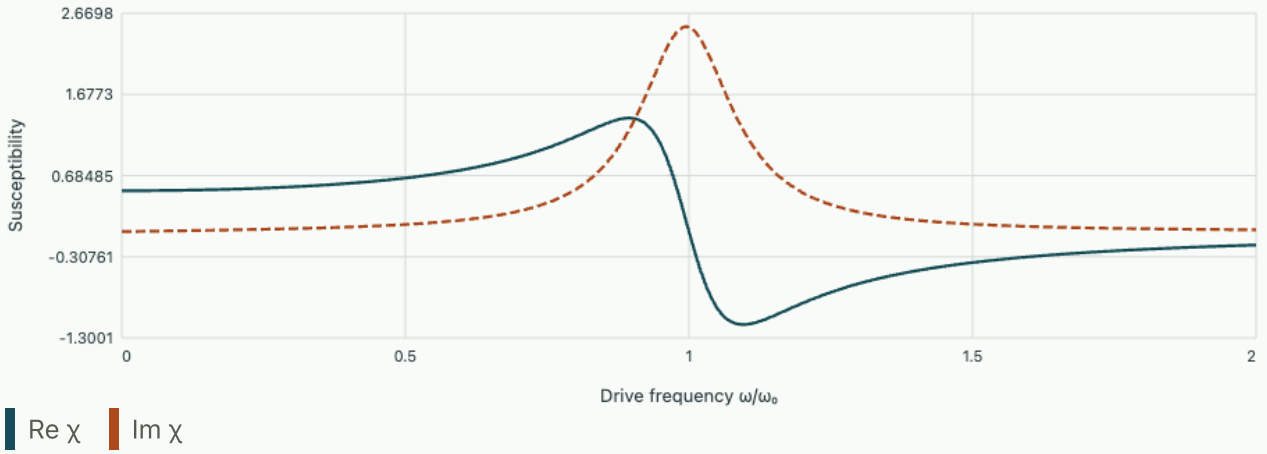
Sweep the drive through a material resonance. Follow the phase response and loss separately before interpreting a pulse arrival time.

Open this experiment online

Try the calculation

Damping γ/ω₀ **0.2**

Oscillator strength ωₚ²/ω₀² **0.5**



At ω = ω₀ the susceptibility is purely imaginary with Im χ = 2.5. At zero frequency the real susceptibility is 0.5.

Worked pulse problem · bound a group-delay measurement

A pulse traverses two samples cut from one transparent material, of lengths 10.0 and 30.0 mm. The measured difference between the carrier-envelope peak times is 103 ps. The extra material replaces air in a fixed total optical path. Estimate the group index. A student obtains a substantially different value from the phase velocity at the carrier. Specify two physical measurements and one numerical test that can resolve the disagreement.

▼ **Use the replaced air and examine the pulse spectrum**

The incremental delay is $(n_g - n_{g,\mathrm{air}})\Delta L/c$, with $n_{g,\mathrm{air}}$ near one in this idealised setting. Thus $n_g \simeq 1 + c(103\,\mathrm{ps})/(20.0\,\mathrm{mm}) = 2.54$. Treating the extra material path as an added path in an otherwise unchanged optical train would give 1.54 instead. A genuinely moved detector would require that different time reference and geometry. A 2 ps independent uncertainty in this differential delay alone contributes about 0.030 to n_g.

First measure phase at several closely spaced frequencies and unwrap its slope. The slope gives group delay, whereas phase at one frequency determines a phase index only after fixing an integer number of waves. Second, shorten the spectrum while keeping its central frequency fixed and compare peak delay with a cross-correlation delay and a centre-of-energy delay. A strongly distorted pulse can yield three different answers even with perfect electronics. Record a transmitted spectrum for each sample to establish whether selective attenuation is significant.

As a numerical test, propagate the measured incident spectrum using the fitted complex wavenumber and reconstruct the physical detector observable, including its impulse response. Compare with the analytic linear-phase translation after suppressing higher dispersion derivatives and frequency-dependent attenuation. Keep the computational time window large when narrowing the spectrum, since the pulse then extends further in time. A wrapped pulse in a periodic FFT window can acquire an apparently early peak. Looking at the first and last portions of the unwrapped time trace is an inexpensive way to rule that out.

A deliberately chirped input reveals the sign of dispersion

The temporal width of an initially transform-limited Gaussian after propagation depends on squared group-delay dispersion. Measuring one broadened width therefore loses the sign of the quadratic phase. Add a known adjustable quadratic spectral phase before the sample. For a Gaussian with fixed spectral intensity, the broadening now depends on the square of the sum of the imposed and sample phase curvatures. Plot measured pulse width versus the imposed curvature. The minimum identifies the setting that compensates the sample and hence gives the opposite of its signed dispersion.

For example, suppose independent calibration of the phase device assigns its minimum output width to an imposed minus four hundred femtoseconds squared after inserting a plate. If the optical train without the plate minimises at minus one hundred femtoseconds squared, the plate adds plus three hundred femtoseconds squared. The displacement of the minimum removes a common phase offset from the optical train. The widths at their respective minima need not be exactly equal if the plate changes the transmitted spectrum.

Measure that spectrum before interpreting a displaced or broadened minimum. A filter narrowing the spectrum increases even the best possible transform-limited pulse width. Cubic spectral phase and detector response may make the measured width curve asymmetric or leave a long tail near the nominal optimum. Comparing a reconstructed complete temporal profile with its spectrum gives information that a single width at one propagation distance cannot contain.

Tutorial · infer a material slope with a pair of delay measurements

Two samples of the same transparent glass have lengths 20.0 and 70.0 mm. With the receiver and external air path arranged so that the added glass replaces an equal vacuum length, the narrowband arrival time differs by 90.0 ps. At 1.00 µm the independently measured phase index is 1.500. Determine group index and the local index slope. Estimate how an unaccounted fixed electronic time offset affects a two-length measurement.

▼ **Use the difference to eliminate common delays**

With $\Delta L = 0.0500$ m and $\Delta t = (n_g - 1)\Delta L/c$, the inferred index is

$$n_g = 1 + c\Delta t/\Delta L = 1.540.$$

Therefore $dn/d\lambda_0 = (n - n_g)/\lambda_0 = -0.040$ per µm. A fixed additive trigger latency appears in both raw arrival times and cancels in the difference. A sample-dependent connector or surface coating that changes the pulse transmission phase need not cancel. Comparing a third length tests the assumed affine dependence of delay on material length, while measuring transmitted spectra can reveal frequency-dependent filtering.

Tutorial · distinguish dispersion from simple attenuation

An initially transform-limited Gaussian pulse has intensity FWHM 50 fs. After a transparent segment its width is 80 fs and its peak intensity is 0.50 times the original. Estimate $|k_2 L|$ under the Gaussian model, then infer the additional frequency-independent intensity transmission needed to account for the peak. Could the measured peak alone determine dispersion?

▼ **Recover a width and then an energy ratio**

The initial amplitude parameter is $T_0 = 50/(2\sqrt{\ln 2}) = 30.0$ fs. The width ratio 1.60 gives

$$|D| = \sqrt{1.60^2 - 1} = 1.249.$$

Thus $|k_2 L| \simeq 1.13 \times 10^3$ fs². At conserved energy the peak would be $1/1.60 = 0.625$, so an additional intensity transmission of $0.50/0.625 = 0.80$ reconciles both observations. Equivalently the energy ratio is peak ratio times width ratio, 0.80. A peak value alone admits many combinations of filtering, dispersion and attenuation. The inference additionally assumes that higher-order spectral phase and spectral amplitude filtering do not change the Gaussian shape.

David Tong, University of Cambridge · electromagnetism in matter, dispersion and causality

A guided mode, a cavity line and the energy they carry

Waves, radiation and relativity

A microwave pulse sent through a rectangular guide arrives late even though the guide is evacuated. Terminating the guide converts the travelling pattern into a family of resonances. Use the same transverse eigenproblem to predict its cutoff, pulse delay, cavity frequency and measured linewidth.

State the longitudinal dependence before choosing a transverse basis

Let a hollow rectangular perfect-conductor guide occupy $0 < x < a,\ 0 < y < b$ and extend in z. Fill it uniformly with a lossless dielectric. A single-frequency field varies as $e^{i\beta z - i\omega t}$. In the interior each Cartesian component satisfies the Helmholtz equation. Separating the z dependence leaves the transverse equation $\nabla_t^2 F + k_c^2 F = 0$, with $k_c^2 = k^2 - \beta^2,\ k = \omega\sqrt{\epsilon\mu}$. Conducting walls impose zero tangential E at the perimeter. They do not impose zero H there, since surface current can flow and supply the required magnetic discontinuity.

A transverse-magnetic (TM) mode has $H_z = 0$ and uses E_z as its longitudinal scalar. Because z lies tangential to every side wall, $E_z = 0$ there. The allowed nonzero separable solutions are sine-sine products with positive integers m,n. A transverse-electric (TE) mode has $E_z = 0$ and uses H_z with vanishing outward normal derivative on the side walls. Its scalar modes are cosine-cosine products. One index may be zero but both cannot be zero in this hollow single-conductor guide. The apparent constant H_z TE00 solution fails to reconstruct a nonzero propagating transverse field from Maxwell's equations. A TEM mode can exist in a coaxial line with two conductors maintained at different potentials. A connected perfect conductor surrounding empty simply connected space admits only constant electrostatic boundary potential, precluding a nontrivial transverse potential difference for TEM.

$$k_{c,mn}^2 = (m\pi/a)^2 + (n\pi/b)^2,$$

$$\beta^2 = \omega^2\epsilon\mu - k_{c,mn}^2,$$

$$f_{c,mn} = \frac{1}{2\sqrt{\epsilon\mu}}\sqrt{(m/a)^2 + (n/b)^2}.$$

The cutoff emerges from the room a transverse standing field needs to fit between walls. It is a condition on frequency, geometry and filling material together. Below cutoff a decaying longitudinal solution can still store near fields and mediate coupling over finite lengths. A reading on the far side of a short below-cutoff section therefore need not be exactly zero.

Reconstruct one measurable field from the vector equations

Take $a > b$ and the lowest TE10 mode. A useful choice is $\widetilde{E}_y = E_0 \sin(\pi x/a) e^{i\beta z}$, with $E_x = E_z = 0$. It vanishes at x walls as required. At the y walls E_y is normal and can end on surface charge. Use $\nabla \times \widetilde{\mathbf{E}} = i\omega\mu\widetilde{\mathbf{H}}$ to reconstruct both magnetic components:

$$\widetilde{H}_x = -\frac{\beta}{\omega\mu}\widetilde{E}_y,$$

$$\widetilde{H}_z = -\frac{i\pi E_0}{\omega\mu a}\cos(\pi x/a) e^{i\beta z}.$$

Ampère's law then reproduces the dispersion relation. The negative x direction of H for positive E_y and positive β gives positive S_z. The TE wave impedance is $Z_{\mathrm{TE}} = \omega\mu/\beta$ using transverse magnitudes. Unlike a plane-wave impedance in a homogeneous bulk medium, this one changes with frequency even when ϵ, μ are constant. At cutoff transverse magnetic H_x vanishes for a fixed finite E_y, while H_z remains finite. This is why large field amplitude near cutoff need not imply substantial transmitted axial power.

Integrate power through the cross-section using $\langle S_z \rangle = \beta |\widetilde{E}_y|^2/(2\omega\mu)$. The integral of $\sin^2(\pi x/a)$ is $a/2$. The mode therefore carries $P = ab\,|E_0|^2/(4Z_{\mathrm{TE}})$. Using simply the area times the central peak flux would overestimate power by a factor of two. A small receiving probe samples a spatially weighted field. Its calibration can be checked by moving it from the centre to x=a/6, where field amplitude halves and local axial flux falls by a factor of four.

Compute two velocities and check them by energy transport

With bulk phase speed $v = 1/\sqrt{\epsilon\mu}$, the dispersion relation is

$$\omega^2 = v^2(\beta^2 + k_c^2).$$

Taking the frequency-to-wavenumber ratio and its derivative gives

$$v_{\mathrm{ph}} = \frac{\omega}{\beta} = \frac{v}{\sqrt{1 - (\omega_c/\omega)^2}},$$

$$v_g = \frac{d\omega}{d\beta} = v\sqrt{1 - (\omega_c/\omega)^2},$$

$$v_{\mathrm{ph}} v_g = v^2.$$

Phase fronts can move faster than the bulk wave speed along the guide. They are surfaces of equal phase and carry no independent message. A narrowband unattenuated envelope travels at the smaller group speed. An energy calculation provides a second route to the same quantity. Insert all three TE10 field components into $\overline{u} = (\epsilon|\widetilde{E}|^2 + \mu|\widetilde{H}|^2)/4$, integrate over the cross-section, and use $\beta^2 + (\pi/a)^2 = \omega^2\epsilon\mu$. Electric and magnetic energies per length are each $\epsilon ab|E_0|^2/8$. Dividing P by total energy per length yields

$$\beta/(\omega\epsilon\mu) = v_g.$$

Keeping only the transverse magnetic component would spoil the equality.

For an evacuated guide $a = 30.0$ mm, $b = 12.0$ mm, $f_{c,10} = 4.997$ GHz. At 8.00 GHz, $\sqrt{1 - (f_c/f)^2} = 0.781$, $v_g = 2.34 \times 10^8$ m/s and guide wavelength $2\pi/\beta = 48.0$ mm. A 2 m section gives 8.54 ns group delay, compared with 6.67 ns in bulk vacuum. TE20 turns on at 9.99 GHz and TE01 at 12.49 GHz, so the chosen frequency is inside the single-propagating-mode band. With central electric peak 100 V/m, $Z_{\mathrm{TE}} \simeq 482\ \Omega$ and $P \simeq 1.87$ mW. Knowing frequency is essential to infer power from the same central probe in another band.

Close the guide and quantize its longitudinal phase

Add perfect end walls at z=0 and z=L. Tangential E_y must vanish at each end, so superpose equal counterpropagating modes into $\sin(p\pi z/L)$ with $p = 1, 2, \ldots$ for this TE10 family. The resonance frequencies become

$$f_{10p} = \frac{v}{2}\sqrt{a^{-2} + (p/L)^2}.$$

The common estimate $c/(2L)$ is the free spectral range for a nondispersive one-dimensional TEM cavity. It omits the transverse wavenumber in this example. For the guide above closed at $L = 60$ mm, $f_{101} = 5.59$ GHz and $f_{102} = 7.07$ GHz. Their separation is about 1.48 GHz, considerably smaller than the nominal 2.50 GHz TEM spacing. At sufficiently high longitudinal order the separation approaches $v/(2L)$, though other transverse families can then populate the spectrum.

The mode's field has nodes and antinodes in two coordinates. A tiny probe on an electric node may leave a resonance undetected, and moving a lossy probe can also shift its frequency and damping. Two nearby spectral peaks could be adjacent p indices, two different transverse modes or a degeneracy split by manufacturing errors. Compare measured spatial patterns and polarization selection rules before indexing solely by a frequency spacing.

Derive amplitude decay, stored-energy decay and linewidth

In practice wall conduction, dielectric loss and deliberate port coupling all remove stored energy. If a lightly damped mode retains the same shape as it decays, energy obeys $dU/dt = -\kappa U$. Then $U = U_0 e^{-\kappa t}$, field amplitude falls as $e^{-\kappa t/2}$, and loaded quality factor is

$$Q_L = \omega_0/\kappa.$$

Energy lost per cycle is approximately $2\pi U/Q_L$ for small fractional loss. The exact integrated energy lost over a cycle is $U[1 - e^{-2\pi/Q_L}]$. These expressions distinguish a rate in radians per second from linewidth in ordinary hertz.

Write a complex amplitude in a frame rotating at ω_0. Under a weak frequency-swept harmonic drive its steady amplitude is proportional to $[\,\kappa/2 - i(\omega - \omega_0)\,]^{-1}$. The power response is a Lorentzian. Its half-maximum points have $|\omega - \omega_0| = \kappa/2$, so the full power width is κ in angular frequency, or f_0/Q_L in hertz. A field-amplitude reading falls to $1/\sqrt{2}$ at those same frequencies. Half the field magnitude corresponds to a different bandwidth.

Consider $f_0 = 6.00$ GHz and $Q_L = 6000$. The power-line FWHM is 1.00 MHz, the energy e-folding time $1/\kappa = 159$ ns, and the field e-folding time 318 ns. Observe a field e-folding time of 159 ns instead, and the inferred Q would be 3000, even though the number printed on the oscilloscope is unchanged. Always determine whether an instrument's output voltage is proportional to field, instantaneous power or a filtered logarithm of either.

Separate a feed port from dissipative losses

Normalize a mode amplitude a so $|a|^2 = U$, and an incident port amplitude s_+ so $|s_+|^2 = P_{\mathrm{in}}$. In a simple one-port cavity the internal loss rate κ_i and external escape rate κ_e add. A consistent choice of port phase gives

$$\dot{a} = [i\Delta - (\kappa_i + \kappa_e)/2]a + \sqrt{\kappa_e}s_+,$$

$$s_- = s_+ - \sqrt{\kappa_e}a.$$

At steady on-resonance drive the reflected amplitude fraction is $(\kappa_i - \kappa_e)/(\kappa_i + \kappa_e)$. Exact cancellation at critical coupling is compatible with continued reflection when driving off resonance, and with finite resonant energy stored in the cavity. Once the drive is removed that energy decays through both channels.

With rates $\kappa_i/(2\pi) = 0.30$ MHz and $\kappa_e/(2\pi) = 0.70$ MHz, the resonant reflected power fraction is 0.16. The remaining 84% of the steady incoming power is dissipated internally in this one-port model. Interchanging the two rates preserves both the linewidth and reflected power at resonance but reverses the reflected phase. A calibrated coherent reflection or a second measurement after deliberately changing coupling can resolve this ambiguity.

Compare field and energy ring-down and linewidth at several loaded quality factors

Change guide width and follow cutoff and axial propagation

Use a small bead to map a cavity mode

A resonator offers a way to measure a field without placing a conventional field sensor at its maximum. Insert a small dielectric perturbation, record the change in resonance frequency, then scan its position. The derivation requires a closed, lossless cavity or a sufficiently high-Q nearly closed cavity for which one isolated mode is meaningful. Take fixed conducting walls, nondispersive permittivity and permeability, and a perturbation small enough to change the frequency and the shape smoothly. A nearby degeneracy requires a coupled-mode treatment.

In the source-free eigenproblem $\nabla \times \mu^{-1}\nabla \times \mathbf{E} = \omega^2\epsilon\mathbf{E}$, multiply by the complex conjugate of the unperturbed mode and integrate over the volume. Subtract the conjugate unperturbed equation multiplied by the perturbed mode. Integration by parts transfers the double curl with no surface contribution at a fixed ideal conducting wall. Expand ω^2 and $\epsilon + \delta\epsilon$ through first order. The unknown first-order change of mode shape cancels from the difference, leaving

$$\frac{\delta\omega}{\omega} = -\frac{1}{2}\frac{\int \delta\epsilon(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2\,dV}{\int \epsilon(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2\,dV}.$$

A positive permittivity increment lowers the resonance frequency. The denominator is proportional to total electric energy for a nondispersive mode, and the fractional shift is the electric-energy overlap with the perturbation. The result describes a small perturbation of the permittivity distribution, not automatically a high-contrast conducting bead. A high-index sphere changes its own internal field even when it occupies a tiny volume; its electrostatic polarizability can then replace the simple weak-contrast volume approximation. The distinction matters if the shift is used for absolute calibration.

Suppose a weak dielectric insert raises local permittivity by 10.0 percent and covers a region containing 0.40 percent of the unperturbed electric energy. At 3.00 GHz the predicted fractional shift is -0.00020, or -600 kHz. Moving it to a position where the electric amplitude is half as large gives a 150 kHz shift for the same orientation, volume and local background. A map of resonance displacement measures squared field magnitude, so by itself it cannot determine the sign of an oscillating electric component. Two neighbouring antinodes separated by a nodal plane can produce nearly equal shifts with opposite electric phases.

A mode with loaded $Q = 3000$ at 3.00 GHz has a full linewidth of about 1.00 MHz. Its centre can still be fitted more accurately than the linewidth if signal-to-noise ratio and frequency stability are adequate. Record an unloaded baseline between bead positions and keep the probe coupling fixed. A systematic drift of 50 kHz is a small part of a linewidth but already one third of the expected shift at the lower-field position. If two modes draw close when the bead is moved, fitting one Lorentzian over their sum can imitate a position-dependent loss as well as a frequency shift.

A finite below-cutoff length is a boundary problem

A guide section whose infinite-guide propagation constant is imaginary does not simply have zero field. It supports decaying and growing spatial solutions. In a semi-infinite region an outgoing bounded solution picks the decaying branch. A finite narrowed section joined to a second propagating guide needs both branches to satisfy the two interfaces. This is the electromagnetic version of transmitting through a classically forbidden interval in a one-dimensional wave equation. No energy flows out of an isolated purely decaying transverse mode of the uniform lossless section, yet a properly driven combination can carry power through the interval.

Write a scalar model of one isolated mode as $\psi'' + k^2\psi = 0$ in identical leads and $\psi'' - \kappa^2\psi = 0$ in the central interval of length d. Assume continuity of ψ and ψ' at both seams. This mimics a single matched constitutive parameter and deliberately neglects higher transverse modes excited at a real step. Propagation of the pair (ψ, ψ') across the barrier uses hyperbolic functions. Matching it to an incident plus reflected wave on the left and one transmitted wave on the right gives

$$T = \left[1 + \frac{(k^2 + \kappa^2)^2}{4k^2\kappa^2}\sinh^2(\kappa d)\right]^{-1}.$$

At $k = \kappa$ this simplifies to $T = \operatorname{sech}^2(\kappa d)$. For $\kappa = 40$ per metre and lengths 25 and 50 mm, the transmitted powers are 0.420 and 0.0707 times incident. Doubling length does not square the transmission, because the reflected wave and the coupling at both interfaces matter. At large κd an exponential remains, multiplied by a matching-dependent prefactor. Extrapolating a long-section straight line on a log-power plot to zero length should therefore not be interpreted as an additional physical loss in a zero-length guide.

For a real device use modal matching, or a numerical Maxwell solution retaining evanescent transverse modes at a discontinuity. Check convergence against the number of included modes at each step, not just the resolution along the uniform section. Measurements with several lengths can separate the uniform attenuation constant from entrance and exit effects. Measuring reflection as well as transmission discriminates a lossless cutoff barrier, where rejected energy returns to the source, from an absorber that dissipates it. Finally use a time-domain envelope narrow enough in frequency that the fraction above the cutoff is negligible. A small above-cutoff spectral tail can dominate a pulse recorded after a long barrier and give a misleading apparent decay length.

Infer a resonator timescale from its line

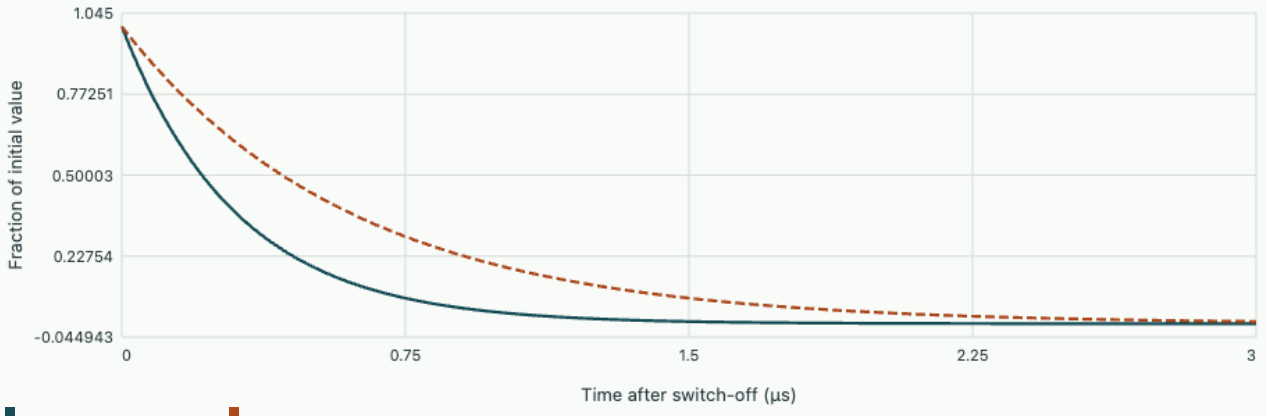
Change loss at a selected cavity mode and compare linewidth with energy decay. State whether each reported rate describes amplitude or energy.

Open this experiment online

Try the calculation

Loaded quality factor Q **6000**

Cavity length d (mm) **80**



The mode frequency is 3.123 GHz with full power linewidth 520 kHz. The energy e-folding time is 0.306 µs.

Worked resonance problem · a square guide with a slight width error

A nominally square air-filled guide of side 20.0 mm supports two orthogonal fundamental patterns with identical cutoff in the ideal square section. The fabricated horizontal dimension is 20.10 mm; the other remains 20.00 mm. Estimate the two cutoffs and the polarisation beating length at 10.0 GHz. How could a single scalar power measurement miss the effect?

▼ Follow the phases of the two modes

For the rectangular TE modes with one half-cycle across either side, the cutoffs are $c/(2a)$ and $c/(2b)$. These are 7.458 and 7.495 GHz, separated by about 37 MHz. At 10.0 GHz the longitudinal propagation constants are $k_{zj} = (2\pi/c)\sqrt{f^2 - f_{cj}^2}$. Their difference is about 0.88 per metre, corresponding to a full relative phase cycle over approximately 7.1 m. It takes only one quarter of this length to build a quarter-cycle retardance between equally excited components.

Near cutoff the propagation-constant sensitivity to a small width change is enhanced because the longitudinal wavenumber appears in the denominator of its derivative. Approaching the higher cutoff too closely also makes the interpretation vulnerable to wall loss, poorly matched launchers and higher-order evanescent contributions at the ends. Above both cutoffs in a sufficiently long uniform section, project the

exit field on a pair of known orthogonal receiving patterns and measure their coherent phase difference. Two power readings along the eigenaxes remain unchanged in the lossless model even while a diagonal analyser shows oscillating power.

Rotating the input launcher onto one eigenpattern should suppress the beating, since no second component then accumulates a relative phase. Repeating the experiment at several frequencies tests the predicted dispersion of $k_{z1} - k_{z2}$, distinguishing a distributed width mismatch from a nearly constant phase offset in two receiver cables. A slow twist of the guide can additionally couple the patterns instead of merely advancing their phases. In that situation first measure the full two-by-two complex transfer matrix; treating every disappearing component as attenuation would misidentify coherent mode conversion.

A vanishing resonance can be a coupling zero

A cavity spectrum measures modes through a particular feed and pickup. For a short aligned electric probe, coupling to a mode is proportional at leading order to the projection of the modal electric field onto the probe's current distribution. A probe at a nodal plane or oriented perpendicular to that field can therefore fail to display a well-defined cavity mode. Moving the receiver away from the node or rotating it can restore the spectral line without changing the cavity eigenfrequency.

As a quantitative illustration consider a mode whose electric field varies as a sine across the width. Move a small probe from one quarter of that width to one twelfth. The ratio of its squared coupling magnitudes is approximately 0.134. With an unchanged weak drive and an effectively nonloading pickup, the transmitted power in that mode can shrink by almost an order of magnitude. If the same port also substantially loads the cavity, its movement changes the linewidth and the resonant intracavity field as well.

A mapping experiment should record peak frequency, linewidth and integrated spectral response alongside peak power. A frequency shift with little change in linewidth supports a predominantly reactive perturbation, whereas linewidth changes signal modified damping or port coupling. A changing peak alone cannot distinguish a weak field region, a rotated polarisation and an altered loaded resonance. Combining two feed orientations with a weak scanning pickup supplies a more informative field map.

Tutorial · infer guide width from a delay without assuming a vacuum transit

A pulse at 9.00 GHz traverses 1.50 m of evacuated rectangular guide in 6.00 ns after cable delays are removed. Assume a narrowband TE10 packet with negligible loss and $b = 10.0$ mm. Infer its cutoff and guide width, check the next two candidate cutoffs and estimate its modal impedance.

▼ **Invert the dispersion and verify the mode assignment**

The group speed is 2.50×10^{8} m/s, or 0.834 of c. Hence $f_c = f\sqrt{1 - (v_g/c)^2} \simeq 4.97$ GHz and $a = c/(2f_c) \simeq 30.2$ mm. TE20 cuts on at about 9.94 GHz and TE01 at 15.0 GHz, both above the test. The TE impedance is $Z_0/(v_g/c) \simeq 452\ \Omega$. A broad pulse overlapping 9.94 GHz would require a mode-launch and dispersion analysis before interpreting one transit time with this inversion. Measuring delay at a second centre frequency tests the inferred cutoff without another width parameter.

Tutorial · separate an internal Q from a port Q

A one-port cavity has $f_0 = 5.00$ GHz and measured power linewidth 0.80 MHz. Halving the port coupling rate with negligible resonance shift reduces linewidth to 0.60 MHz. Infer both initial rates in hertz, both quality factors, the initial resonant reflected fraction and the energy lifetime after the coupling change.

▼ **Combine the rate sums from both tests**

Let $g_i = \kappa_i/(2\pi)$ and $g_e = \kappa_e/(2\pi)$. Solve $g_i + g_e = 0.80,\ g_i + g_e/2 = 0.60$ MHz to get $g_e = g_i = 0.40$ MHz. Both initial individual quality factors are $5.00\,\mathrm{GHz}/0.40\,\mathrm{MHz} = 12500$, and initial loaded Q is 6250. Initial reflection vanishes by critical coupling. After the change the resonant reflected amplitude is 1/3 and reflected power 1/9. The energy lifetime is

$$1/(2\pi\,0.60\,\mathrm{MHz}) = 265\,\mathrm{ns}.$$

This result relies on a fixed internal loss. A coupler adjustment that introduces dielectric loss would invalidate the two-equation inference.

David Staelin, MIT 6.013 · chapter 9 on waves guided by planar and rectangular structures

When a periodic dielectric has no propagating Bloch wave

Fields and the response of matter

Derive a band gap using two lossless transparent materials and distinguish evanescent Bloch propagation from absorption of optical power.

Construct a unit cell whose waves interfere

Stack two transparent, lossless, nonmagnetic dielectrics periodically in one dimension. Their refractive indices are n_1, n_2, their physical thicknesses d_1, d_2, and the period $a = d_1 + d_2$. Work at normal incidence, so polarisation directions are equivalent for isotropic layers. Each separate material admits a propagating wave at the chosen optical frequency, yet the periodic stack may not. The effect comes from coherent interface reflections, not from a negative refractive index or an absorptive constituent.

Describe the local field by its tangential electric and magnetic components. Propagation through one homogeneous layer connects them by a 2×2 transfer matrix with diagonal $\cos\phi_j$, off-diagonal terms proportional to $iZ_j \sin\phi_j$ and $i\sin\phi_j / Z_j$, where $\phi_j = \omega n_j d_j / c$ and $Z_j = Z_0 / n_j$. Changing the chosen propagation direction or time convention may reverse the two imaginary signs together. The real trace and unit determinant of the cell matrix below are unaffected.

Read the dispersion relation from the trace

At an interface both tangential components are continuous in the absence of imposed sheet currents. Multiply the two propagation matrices to form one cell. Its determinant is one because each layer's determinant is

$$\cos^2\phi + \sin^2\phi = 1.$$

Thus the two eigenvalues are reciprocal. For a Bloch mode they are $e^{\pm i k_B a}$ on a pass band. Taking half the trace gives

$$\cos(k_B a) = \cos\phi_1 \cos\phi_2 - \frac{1}{2}\left(\frac{n_1}{n_2} + \frac{n_2}{n_1}\right)\sin\phi_1 \sin\phi_2.$$

When the right side lies between –1 and 1 a real Bloch phase exists. If its magnitude exceeds one, a decaying and a growing Bloch factor replace a pair of unit-modulus factors. For a bounded excitation of a semi-infinite stack choose the decaying direction. The attenuation per cell is $\operatorname{arcosh}\left|\operatorname{trace}(M)/2\right|$. A lossless semi-infinite crystal in a stop band reflects incident power. Spatial attenuation of a Bloch amplitude in this region is compatible with zero volumetric heat production.

Obtain a quantitative stop-band width

Choose $n_1 d_1 = n_2 d_2 = \lambda_0/4$. Writing $s = \omega/\omega_0$, both layer phases are $\phi = \pi s/2$. At $s = 1$ the half trace is $-\frac{1}{2}(n_2/n_1 + n_1/n_2)$. Recognising $\cosh(\log r) = (r + r^{-1})/2$ gives centre attenuation $|\log(n_2/n_1)|$ per cell. For $n_1 = 1.5, n_2 = 2.0$ this is 0.2877, so a decaying eigencomponent has amplitude factor 0.75 per cell. This statement about an eigencomponent should not be confused with the exact transmittance of a specific finite stack, which also depends on entrance and exit impedances.

The first gap edges obey half trace −1. With $\cos^2\phi = 1 - \sin^2\phi$, solve for the edge sine or cosine. The two edge phases are symmetric about $\pi/2$ and satisfy

$$|\cos\phi| = |n_2 - n_1|/(n_1 + n_2).$$

Hence the fractional angular-frequency width is $(4/\pi)\arcsin(|n_2 - n_1|/(n_1 + n_2))$, about 0.183 for the example. The lowest gap width tends continuously to zero as index contrast vanishes. A gap in a plotted numerical spectrum at $n_1 = n_2$ would indicate an implementation or plotting problem.

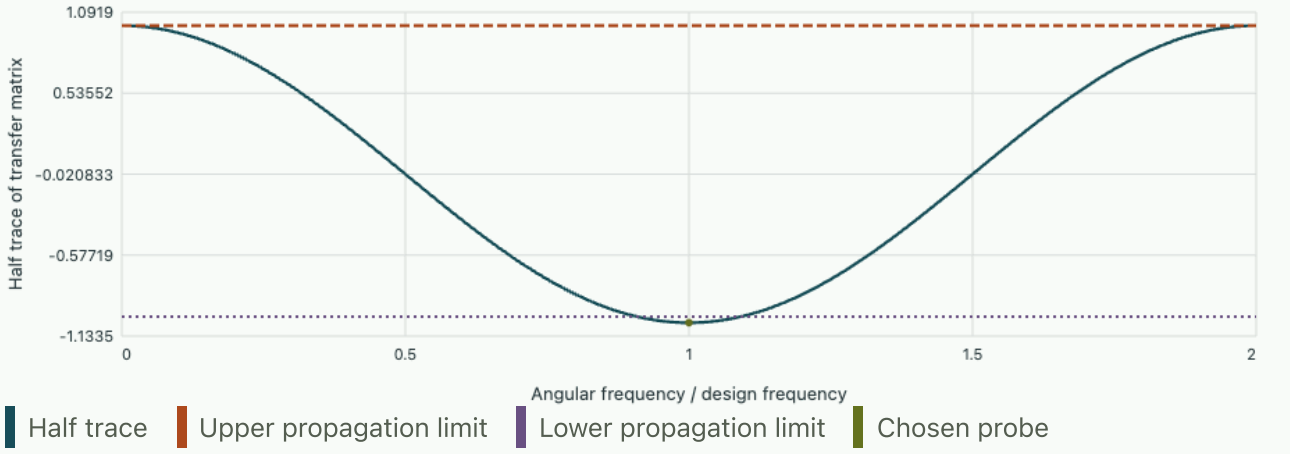
Compare a frequency scan with a finite mirror

Open the live experiment

Try the calculation

High index n_2 (n_1 = 1.5) 2

Probe frequency / design frequency 1



First gap width / design frequency 0.18252. Probe half trace -1.0417. Attenuation per cell 0.28768.

The graph plots half the transfer-matrix trace against normalised frequency and marks the admissible interval for a real Bloch phase. Change the index contrast first, keeping the layers quarter wave at the same reference frequency. The physical layer thickness adjusts in this family; it is a different experiment from changing a material while manufacturing all thicknesses fixed. Use the frequency control to read an attenuation exponent at the centre and just outside the gap.

At low frequency a wavelength spans many cells and a homogenised description can predict the long-wave speed. At the first gap the field changes phase by order one across a cell, exactly where the homogenisation scale separation fails. Conversely, an electromagnetic calculation showing multiple bands at very low frequency should be checked against a uniform-medium or long-wavelength effective-index limit. Compare a finite stack's reflection with the Bragg-mirror experiment. The bulk trace determines whether a propagating state can exist. A finite sample adds interference between boundaries and possible surface or defect effects.

Tutorial

If the two refractive indices become equal, simplify the trace formula for arbitrary thicknesses.

▼ **Work through the solution**

The impedance factor becomes one, so the half trace is $\cos\phi_1\cos\phi_2 - \sin\phi_1\sin\phi_2 = \cos(\phi_1+\phi_2)$. It never exceeds unit magnitude. Labelling a uniform medium by an artificial period folds the dispersion into zones but does not create a physical gap.

At the central quarter-wave frequency for index ratio 2, how many periods reduce a decaying Bloch component's amplitude by a factor of 100?

▼ **Work through the solution**

The amplitude factor per cell is

$$e^{-\log 2} = 1/2.$$

Require $2^{-N} \le 0.01$, giving $N \ge 7$. A prediction for transmitted optical power also needs termination and impedance matching.

A measured stack exhibits both a transmission minimum and measurable heating. Can the minimum be attributed entirely to a band gap?

▼ **Work through the solution**

A lossless band gap can reflect power without heating. Measure reflected and transmitted flux under consistent normalisation, then infer absorption from $1 - R - T$ after accounting for diffuse scattering and other exit channels. The bulk stop-band model and material absorption can coexist.

Compare with band calculations

MIT's mathematical nanophotonics course computes full periodic Maxwell modes. The one-dimensional trace above is a small benchmark with known pass and stop boundaries that a numerical eigenproblem should reproduce.

Why rotating a layered dielectric changes a capacitance measurement

Fields and the response of matter

Derive the two principal permittivities of a fine laminate, rotate the constitutive tensor and distinguish the applied electrode condition from an assumed field direction.

Compare the same material in two orientations

Build a specimen by repeating layers with scalar permittivities $\epsilon_1, \epsilon_2 > 0$, occupying fractions f and $1 - f$. The period is much smaller than both the specimen scale and the spatial variation of the applied field. Suppose both materials are linear, free of mobile charge and effectively lossless in the measurement band. An observer using a large capacitor wants one law relating averaged displacement and electric field. A scalar average cannot describe both arrangements of the same layers relative to the plates.

Place the field tangential to all internal interfaces first. A narrow rectangular electrostatic loop straddling an interface gives continuity of tangential E. Both layers see the same tangential field, but their displacements differ in proportion to permittivity. Averaging gives

$$\overline{D}_t = [f\epsilon_1 + (1 - f)\epsilon_2]\overline{E}_t.$$

The tangent-direction effective permittivity is an arithmetic mean. This is a parallel connection of capacitances at shared voltage in the special capacitor geometry.

Apply a purely normal displacement instead. No free charge lies at the internal interfaces, so their normal displacement is continuous. Normal fields differ as $E_i = D_n/\epsilon_i$. Averaging the voltage drops gives

$$\overline{E}_n = [f/\epsilon_1 + (1 - f)/\epsilon_2]\overline{D}_n,$$

with a harmonic-mean effective permittivity. The local continuity condition changes which field is shared between constituents and thereby changes the relevant averaging operation.

Check the amount of anisotropy from field energy

For $\epsilon_1 = 2\epsilon_0, \epsilon_2 = 10\epsilon_0$ at equal fractions,

$$\epsilon_t = 6\epsilon_0,$$
$$\epsilon_n = (10/3)\epsilon_0.$$

The arithmetic-to-harmonic ratio is 1.8. A constituent does not need to be crystalline or anisotropic for the composite to show distinct principal responses. The geometry of the internal boundaries is enough.

Average the energy density for the tangential experiment,

$$u = f\epsilon_1 E_t^2/2 + (1 - f)\epsilon_2 E_t^2/2 = \epsilon_t E_t^2/2.$$

For the normal experiment,

$$u = D_n^2[f/\epsilon_1 + (1 - f)/\epsilon_2]/2 = D_n^2/(2\epsilon_n).$$

These expressions supply a check independent of the capacitor-circuit analogy and identify which energy should be used when field or free charge is controlled. Energy positivity follows directly from the positive constituent permittivities. It does not require guessing an effective law from the largest measured component.

Rotate the law, then specify what the electrodes control

Use principal axes $\hat{n}, \hat{t}$ and suppose a macro-field makes angle θ to the layer normal. If $\overline{\mathbf{E}} = E(\cos\theta\,\hat{n} + \sin\theta\,\hat{t})$, then $\overline{\mathbf{D}} = E(\epsilon_n \cos\theta\,\hat{n} + \epsilon_t \sin\theta\,\hat{t})$. Project along the imposed field and divide by E:

$$\epsilon_E(\theta) = \epsilon_n \cos^2\theta + \epsilon_t \sin^2\theta.$$

The perpendicular displacement is $E(\epsilon_t - \epsilon_n)\sin\theta\cos\theta$. It is generally nonzero. A device that instead fixes the direction of averaged displacement while allowing a transverse field measures $\epsilon_D(\theta) = [\cos^2\theta/\epsilon_n + \sin^2\theta/\epsilon_t]^{-1}$. These two directional responses correspond to different constraints. They agree at principal axes and for isotropic material. A pair of unspecified “angled capacitor” measurements cannot be interpreted uniquely without describing the contacts and edge conditions.

At 45° in the example, the field-controlled projection gives about $4.667\epsilon_0$, while the displacement-controlled projection gives $4.286\epsilon_0$. These are both legitimate scalar ratios from one symmetric positive-definite tensor. Neither supplies a universal scalar permittivity for the rotated material. The complete vector response also predicts displacement tilt and field energy for any orientation.

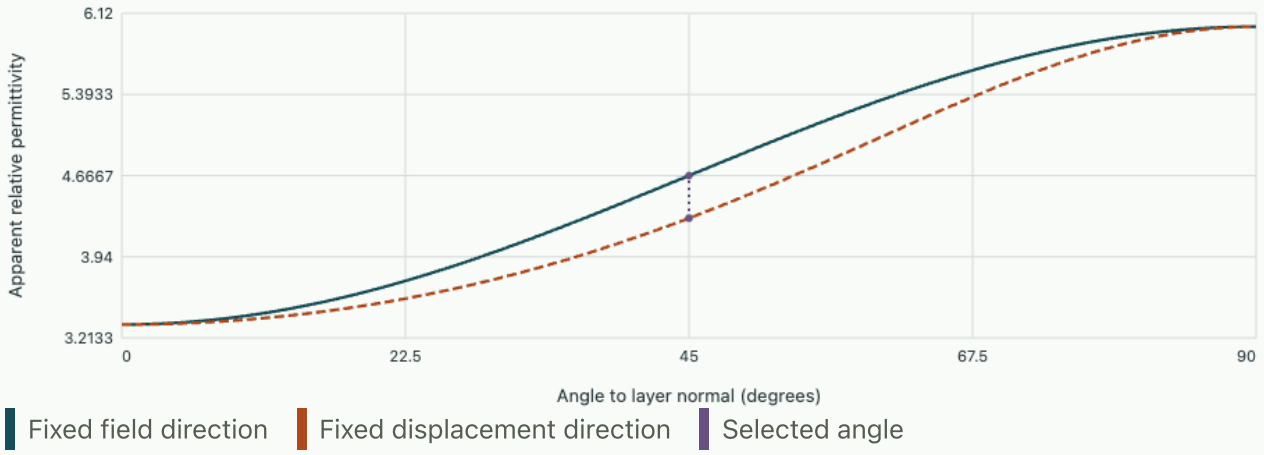
Test the measurement against the unit-cell model

Open the live experiment

Try the calculation

Fraction of low-permittivity phase **0.5**

Measurement angle (degrees) **45**



Principal relative permittivities: normal 3.3333, tangential 6. At 45° the two responses are 4.6667 and 4.2857.

Change the volume fraction while keeping the constituent values fixed. At a pure phase all orientations and both control conditions must collapse to a single value. At intermediate fractions, find the angle where the transverse component would be largest. It occurs where $|\sin\theta\cos\theta|$ is largest, but its magnitude also reflects the material contrast. Now reduce the contrast mentally to one and predict all curves without recomputing a single cell.

A numerical field model offers an independent homogenisation experiment. Use periodic tiling in the material directions and impose a macroscopic voltage difference plus periodic fluctuations on opposite faces. For a unit imposed field along each coordinate direction, average the resulting displacement over the cell. The resulting vectors supply the tensor columns. A grounded box on every side is a different boundary problem and can suppress the tangential response. Increase the number of material periods at fixed specimen scale and refine the individual period separately. These two limits test homogenisation and numerical representation, respectively.

How far can the effective law travel?

Electrostatic homogenisation assumes the cell fields respond quasistatically. In a wave experiment near the Bragg gap, phase changes substantially across one period and a local constant tensor is inadequate. A finely layered capacitor and a dielectric mirror can be made from the same constituents while requiring different models at different frequencies. Test the Bloch spectrum to identify the regime where layering begins selecting propagation bands. A low-frequency effective index alone cannot predict that gap.

Large spatial contrast also places a practical demand on mesh and interface handling. A normal gradient can jump strongly even when potential and displacement are continuous. A smooth-looking interpolated voltage surface is weak evidence for an accurate energy density. Compare constituent fields, total flux and energy consistency before accepting the extracted effective tensor.

Tutorial

Show that $\epsilon_t - \epsilon_n$ is nonnegative for every positive pair of constituents without appealing to a named mean inequality.

▼ **Work through the solution**

Put both terms over the positive denominator $f\epsilon_2 + (1-f)\epsilon_1$. The numerator after cancellation is $f(1-f)(\epsilon_1 - \epsilon_2)^2$, so the difference is nonnegative and vanishes for identical materials or a pure phase.

Equal fractions give measured relative values 6 tangentially and $10/3$ normally. Can you infer both constituents if the laminate assumptions hold?

▼ **Work through the solution**

Their sum is 12. The harmonic mean is $2ab/(a+b) = 10/3$, giving product 20. The positive roots of $x^2 - 12x + 20 = 0$ are 2 and 10. The data do not identify which label belongs to which phase without another observation.

Why can a constituent occupy very little volume yet strongly affect the normal response?

▼ **Work through the solution**

Every normal field line in this lamination crosses each layer. The inverse response weights f/ϵ; even a thin, low-permittivity layer can contribute a large voltage drop. A field along the layers distributes displacement differently and may be much less sensitive to that same phase.

How a magnetic field rotates the polarisation of transmitted light

Fields and the response of matter

Resolve a linearly polarised wave into circular eigenmodes, infer a refractive-index difference from an analyser and distinguish nonreciprocal rotation from a simple change of reference axes.

Derive rotation from propagation rather than drawing an arrow

Let monochromatic light propagate along $+z$ through a transparent magneto-optic sample, with an externally prepared axial magnetic field. Define fixed laboratory vectors $\mathbf{e}_\pm = (\hat{x} \pm i\hat{y})/\sqrt{2}$ and use phasors $e^{i(kz-\omega t)}$. These definitions specify the handedness algebraically. Calling a vector simply “right circular” without stating the viewing and time conventions can reverse a reported rotation sign.

An input linearly polarised in x is $(\mathbf{e}_+ + \mathbf{e}_-)/\sqrt{2}$. Suppose the medium supports wave numbers $k_\pm = 2\pi n_\pm/\lambda_0$ for the two bases, with negligible absorption. After a distance L, factor out the common phase $e^{i(k_+ + k_-)L/2}$. The remaining Cartesian phasor is $\hat{x}\cos\Delta - \hat{y}\sin\Delta$, where $\Delta = (k_+ - k_-)L/2$. Thus the polarisation angle in the fixed x, y plane is

$$\theta = \frac{\pi L}{\lambda_0}(n_- - n_+).$$

Equal transmissions of the two components preserve a linearly polarised output. Unequal absorption makes the circular amplitudes unequal and generally produces an ellipse. Rotating a linear analyser then finds a finite minimum, even for a perfectly coherent field. Residual intensity at the nominal extinction angle can also arise from stray light or imperfect polariser extinction, which must be measured independently.

Derive a magneto-optic eigenbasis

For axial magnetisation, a simple lossless transverse relative-permittivity block is $\left(\begin{smallmatrix} \epsilon & ig \\ -ig & \epsilon \end{smallmatrix}\right)$ with real g at the frequency of observation. Acting on $(1, i)$ gives $(\epsilon - g)(1, i)$; acting on $(1, -i)$ gives $(\epsilon + g)(1, -i)$. For relative permeability near one, $n_+^2 = \epsilon - g$, $n_-^2 = \epsilon + g$. At small g their difference $n_- - n_+ \simeq g/\sqrt{\epsilon}$. Reversing magnetisation reverses the sign of the off-diagonal coefficient and interchanges the two eigenvalues. Both the existence of circular modes and the sign-reversal test now follow from a constitutive model.

If $n_- - n_+$ is proportional to magnetic field over a specified range, write $\theta = VBL$, with Verdet constant V in rad/(T m). This is a measurable material and wavelength dependent coefficient. It is not determined by the mean refractive index alone. A strong nearby absorption line can give substantial frequency dependence and dichroism, invalidating a lossless narrow-band extrapolation.

Work through an analyser calibration

For $\lambda_0 = 600$ nm, $L = 0.20$ mm and measured index difference 5.0×10^{-4}, the rotation is $\pi(2.0 \times 10^{-4})(5.0 \times 10^{-4})/(6.0 \times 10^{-7}) = \pi/6$, or 30°. An ideal analyser at angle α transmits $I/I_{\text{in}} = \cos^2(\alpha - \theta)$. Its maxima occur at 30° and 210°; an intensity measurement determines a linear

polarisation angle modulo 180°. For a sample thick enough to rotate several half turns, an intensity maximum alone cannot give an unwrapped θ. Vary field or path length continuously from a known reference.

If the applied field is independently 0.40 T in that example, $V = \theta/(BL) \simeq 6.54 \times 10^3$ rad/(T m). The arithmetic is dimensionally correct, but the large value warrants a physical check of thickness, field calibration, units and material spectrum before reporting it as an ordinary glass coefficient. Degrees substituted for radians would make the inferred material coefficient too large by $180/\pi$.

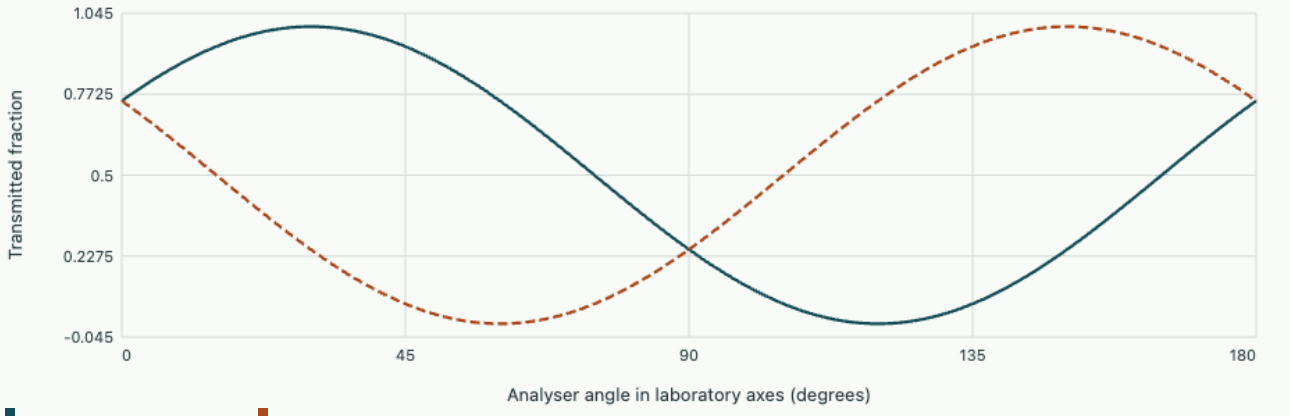
Use field reversal to reject alignment errors

Open the live experiment

Try the calculation

Index difference (10^{-4}) **5**

Sample length (mm) **0.2**



Unwrapped single-pass rotation 30°. At a 45° analyser the two transmitted fractions are 0.93301 and 0.066987.

The experiment shows the predicted analyser traces for opposite index splitting, with the same input polariser. Increase sample thickness to see why tracking the maximum by one measurement can select the wrong branch. Reversing the splitting reverses the tilt in fixed laboratory axes. A fixed analyser offset shifts both inferred angles equally and cancels in half their difference, provided the branch has been unwrapped consistently.

For small rotation, an analyser exactly at a crossed position transmits intensity proportional to θ^2. Its response has no sign sensitivity close to zero. Set the analyser at 45° instead:

$$\cos^2(\pi/4 - \theta) = [1 + \sin(2\theta)]/2 \simeq 1/2 + \theta.$$

The difference between fields $+B$ and $-B$ then measures a signed signal linear in small rotation. This trades extinction of the background for sensitivity to sign. Balance the two orthogonal analyser outputs if source intensity fluctuates appreciably during a sweep.

Ask what happens on a return pass

For the ideal Faraday medium at fixed magnetisation, the circular eigenmodes relative to a fixed laboratory basis are determined by its transverse material tensor. On propagation back through the sample, the accumulated positive propagation path again distinguishes the same laboratory circular basis after accounting for the reflection boundary. With a suitable simple mirror at normal incidence, the resulting material rotation doubles on a round trip in fixed transverse coordinates. In a reciprocal optically active material, the eigenvalue assignment with respect to the laboratory basis reverses on reversing propagation, and the single-pass rotation cancels on the ideal return. Write both transformations in one fixed basis before comparing signs.

A real double-pass setup includes mirror phase shifts, stress birefringence of windows and possible ellipticity. Reversing magnetic field, removing the sample and recording both analyser quadratures provide separate checks. A nonzero rotation in one transmitted beam proves only that its polarisation changed; establishing nonreciprocity requires comparing a reversed propagation experiment with a stated polarisation convention.

Tutorial

Two analyser maxima occur at 47° and 39° for opposite weak fields. If the response is linear and the same branch is tracked, what field-induced rotation and static offset do they imply?

▼ **Work through the solution**

Half the difference gives 4 degrees of signed rotation for the first field relative to zero. The mean is 43 degrees, the combined input polarisation and analyser zero. Conversion to a Verdet constant additionally needs the applied field and path length.

In the 600 nm example, what thickness uncertainty corresponds to 1 degree of rotation uncertainty at fixed index difference?

▼ **Work through the solution**

Rotation is proportional to thickness. Relative uncertainty $1/30$ therefore gives about $0.20/30 = 0.0067$ mm, or 6.7 µm. Uncertainty in wavelength or field would add additional contributions.

An ideal analyser gives a minimum 4% of its maximum. Can this be explained solely by a rotation of an ideal linearly polarised monochromatic input in a lossless circularly birefringent sample?

▼ **Work through the solution**

Pure rotation preserves a linear state, so an ideal analyser should extinguish it completely. Differential loss, ellipticity, depolarisation, finite bandwidth or a detector background are possible alternatives. Use circular-analysis or full Stokes measurements and an empty-path baseline to distinguish them.

Compare with a full polarisation measurement

Use the Stokes reconstruction experiment to determine which intensities reveal a circular component. A fitted linear rotation alone cannot measure that missing component.

What a detector samples

Compare radiation, scattering and the measured diffraction pattern. Examine how aperture sampling, transfer calibration and finite coherence alter the observable.

1. Radiating multipoles and what a detector actually integrates
2. Scattering, extinction and the phase of the forward wave
3. Diffraction from the near field to the focal plane
4. Measure the phase speed of a numerical electromagnetic wave
5. Separate diffraction from a finite observation window
6. Estimate a cavity linewidth from a finite, biased record
7. Reconstructing a partially polarised electromagnetic wave
8. Two polarisations moving through a uniaxial crystal
9. What a momentum eigenstate means in a finite experiment
10. From a finite pulse to a continuum transition rate
11. Which part of a Bell correlation survives imperfect state preparation?
12. Recover a transition frequency from two separated pulses

Radiating multipoles and what a detector actually integrates

Waves, radiation and relativity

A small oscillating source looks point-like from a distant detector, yet its angular pattern reveals its structure. Derive electric and magnetic dipole radiation from a conserved current, identify what survives when electric dipoles cancel, and turn a detector aperture into a calibrated estimate of total power.

Name the small parameters before discarding a part of the field

Let d be the characteristic source size, r the source-to-detector distance, and $k = \omega/c$. The condition $d/r \ll 1$ permits geometric expansions of the radial amplitude. The condition $kr \gg 1$ selects the radiation field over induction and quasistatic fields. The condition $kd \ll 1$ permits a multipole expansion of the source's directional phase. These inequalities concern different pairs of lengths. A source can be electrically small while a probe sits in its near field. A large optical aperture can be many wavelengths wide while its far field is measurable at a sufficiently large r, though it then requires many multipoles or a full Fourier integral.

For harmonic localized sources in vacuum the outgoing far-zone vector potential is proportional to $\mathbf{I}(\mathbf{n}) = \int \widetilde{\mathbf{J}}(\mathbf{r}')e^{-ik\mathbf{n}\cdot\mathbf{r}'}d^3r'$. Differentiating the rapidly varying e^{ikr}/r factor gives the magnetic field; using the radiation plane-wave relationship gives the electric field. This removes the component of $\mathbf{I}$ parallel to the observation direction:

$$\widetilde{\mathbf{E}}_{\mathrm{rad}} = \frac{i\omega\mu_0 e^{ikr}}{4\pi r}\,[\mathbf{I} - \mathbf{n}(\mathbf{n}\cdot\mathbf{I})],$$

$$\widetilde{\mathbf{B}}_{\mathrm{rad}} = \frac{1}{c}\mathbf{n}\times\widetilde{\mathbf{E}}_{\mathrm{rad}}.$$

The transverse projection is essential. It forces zero electric-dipole radiation along the dipole axis. Treating the current integral as an already transverse scalar misses this geometrical zero and produces the wrong total power.

Eliminate an integrated current with charge conservation

The electric dipole moment is

$$\mathbf{p} = \int \mathbf{r}'\rho\, d^3r'.$$

Differentiate this definition and use $\dot\rho = -\nabla'\cdot\mathbf{J}$. For component j, integration by parts changes $-\int x'_j\partial'_l J_l$ into $\int J_j$, since the surface term vanishes outside a localized current. Thus $\int\widetilde{\mathbf{J}} = -i\omega\widetilde{\mathbf{p}}$. For $kd \ll 1$ the leading term of the phase expansion is one. Inserting it gives

$$\widetilde{\mathbf{E}}_p = \frac{k^2 e^{ikr}}{4\pi\epsilon_0 r}[\mathbf{p} - \mathbf{n}(\mathbf{n}\cdot\mathbf{p})],$$

$$\frac{d\overline{P}}{d\Omega} = \frac{\omega^4|\mathbf{n}\times\mathbf{p}|^2}{32\pi^2\epsilon_0 c^3}.$$

The phasor p is a peak moment. For a linearly oscillating real-axis dipole along z the intensity follows $\sin^2\theta$, independent of azimuth. Integrating $\sin^2\theta\sin\theta\,d\theta d\phi$ gives $8\pi/3$. Hence the cycle-averaged total is

$\overline{P} = \omega^4 |\mathbf{p}|^2 / (12\pi\epsilon_0 c^3)$. A complex vector p describing an elliptical source is allowed in the vector formula. With equal orthogonal components in temporal quadrature, radiation along the normal to their plane is circularly polarized and there is no direction parallel to both components at once.

Consider an effective moment $p_0 = 1.0 \times 10^{-12}$ C m at 100 MHz. A vacuum far-field receiver at 10 m on the equator sees peak $E_0 = k^2 p_0/(4\pi\epsilon_0 r) = 3.95$ mV/m. Mean flux is $E_0^2/(2Z_0) = 20.7$ nW/m². Total power is $(8\pi/3) r^2 S_{\mathrm{eq}} = 17.4$ µW. At the same distance but a 30° polar angle the field peak halves and flux falls to a quarter. Increasing frequency while holding moment fixed raises power as the fourth power. Maintaining that same moment with a specified driven current introduces its own frequency dependence, so a frequency sweep must state the held source variable.

Recover the magnetic dipole from the antisymmetric current moment

Expand the phase as $1 - ikn_j x'_j + \cdots$. The first directional correction involves $M_{ij} = \int x'_j J_i \, d^3r'$. Its symmetric and antisymmetric parts have distinct origins. Define $\mathbf{m} = \frac{1}{2}\int \mathbf{r}' \times \mathbf{J} \, d^3r'$. Then the antisymmetric contraction with n gives a current contribution $ik\mathbf{n} \times \mathbf{m}$ for the index convention above. Substituting into the far-field electric expression yields

$$\widetilde{\mathbf{E}}_m = -\frac{Z_0 k^2 e^{ikr}}{4\pi r} \mathbf{n} \times \mathbf{m},$$

$$\overline{P}_m = \frac{\mu_0 \omega^4 |\mathbf{m}|^2}{12\pi c^3}.$$

A small planar current loop has $\mathbf{m} = I\,\mathbf{A}$ with oriented area vector. For a loop in the xy plane its magnetic-dipole intensity has the same $\sin^2\theta$ dependence as a z electric dipole, while its electric polarization is azimuthal instead of meridional. An intensity detector without polarization analysis can therefore confuse two fundamentally different sources. A rotating linear analyser placed at the equator can distinguish them even if each delivers the same incident power.

For a loop of radius 10 mm carrying a 10 mA peak sinusoidal current at 100 MHz, $m_0 = 3.14 \times 10^{-6}$ A m² and the equivalent effective electric-moment magnitude for equal power is $m_0/c = 1.05 \times 10^{-14}$ C m. Its power is about $17.4\,\mu\mathrm{W}(0.0105)^2 = 1.91$ nW. The current and geometry are assumed uniform around the loop. Larger loops require retaining retardation along the current path and a feed model before scaling this result.

Keep the symmetric part when the electric dipole vanishes

Define $D_{ij} = \int \rho x'_i x'_j \, d^3r'$. One integration by parts gives

$$\dot{D}_{ij} = \int (x'_i J_j + x'_j J_i) d^3r'.$$

Therefore $M_{(ij)} = -i\omega D_{ij}/2$ for harmonic sources. Split D into trace and traceless parts, defining $Q_{ij} = 3D_{ij} - \delta_{ij}\,\mathrm{tr}\, D$. The trace contribution after contraction is parallel to n and vanishes under transverse projection. Relative to the electric dipole field the effective transverse source is consequently

$$\widetilde{\mathbf{E}}_{p,Q} = \frac{k^2 e^{ikr}}{4\pi\epsilon_0 r}\,\mathsf{P}_\perp\left[\mathbf{p} - \frac{ik}{6}Q\mathbf{n}\right],$$

$$\mathsf{P}_\perp\mathbf{v} = \mathbf{v} - \mathbf{n}(\mathbf{n}\cdot\mathbf{v}).$$

This coefficient belongs to the stated quadrupole convention. Changing the factor three in Q changes the prefactor in the radiation field. Comparing numerical tensors from different sources therefore requires checking their definitions. Electric quadrupole and magnetic dipole fields both arise at the same first correction to a general localized current. One cannot always describe a source whose electric dipole vanishes as purely quadrupolar without examining the current's antisymmetric moment.

A transparent demonstration uses two opposite, equal z-directed electric dipoles separated along z by distance d. Their individual small-source patterns are proportional to $\sin\theta$. The pair's far-field phase sum is $-2i\sin(kd\cos\theta/2)$ up to an overall phase. When $kd \ll 1$ the field pattern is proportional to $\sin\theta\cos\theta$, with zeros on the axis and equator and maxima at 45° and 135°. Expansion of the exact array factor gives a controlled test of the approximation. At $kd = 0.4$, $\theta = 45^\circ$, the argument of the sine is 0.1414. Replacing the sine by its argument overestimates amplitude there by about $0.1414^2/6 = 0.33\%$. A small amplitude imbalance in the opposite dipoles can reintroduce an electric-dipole contribution even where the ideal quadrupole vanishes.

Change the symmetry of a dipole pair and test which angular nulls survive

Integrate over the detector aperture

A detector collecting directions within a cone normally measures an integral of $dP/d\Omega$ times throughput and polarization acceptance. For a z electric dipole and a complete azimuthal cone of half-angle θ_a centred on +z, the fraction is

$$f_{\mathrm{pole}} = \frac{\int_0^{\theta_a} \sin^3\theta\, d\theta}{\int_0^{\pi} \sin^3\theta\, d\theta} = \frac{2 - 3\cos\theta_a + \cos^3\theta_a}{4}.$$

For small θ_a expand the integrand rather than subtracting nearly equal cosines in floating point: $f_{\mathrm{pole}} \simeq 3\theta_a^4/16$. Although a true axial ray carries no power, a finite cone collects a small positive flux varying as the fourth power of its opening angle. An isotropic-emission approximation would instead scale as $\theta_a^2/4$, badly overestimating the near-null signal. An equatorial belt between $\theta = \pi/2 - \delta$ and $\pi/2 + \delta$, spanning all azimuths, captures fraction $(3\sin\delta - \sin^3\delta)/2$. For $\delta = 10^\circ$ this is 0.258, even though the belt covers only 0.174 of the full solid angle.

To convert a small planar detector into solid angle, include its projected area. For a detector of area $A_d = 1.0$ cm² normal to the line of sight at 2.00 m, $\Delta\Omega \simeq A_d/r^2 = 2.5 \times 10^{-5}$ sr. At an equatorial maximum the normalized dipole density is $3/(8\pi)$ per steradian, giving a collected fraction 2.98×10^{-6}. Rotating its face 60° to the line of sight halves this fraction before considering polarization-sensitive throughput. An extrapolation from a small aperture to total emission is therefore a model-based estimate, whose validity can be assessed with readings at additional angles.

Separate geometrical zeros from coherence and background

Suppose equal geometry and detectors at polar angles 90°, 45° and 0° return powers 100, 50 and 5 in arbitrary units. An ideal point dipole with no offset would suggest 100, 50 and zero. Fitting a constant isotropic additive background using the axial reading leaves 95, 45 and zero. A pure dipole would predict 47.5 at 45° after subtraction. The discrepancy is small but informative if calibration uncertainty is smaller. Check finite aperture before interpreting the axial excess as a second emitter; make the aperture twice as narrow and see whether its captured fraction near the null falls by roughly sixteen, as for an ideal axial dipole cone.

If two excited source moments are coherent, add their complex fields before squaring. Their cross term can create a fore-aft asymmetry that disappears upon incoherent averaging. If their relative phase changes randomly over the integration window, add the angle-dependent intensities instead. A single angular sweep measures some combinations of magnitudes and relative phases and leaves others degenerate. Repeating with an analyser or a known phase reference helps distinguish a weak quadrupole contribution from an independent weak background. A total-power reading alone discards even more directional information.

Compare radiated power and angular dependence against a dipole calculation

Directionality from coherent electric and magnetic moments

A compact source can have two comparable leading moments of different type. This can happen if a resonant dielectric particle supports electric and magnetic responses at the same optical frequency. The long-wavelength multipole expansion still applies to the exterior observation, but symmetry or resonance can change the relative magnitudes of the moments. Add the electric fields before calculating power whenever the moments are coherently driven. Adding two separately calculated angular powers would erase a measurable directional effect.

Take an electric moment $p\mathbf{e}_x$ and a magnetic moment $m\mathbf{e}_y$, both complex peak amplitudes with the same time convention. Their leading radiation fields share the common factor $k^2 e^{ikr}/(4\pi\epsilon_0 r)$. The electric vector is $\mathbf{p} - \mathbf{n}(\mathbf{n}\cdot\mathbf{p})$. The magnetic contribution in these units is $-\mathbf{n}\times\mathbf{m}/c$. Thus in the $+z$ and $-z$ directions their electric polarisations coincide with x, but one magnetic contribution changes sign. After removal of the common amplitude their intensities are proportional to

$$I_+ = |p + m/c|^2,$$

$$I_- = |p - m/c|^2,$$

$$\frac{I_+ - I_-}{I_+ + I_-} = \frac{2\,\mathrm{Re}(p^* m/c)}{|p|^2 + |m/c|^2}.$$

If $m/c = p$, the intensity vanishes exactly in the backward direction. A quadrature phase between equal magnitudes removes the forward–backward difference, even though both moments radiate. A pair of intensity measurements therefore constrains a real cross term but does not determine a unique pair of complex moments. Rotating analysers and adding another angular direction can supply additional independent constraints.

To understand the full angular pattern, sum the squared magnitudes of the two vector contributions for an arbitrary direction. The separate terms are $|p|^2(1-n_x^2)$ and $|m/c|^2(1-n_y^2)$; their cross term is $2n_z\,\mathrm{Re}(p^* m/c)$. For the equal in-phase case the sum reduces to $|p|^2(1+\cos\theta)^2$. Its integral over the sphere is $16\pi|p|^2/3$, twice the integral for one electric dipole alone at this amplitude. The cross term

integrates to zero because it reverses sign between opposite hemispheres. Directional suppression redistributes existing radiation while leaving the sphere-integrated powers of these two orthogonal multipole types additive.

The mean angular intensity is proportional to $4|p|^2/3$, and the forward value to $4|p|^2$. The corresponding directivity is three. The backward hemisphere still contains radiation even when its axial intensity is exactly zero. Integrating $(1+u)^2$ from $u=-1$ to 0 and dividing by the integral from -1 to 1 gives a fraction $1/8$. A detector measuring integrated backward hemispherical power would report a clear finite signal. Two teams reporting an axial null and one eighth backward power are describing consistent observations if their angular acceptances differ.

The extinction of a null is exceptionally sensitive to field errors

Set $m/(cp) = (1+\epsilon)e^{i\varphi}$, with real small amplitude imbalance ϵ and phase error φ in radians. Expanding the backward intensity to the first nonzero order gives

$$I_-/|p|^2 \simeq \epsilon^2 + \varphi^2;$$

the forward intensity is about $4|p|^2$. Misalignment, background scattering and angular acceptance can also fill the nominal null. The ratio alone cannot identify an electric–magnetic amplitude mismatch unless phase and background are independently known.

A 5.0 percent amplitude error together with a five-degree phase error gives a leading estimate $I_-/I_+ \simeq 0.00253$. Using the exact formula gives about 0.00250. With 100,000 forward counts in matched exposures a perfectly calibrated backward detector would register about 250 signal counts. A background of 200 counts is comparable to this signal, and subtracting it adds uncertainty from the background exposure. Doubling the integration time does not remove an incorrectly modelled constant fraction of forward stray light reaching the backward detector.

Finite acceptance near an exact null has a different scaling. In the ideal balanced case write $\theta = \pi - \delta$. The intensity $(1-\cos\delta)^2$ grows as $\delta^4/4$. Integrated over a cone of half-angle α it scales as α^6, because $d\Omega$ contributes a further factor of δ. A small constant background integrated over the same cone instead scales as α^2. Measuring with several aperture stops can distinguish these trends only if the instrument's point-spread function and angular centring are known. For a simple electric dipole aligned with the viewing direction, the zero is quadratic in angle, giving yet another scaling.

A set of checks can distinguish a physical interference null from an insensitive receiver. Rotate the source by half a turn about the y axis so that its electric moment changes sign but its magnetic moment does not. The predicted dark direction should reverse. Reverse one measurement cable's reference phase in a coherent receiver and verify that the reported phase changes while intensity does not. Use an independent isotropic calibration source or a known dipole orientation to calibrate gain in both directions. Finally scan frequency through the resonances that create the two moments. The null should follow the relative amplitude and phase inferred from independently measured spectral responses, not just one accidentally quiet detector channel.

Cancel the leading dipole contribution

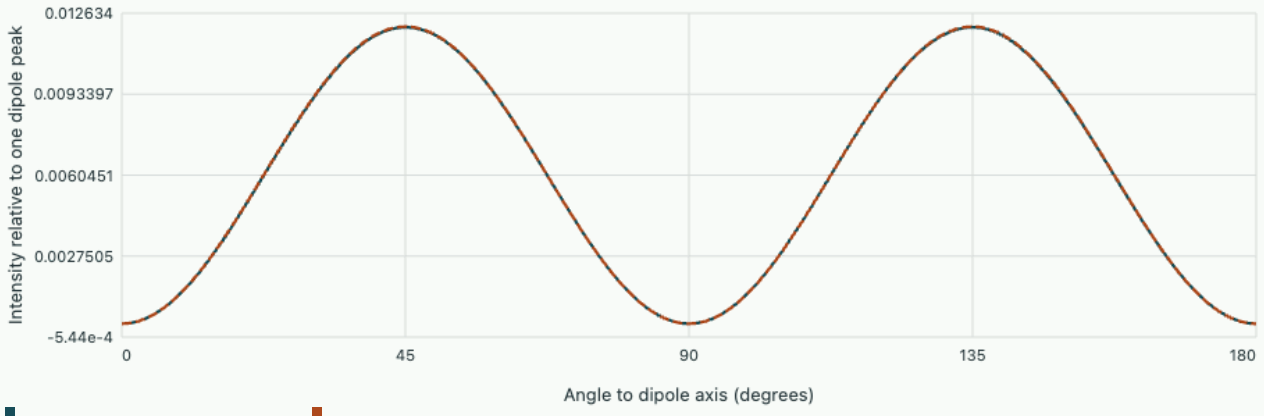
Change the separation or relative phase of the source elements and locate directions in which the leading radiation vanishes.

Open this experiment online

Try the calculation

Separation d/λ **0.035**

Opposing amplitude deficit **0**



Integrating over solid angle gives 0.9656% of the total power of one dipole. At 90° the intensity relative to its equatorial peak is 0.

Worked angular problem · fit an axis tilt from two detectors

A linearly oscillating electric dipole intended to lie along z is tilted by a small unknown angle ϵ toward $+x$. Matched detectors in the xz plane are placed at $+45^\circ$ and -45° relative to z. Their background-subtracted counts are 4200 and 5800 in equal exposures. Infer the tilt and its approximate counting uncertainty. Would reversing the electric dipole's oscillation phase change the answer?

▼ **Keep the two detector directions separate**

With positive angles toward $+x$, the expected intensities vary as $\sin^2(\pi/4 - \epsilon)$ and $\sin^2(\pi/4 + \epsilon)$. Their sum is independent of tilt, and the signed difference divided by their sum is $\sin(2\epsilon)$ when the difference is defined as $I_- - I_+$. Its measured value is 0.160. The small solution is

$$\epsilon = \arcsin(0.160)/2 = 4.60^\circ.$$

Additional mathematical solutions are excluded by knowing the source is close to the desired initial axis.

Treat the counts as independent Poisson measurements with background uncertainties negligible for this calculation. Condition on their total of ten thousand. The normalised difference has estimated standard error

$\sqrt{(1-0.160^2)/10000} = 0.00987$. Propagating through the inverse sine gives a tilt error of about 0.0050 radians or 0.29 degrees. A relative detector-gain error can imitate part of this difference. Swap the detectors physically while preserving source orientation or interchange the source tilt to separate a fixed gain bias.

Reversing the dipole phase multiplies the entire electric field by minus one, leaving both intensities unchanged. A phase-sensitive receiver with a common reference could detect the reversal. A finite aperture averages a range of angles, reducing the sensitivity if the apertures are wide. Integrate the angular predictions across the measured fields of view before applying the point-detector formula to a precision tilt.

A translated expansion can mix multipole labels

The exact far field is independent of where the expansion origin is drawn. Individual moments need not be. A translation of the source origin by a vector changes the electric dipole by the total source charge times that vector, with sign depending on the translation convention. For an electrically neutral compact source the dipole is therefore origin independent. A corresponding expansion of the quadrupole contains products of the translation with the dipole as well as the monopole. If both lower moments vanish, the first surviving electric quadrupole is likewise robust under a shift.

In a directional Fourier integral, changing the origin also changes the external phase that multiplies the expanded moments. At a fixed observation direction this common phase has no effect on intensity, but at a given truncation order its series must be kept consistently when comparing complex amplitudes. An origin-dependent angular intensity at the nominal retained order is a warning that a translated lower-order contribution or a same-order magnetic term may have been dropped.

A numerical check translates the integration grid by a known small vector while preserving the physical source and detector positions. Recompute both the unexpanded far-field integral and its truncated multipole prediction. The exact physical comparison should agree after phase referencing. The discrepancy between a leading-order expansion and the exact result should scale with the neglected power of source size over wavelength. This test probes moments, phase conventions and coordinate bookkeeping with one calculation.

Tutorial · infer a dipole moment and estimate detector uncertainty

A calibrated equatorial probe at 5.0 m observes a 300 MHz linearly oscillating dipole with RMS electric amplitude 2.00 mV/m. Determine peak moment and total emitted power. The receiver angle is uncertain by ±5° around the equator. Estimate the peak possible relative bias in a power estimate that assumes exact equatorial placement.

▼ **Keep the RMS calibration through the flux calculation**

Peak E is $\sqrt{2}(0.00200) = 0.002828$ V/m. With $k = 2\pi f/c \simeq 6.29$ m⁻¹, $p_0 = 4\pi\epsilon_0 r E_{\text{peak}}/k^2 \simeq 3.98 \times 10^{-14}$ C m. Alternatively use RMS E directly to find

$$S = E_{\text{rms}}^2/Z_0 = 1.06 \times 10^{-8}\,\text{W/m}^2.$$

The inferred total is $(8\pi/3)(25)S = 2.22$ μW. At a 5° displacement from the maximum the measured flux is multiplied by $\cos^2 5^\circ = 0.9924$, so the inferred total can be low by 0.76%. A comparable angular error near an axial null produces a much larger relative uncertainty.

Tutorial · use a symmetry-breaking parameter as a diagnostic

Two opposing parallel dipoles have moment magnitudes $p(1+\delta)$ and $p(1-\delta)$, with $kd = 0.20$. Derive the exact direction-dependent pair factor, estimate the equatorial-to-45° intensity ratio for $\delta = 0.010$, and explain why this experiment cannot determine the sign of a small real imbalance from total intensities alone in this geometry.

▼ **Retain the phase of each current element**

With $q = kd\cos\theta/2$, choose pair factor $(1+\delta)e^{-iq} - (1-\delta)e^{iq} = 2\delta\cos q - 2i\sin q$. The intensity pattern is proportional to $4\sin^2\theta(\delta^2\cos^2 q + \sin^2 q)$. At the equator q=0, so the intensity factor is

$$4\delta^2 = 0.0004.$$

At 45° q=0.07071 and the factor is about 0.01018. The ratio is 0.0393. The residual electric-dipole and leading quadrupole terms here occur in quadrature, eliminating their cross term for a real imbalance. Reversing its sign leaves each intensity unchanged. A phase-sensitive reference at the equator could recover that sign.

David Tong, University of Cambridge · radiated fields of localized currents and multipole radiation

Scattering, extinction and the phase of the forward wave

Waves, radiation and relativity

A suspension of transparent particles dims a narrow transmitted beam while absorbing almost none of its energy. Derive the scattered angular power, identify the interference term seen in transmission, and use energy conservation to put a bound on the response of a single small particle.

Build the cross sections from a specified incident beam

Illuminate an isolated target in a uniform transparent host by a monochromatic plane wave of incident intensity I_0. A far-away receiver collecting solid angle $d\Omega$ measures scattered power dP_{sca}. Define $d\sigma_{\mathrm{sca}}/d\Omega = (dP_{\mathrm{sca}}/d\Omega)/I_0$ and integrate over direction to obtain σ_{sca}. Dividing internally dissipated power by the same incident intensity defines σ_{abs}. Extinction is $\sigma_{\mathrm{ext}} = \sigma_{\mathrm{sca}} + \sigma_{\mathrm{abs}}$ for a passive isolated object in a lossless surrounding medium. All three have dimensions of area, even for a particle that has no sharply defined physical footprint.

The geometrical area of an object and its cross section answer different questions. A transparent object much smaller than its wavelength may scatter only a small fraction of the power geometrically incident on its silhouette. A resonant dipole can draw energy from an effective area comparable to a wavelength squared, much larger than a nanoparticle's silhouette. The incident field and the disturbance it generates coexist in the forward direction. Removing a fixed area from a ray diagram cannot represent the interference between those fields.

Represent the incident polarization by a normalized complex vector $\mathbf{e}_0 \perp \widehat{\mathbf{z}}$. Define a vector scattering amplitude of dimensions length by $\widetilde{\mathbf{E}}_{\mathrm{sca}} = E_0 e^{ikr}\mathbf{f}(\mathbf{n})/r + O(r^{-2})$ for the same $e^{-i\omega t}$ convention. The scattered wave is transverse to n; for a polarization-insensitive receiver $d\sigma/d\Omega = |\mathbf{f}|^2$. With a receiving analyser $\mathbf{a}$ only $|\mathbf{a}^{\dagger}\mathbf{f}|^2$ is detected. Thus an intensity pattern alone loses information about phase and possibly polarization. The forward extinction will depend on one particular quadrature of that phase.

Derive the small sphere's polarizability from two boundary equations

In vacuum take a dielectric sphere of radius a with scalar relative permittivity ϵ_r at the driving frequency. For $ka \ll 1$ and negligible retardation inside the sphere, solve the instantaneous-looking boundary problem for a phasor incident field $E_0\hat{z}$. Axial symmetry and the degree-one angular dependence admit $\Phi_{\mathrm{in}} = -Cr\cos\theta$ and $\Phi_{\mathrm{out}} = -E_0 r\cos\theta + D\cos\theta/r^2$. At r=a continuity of tangential E is equivalently continuity of these potentials; continuity of normal D supplies the other equation. They give

$$C = E_0 - D/a^3,$$

$$\epsilon_r C = E_0 + 2D/a^3,$$

$$D = a^3\,\frac{\epsilon_r - 1}{\epsilon_r + 2}\,E_0.$$

Comparing the decaying term with a dipole potential gives

$$p = 4\pi\epsilon_0 D.$$

Define the volume polarizability by $p = \epsilon_0 \alpha E_0$, so $\alpha_0 = 4\pi a^3(\epsilon_r - 1)/(\epsilon_r + 2)$ in m³. Some conventions absorb ϵ_0 in α. Naming the convention avoids spurious permittivity factors when scattering formulas are

converted between sources. At $\epsilon_r \to 1$ the particle becomes indistinguishable from its environment and the dipole vanishes. In the formal electrostatic conducting limit the coefficient tends toward $4\pi a^3$.

Radiation from this induced electric dipole gives

$$\mathbf{f}(\mathbf{n}) = \frac{k^2\alpha}{4\pi}\,[\mathbf{e}_0 - \mathbf{n}(\mathbf{n}\cdot\mathbf{e}_0)],$$

$$\sigma_{\mathrm{sca}} = \frac{k^4}{6\pi}|\alpha|^2.$$

The projection shows why scattered light from unpolarized illumination becomes partially polarized at an angle: average the power of two orthogonal transverse incident polarizations. For observation angle θ to the incident wavevector, the normalized angular factor is $(1+\cos^2\theta)/2$. At 90° an analyser can suppress the surviving polarization. The total cross section is independent of the orientation of a single linear polarization because the isotropic sphere can be rotated with that polarization.

Predict an absolute scale as well as a wavelength slope

For a transparent sphere $a = 20$ nm, index 1.50 so $\epsilon_r = 2.25$, and $\lambda_0 = 600$ nm, the contrast factor is 0.2941 and $ka = 0.2094$. The quasistatic volume polarizability is 2.96×10^{-23} m³. Substitution gives $\sigma_{\mathrm{sca}} \simeq 5.6 \times 10^{-19}$ m² compared with geometric area 1.26×10^{-15} m². At otherwise fixed properties, halving radius divides $|p|$ by eight and σ_{sca} by 64. Changing wavelength from 600 to 500 nm instead increases the leading cross section by $(600/500)^4 = 2.074$. This scaling assumes that material dispersion and finite-size corrections have stayed sufficiently small.

For a dilute suspension with $N = 10^{16}$ particles/m³ path loss from independent scattering gives

$$dI/dz = -N\sigma_{\mathrm{ext}} I.$$

Over 0.10 m at the 600 nm parameters, optical depth is about 5.6×10^{-4}, or 0.056% collimated intensity loss. Collecting a wider transmitted cone can admit some of the scattered light and reduce that apparent extinction. At high density fields scattered by one particle drive others and positional correlations can alter angular scattering. The isolated-particle prediction and exponential independent-particle extrapolation should be tested separately.

Change particle radius and compare a calculated scattering cross section

Find an interference signal that survives at arbitrarily large distance

Write the total field outside the scatterer as incident plane wave plus outgoing spherical wave. The Poynting flux through a large sphere contains incident-only, scattered-only and cross terms. The net incident-only flux through a closed sphere vanishes. The scattered-only term tends to $I_0\sigma_{\mathrm{sca}}$. The interference term accounts for the reduction of incident energy in the forward beam. To see its scaling, examine a narrow forward cap; in a scalar polarization-preserving calculation the relevant angular oscillation is $e^{ikr(1-\cos\theta)}$. The interference field is order $1/r$, but the integration area is order r^2; oscillatory phases cancel over most directions and leave a forward contribution of finite order.

Near the cap $\theta \ll 1$, $1-\cos\theta \simeq \theta^2/2$, $d\Omega \simeq 2\pi\theta d\theta$. Introduce a smooth angular acceptance equal to one near zero. The leading forward factor in the integral is

$$2\pi \int_0^\infty \theta e^{ikr\theta^2/2} d\theta = 2\pi i/(kr),$$

with arbitrarily weak positive damping understood before taking the large-radius limit. Magnetic and electric interference terms have the same forward polarization and supply twice the real cross product in total flux. A backward stationary contribution is cancelled by the opposing propagation directions in the net flux. With the outgoing-wave convention this gives

$$\sigma_{\mathrm{ext}} = \frac{4\pi}{k} \Im[\mathbf{e}_0^* \cdot \mathbf{f}(\widehat{\mathbf{z}})].$$

This is the optical theorem. A forward detector sensitive only to $|\mathbf{f}(0)|^2$ is insufficient; the unscattered incident field acts as a coherent reference that picks out the appropriate quadrature. Changing the scattering-amplitude convention changes corresponding signs. Inserting a known passive dipole is a quick way to verify the convention used in a computational implementation.

There is an equivalent small-particle derivation. The cycle-averaged work extracted from a prescribed incident electric field is

$$\langle \mathbf{E}_{\mathrm{inc}} \cdot \dot{\mathbf{p}} \rangle = (\omega/2)\,\mathrm{Im}(\widetilde{\mathbf{E}}_{\mathrm{inc}}^* \cdot \widetilde{\mathbf{p}}).$$

Dividing by $I_0 = |E_0|^2/(2Z_0)$ gives

$$\sigma_{\mathrm{ext}} = k\,\mathrm{Im}\,\alpha.$$

In the forward direction the dipole formula gives $\mathbf{e}_0^* \cdot \mathbf{f}(0) = k^2\alpha/(4\pi)$, reproducing that result exactly. Comparing integrated scattered power with interference-derived extinction tests both amplitude normalization and global energy accounting.

Repair the apparent contradiction of a real static polarizability

A strictly real static α_0 gives positive dipole radiation in σ_{sca}, yet inserting that same real number into $k\Im\alpha$ would give zero extinction. The two formulas then mix orders of approximation. Replacing the driving field by its electrostatic value determined the leading in-phase moment. The small imaginary part needed to supply radiated energy arises at the next electrodynamic order. For a lossless purely dipolar particle energy conservation requires $\sigma_{\mathrm{ext}} = \sigma_{\mathrm{sca}}$, hence

$$\Im\alpha = \frac{k^3}{6\pi}|\alpha|^2,$$

$$\Im(1/\alpha) = -\frac{k^3}{6\pi}.$$

Retaining the static inverse real part yields a simple radiation correction $\alpha = [\alpha_0^{-1} - ik^3/(6\pi)]^{-1}$. For a real α_0 this expression obeys the lossless optical theorem. It captures the radiative imaginary correction while ignoring possible real dynamic depolarization and higher multipoles; it is not an exact all-size sphere solution.

If the material itself absorbs, a passive α_0 contributes additional loss, and $\sigma_{\mathrm{abs}} = k\Im\alpha - k^4|\alpha|^2/(6\pi)$ must remain nonnegative.

Define $b = k^3\alpha/(6\pi)$. Passivity is $\Im b \geq |b|^2$, which is the disk $(\Re b)^2 + (\Im b - 1/2)^2 \leq 1/4$ in a complex-amplitude diagram. Points on the boundary are purely radiative dipoles; interior points have additional internal absorption. The largest possible $|b|$ is one at $b = i$, giving $\sigma_{\mathrm{sca}} = 6\pi/k^2 = 3\lambda^2/(2\pi)$. At that point the particle is lossless and extinction is all scattering. The largest absorption occurs instead at $b = i/2$. There $k\Im\alpha = 3\pi/k^2$, scattering is $1.5\pi/k^2$, and the difference $1.5\pi/k^2$ is absorbed. The result applies to one ideal electric dipole channel with the stated normalization. Additional channels can contribute separately for a larger or more complex object.

Choose the detector geometry that answers the desired question

A narrow-aperture transmission experiment measures an extinction-like signal if it isolates the original mode. A large integrating sphere estimates total outgoing scattered flux over its acceptance and provides an independent σ_{sca}; a temperature-rise measurement estimates heat deposition. For the same incident beam, the combination should satisfy

$$P_{\mathrm{ext}} = P_{\mathrm{sca}} + P_{\mathrm{abs}}.$$

Measuring extinction and assuming it is all absorption can exaggerate heating by orders of magnitude for transparent dielectric particles. Conversely reflection plus a small-angle detector may miss side-scattered light and turn that deficit into an apparent unexplained source term.

Suppose a calibrated intensity is 2.0×10^4 W/m², transmission implies 60 pW extracted per particle, and integrated scattering gives 45 pW. The inferred cross sections are 3.0×10^{-15} and 2.25×10^{-15} m², leaving 0.75×10^{-15} m² for absorption and 15 pW thermal load. If each of the two power readings has independent 5 pW standard uncertainty, the difference has about 7.1 pW uncertainty. The inferred absorption is then much less certain in relative terms than either directly measured quantity. A negative fitted difference within that uncertainty should prompt an examination of collection angles and background subtraction rather than an immediate interpretation as optical gain.

How energy conservation fixes a scattering phase

The forward optical theorem can appear surprising because it relates a single angular amplitude to a sphere-integrated power. For a spherical scatterer this relationship can be seen one angular momentum channel at a time. Begin with a scalar plane wave scattering elastically from a spherically symmetric object in a lossless exterior. This scalar example exposes the mathematics; electromagnetic spherical waves have separate electric and magnetic polarisation families and start at angular momentum one. A scalar zeroth channel should never be inserted directly into a vector light-scattering calculation.

Outside a sphere containing the object, the radial equation for each Legendre component has independent incoming and outgoing spherical solutions. Define S_ℓ as the ratio of outgoing to incoming wave amplitudes relative to its value for free propagation. The radial energy flux in a channel is proportional to the squared magnitude of the outgoing amplitude minus that of the incoming amplitude. For an elastic scatterer flux balance demands $|S_\ell| = 1$. Set $S_\ell = e^{2i\delta_\ell}$ with real phase shift δ_ℓ. The scattered outgoing wave is the change from the unscattered outgoing part, proportional to $S_\ell - 1$.

$$f(\theta) = \frac{1}{k}\sum_{\ell=0}^{\infty}(2\ell+1)e^{i\delta_\ell}\sin\delta_\ell\,P_\ell(\cos\theta),$$

$$\sigma_{\mathrm{sca}} = \frac{4\pi}{k^2}\sum_{\ell}(2\ell+1)\sin^2\delta_\ell.$$

The second result follows by integrating $|f|^2$ and using $\int_{-1}^{1} P_\ell P_m\,du = 2\delta_{\ell m}/(2\ell+1)$. At forward angle all Legendre functions equal one. The imaginary part of $e^{i\delta}\sin\delta$ is $\sin^2\delta$, exactly the quantity appearing in the integrated cross section. This equality in each uncoupled channel proves the elastic optical theorem in the scalar example. It does not require the forward intensity $|f(0)|^2$ to have any particular value. The calibrated phase of the forward amplitude is essential.

For a hypothetical acoustic sphere at wavenumber $k = 20$ per metre, suppose only the scalar zeroth phase shift is appreciable and has $\delta_0 = \pi/6$. Its scattering amplitude has real part 0.0217 m and imaginary part 0.0125 m, independent of viewing angle. Integrating its differential cross section gives $4\pi(0.025\,\mathrm{m})^2 = 0.00785$ square metres. The forward identity gives $4\pi(0.0125)/20 = 0.00785$ square metres as well. Shifting the phase through $\pi/2$ at this wavenumber increases the cross section by four, to the single-channel elastic maximum.

What an incomplete outgoing channel tells the experimenter

For an absorbing or internally converting target write $S_\ell = \eta_\ell e^{2i\delta_\ell}$, $0 \le \eta_\ell \le 1$. The missing outgoing flux in that exterior channel contributes to absorption or to unobserved channels, depending on the physical system. Define $a_\ell = (S_\ell - 1)/(2i)$, so the scalar wave amplitude above uses $a_\ell = e^{i\delta_\ell}\sin\delta_\ell$ in the purely elastic limit. Elementary algebra gives

$$\mathrm{Im}\,a_\ell - |a_\ell|^2 = \frac{1-|S_\ell|^2}{4},$$

$$\sigma_{\mathrm{abs}} = \frac{\pi}{k^2}\sum_{\ell}(2\ell+1)(1-\eta_\ell^2).$$

For a single channel with $\eta = 0$, the scattered amplitude is $i/2$. The residual coherent outgoing component relative to its incident plane-wave background corresponds to a finite scattered power even though this channel has zero total outgoing radial amplitude. Its absorbed and scattered cross sections are equal, while extinction is their sum. This distinction is the angular-channel counterpart of interference between incident and scattered fields in the forward direction.

An experiment may infer $\eta < 1$ because the detector fails to collect another outgoing polarisation, a converted elastic mode or a large-angle part of the pattern. To assign true heat absorption one needs the full exterior channel accounting or a calorimetric measurement. For a homogeneous liquid suspension distinguish energy dissipated in the nominal particles from energy dissipated in the surrounding medium. The preceding cross-section definitions assume a nonabsorbing reference host, so a lossy host requires an explicit incident-reference subtraction and an agreed control surface.

Suppose a measurement models the target only as an isotropic scalar scatterer, yet detects a pronounced forward lobe. A one-channel fit can reproduce a forward complex amplitude by choosing an appropriate complex a_0, but it also predicts the same intensity at every other angle. Add a possible $\ell = 1$ component and

measure at $\theta = 0, \pi/2, \pi$. At right angles the term proportional to P_1 vanishes, which helps isolate the $\ell = 0$ component. Sum and difference of complex forward and backward amplitudes separate even and odd parity. Intensity-only data lose signs and relative phase and may require additional incident conditions or interferometry.

To check a numerical partial-wave calculation, increase the highest retained order at fixed geometry and compare the integrated angular quadrature with the coefficient sum. These diagnostics test the angular grid and truncation separately. A narrow forward peak may contribute little at any grid point if no nodes resolve its centre, yet contribute substantially to extinction. Gauss–Legendre quadrature in $u = \cos\theta$ automatically includes the solid-angle measure. Finally compare the coefficient-sum extinction against the imaginary forward value with the chosen time convention. A sign error in outgoing radial waves can produce a beautifully converged negative extinction for a passive sample.

Track the frequency dependence of a scattering signal

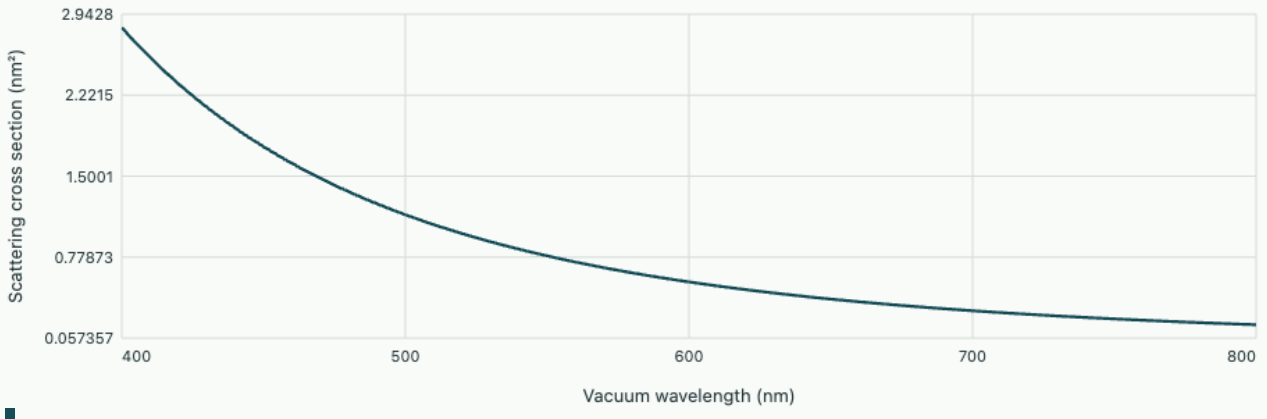
Change the ratio of forcing to resonance frequency. Decide which part of the response determines a scattering measurement and compare it with the power balance in the text.

Open this experiment online

Try the calculation

Sphere radius (nm) **20**

Relative permittivity **2.25**



Rayleigh sphere

The leading cross section at 600 nm is 0.558 nm². The shortest plotted wavelength gives ka = 0.314; finite-size corrections grow as this ratio grows.

Worked suspension problem · infer and check an optical depth

A dilute suspension has independently measured single-particle extinction cross section 5.0×10^{-15} square metres and particle density 3.0×10^{14} per cubic metre. A narrow stop rejects the scattered halo. Predict the

direct transmission through 20 mm and estimate how many particles occupy a cylindrical illuminated volume of radius 0.10 mm. Which approximation is most vulnerable if the radius is increased?

▼ **Derive the transport law and identify the role of the aperture**

In an infinitesimal length dz, the direct-beam intensity loses a fraction $n\sigma_{\mathrm{ext}}dz$ in the independent, uniform, dilute limit. Integrating gives

$$I/I_0 = e^{-n\sigma_{\mathrm{ext}}L}.$$

The extinction coefficient is 1.50 per metre and optical depth 0.0300, implying direct transmission 0.9704. A first-order calculation gives 0.9700. The small difference is consistent with neglect of higher powers in optical depth, but the exponential by itself does not describe multiple scattering into the receiver.

The illuminated volume is $\pi(10^{-4})^2(0.020) = 6.28 \times 10^{-10}$ cubic metres and contains about 188,000 particles on average. Increasing beam radius at fixed density and path length raises this count but leaves ideal optical depth unchanged. Counting more independent scatterers reduces relative sample fluctuations. It can, however, change beam divergence, spatial illumination and what scattered angles reach a finite detector. There is no general reason for local diluteness to fail merely because the beam area is larger.

Interchange cells of several lengths and measure the slope of the logarithm of transmitted intensity. An intercept different from zero suggests interface loss or a calibration offset if both incident and transmitted powers are referenced outside the cell. Repeat the length scan with a wider angular stop. A reduced slope with the wide stop signals that the detector collects part of the scattering halo; a genuine absorption contribution persists. At higher densities, evaluate interactions between particles or coherent structure before extrapolating the dilute-solution cross section from the same attenuation law.

A scattering measurement weights the large end of a distribution

The small-particle scattering law grows as the sixth power of radius for fixed material contrast and wavelength. A suspension with a modest spread of radii can therefore scatter substantially more than a calculation using the mean radius for every particle. If radii are independently distributed with mean a_0 and small fractional standard deviation s, expand $a^6 = a_0^6(1+\delta)^6$. The first correction to the ensemble mean is $15s^2$ times the monodisperse value when the centred deviations average to zero. At ten percent relative spread, the leading increase is fifteen percent.

Separately measuring total particle material changes the comparison. Material volume scales as the third power, whose corresponding correction is $3s^2$. At fixed material volume, the scattering per unit volume therefore increases by approximately $12s^2$ relative to spheres of mean radius, within the same narrow-distribution approximation. Reporting the quantity held fixed changes the numerical prediction meaningfully.

A few larger outliers can invalidate this expansion and even take the sample beyond the Rayleigh regime. Measure a size distribution independently and average the single-particle cross section over that distribution, including detector acceptance if it varies with size. A fit to a single apparent radius may match total intensity at one wavelength while failing its spectral slope or angular pattern. Comparing two wavelengths and a separate volume estimate gives a practical check on that hidden weighting.

Tutorial · examine two proposed dipole fits for passivity

At a fixed wavelength two fits to an angular intensity pattern return $b_1 = 0.30 + 0.10i$ and $b_2 = 0.30 - 0.10i$, with $b = k^3\alpha/(6\pi)$. Can angular intensity distinguish them? Find normalized scattering, extinction and absorption by dividing each cross section by $6\pi/k^2$. Which fit can describe a passive particle?

▼ **Compare interference with scattered intensity**

Both have $|b|^2 = 0.10$, giving the same angular scattering and normalized scattered cross section 0.10. Extinction in these units is $\Im b$, hence +0.10 or −0.10. Absorption is the difference, zero or −0.20. The first sits on the lossless passivity circle. The second requires an active energy supply if described by these conventions. Even an angular pattern measured over the whole sphere would retain that ambiguity because it contains squared magnitudes. The forward interference phase or an independent extinction reading discriminates.

Tutorial · recover a particle radius with a stated systematic uncertainty

At fixed wavelength and index a dilute sample scatters 1.50 times the reference signal per unit volume. Assume initially the same number density and a Rayleigh regime. Infer the radius ratio. Then allow number density to be 20% higher than assumed. Estimate how much of the signal difference could instead be a concentration effect, and propose a phase-sensitive check of the dipole interpretation.

▼ **Track the powers of radius and concentration separately**

Scattered power is proportional to Na^6 under these assumptions, so the radius ratio is

$$1.50^{1/6} = 1.070.$$

If number density is actually 1.20 times the reference, the ratio becomes

$$(1.50/1.20)^{1/6} = 1.038.$$

An apparently precise intensity can therefore constrain a combined quantity quite well without determining either factor separately. Measure an independent particle count or concentration. A forward extinction measurement should satisfy the optical theorem, while the real part of the dilute coherent refractive response scales mainly with $N\Re\alpha \sim Na^3$. Combining these observables is better conditioned than repeating a single intensity measurement.

Barton Zwiebach, MIT · optical theorem and the forward scattering amplitude

David Tong, University of Cambridge · scattering, induced dipoles and electromagnetic radiation

Diffraction from the near field to the focal plane

Waves, radiation and relativity

An illuminated aperture is a boundary condition for a wave. A screen a millimetre away, an intensity camera at a lens focus and a microscope measuring an evanescent field sample different consequences of the same condition. We will derive operators for propagating the field, estimate the approximations each instrument makes and use energy and sampling checks to tell numerical structure from a measurable fringe.

How each spatial frequency propagates

Take a monochromatic scalar field with physical value given by the real part of $U(x,y,z)e^{-i\omega t}$. This approximates one transverse component when the polarisation remains nearly uniform and angles are small. In a homogeneous transparent half-space it obeys the Helmholtz equation. Specify the complex amplitude in the plane $z=0$ and require outgoing or decaying waves for positive z. Fourier transform the two transverse coordinates, leaving a differential equation in the third.

$$\partial_z^2 \widehat{U} + (k^2 - k_x^2 - k_y^2)\widehat{U} = 0,$$

$$\widehat{U}(k_x,k_y,z) = \widehat{U}(k_x,k_y,0)e^{ik_z z},$$

$$k_z = \sqrt{k^2 - k_x^2 - k_y^2}.$$

Choose the root with positive real part for propagation and positive imaginary part for decay. Each propagating component advances in phase. Hence power integrated over a plane is unchanged for a paraxial beam with no absorbing aperture downstream. For angles approaching grazing incidence the weight for normal flux contains the direction cosine. Components with transverse wavenumber exceeding k decay exponentially. An aperture map thus contains structures invisible to a distant measurement even if the map itself is known exactly.

In vacuum a sinusoidal field structure of period 400 nm illuminated at 600 nm has $k_x/k = 1.50$. Its amplitude has fallen to $1/e$ after $1/(k\sqrt{1.5^2-1}) = 85.4$ nm. At 500 nm the remaining amplitude is 0.00287 and the squared amplitude is roughly eight parts in a million of its plane value. A numerical back-propagation that reinstates the original component also multiplies measurement errors by about 350. Improving detector gain does not restore the unknown phase.

A propagation kernel with a controlled phase error

For small transverse wavenumber, expand $k_z = k - k_\perp^2/(2k) - k_\perp^4/(8k^3) + \cdots$. The first correction becomes a chirp in Fourier space. Completing the square in its inverse transform gives the Fresnel kernel. Write $\mathbf{r}$ for a transverse observation point and $\boldsymbol{\xi}$ for a point in the aperture.

$$U(\mathbf{r},z) \simeq \frac{e^{ikz}}{i\lambda z}\int U(\boldsymbol{\xi},0)\exp\left[\frac{ik}{2z}|\mathbf{r}-\boldsymbol{\xi}|^2\right]d^2\xi.$$

This expression retains phase curvature across the aperture. The coefficient ensures that a spatially uniform boundary amplitude reproduces a uniform plane wave. An uncovered infinite plane is a useful test for a numerical convolution, although a finite computation needs padding or a truly periodic uniform test. A

measured slit has sharp edges with broad spatial spectra, so peak error near an edge can exceed the error suggested by a small angle evaluated for the central lobe.

At an observation angle θ a propagation distance z accumulates neglected phase of order $kz\sin^4\theta/8$. For $z = 0.10$ m, $\lambda = 500$ nm and $\theta = 0.01$ rad this is 0.00157 rad. At 0.05 rad it is close to one radian. A faint sidelobe at a moderately large angle can require exact angular-spectrum propagation even while a paraxial estimate describes the central spot beautifully.

Which quadratic term can the far field lose?

Expand the kernel's square into an observation phase, a source phase and a bilinear term. Only the source phase can be dropped without changing the predicted intensity directly through that factor. For source radius a its maximum value is $\pi a^2/(\lambda z)$. The remaining integral is the Fourier transform of the aperture at spatial frequencies $x/(\lambda z), y/(\lambda z)$. Define the Fresnel number using the half-width a. A convention based on full slit width gives a value four times larger.

$$U(\mathbf{r}, z) \simeq \frac{e^{ikz + ikr^2/(2z)}}{i\lambda z}\,\widehat{U}\!\left(\frac{kx}{z}, \frac{ky}{z}, 0\right),$$

$$N_F = \frac{a^2}{\lambda z}.$$

For a uniformly illuminated slit of full width $w = 0.20$ mm at 532 nm and $z = 1.0$ m, $N_F = 0.0188$. Direct integration of the linear aperture phase gives

$$I/I_0 = \operatorname{sinc}^2(\pi w x/\lambda z),$$

where $\operatorname{sinc}(u) = \sin(u)/u$. The first zeros should be 2.66 mm either side of centre. The edge phase is 0.059 rad. Moving the screen to 30 mm gives $N_F = 0.627$ and edge phase 1.97 rad. Substituting the smaller distance in the far-field formula yields a number but no longer a controlled approximation.

Compare near- and far-field predictions in the slit integral

In that experiment reduce the mesh interval before moving the screen. If the exposed variable counts panels, doubling panels at fixed aperture halves the interval. A converged difference between the full and far-field integrals is a physical failure of the approximation. A feature that moves each time the number of source samples doubles is a numerical failure. Keep the same physical horizontal axis and the same normalisation when comparing adjacent screens.

From a circular pupil to a resolution measurement

When the pupil is a disk, polar coordinates considerably simplify the Fourier integral. Align the transverse transform vector along the horizontal axis. The angular average of $e^{-iqr\cos\phi}$ defines $J_0(qr)$. One radial integration, using $d[uJ_1(u)]/du = uJ_0(u)$, then gives the result. Divide by the on-axis amplitude, proportional to aperture area.

$$\frac{U(q)}{U(0)} = \frac{2}{a^2}\int_0^a rJ_0(qr)\,dr = \frac{2J_1(qa)}{qa},$$

$$q = k\sin\theta.$$

At $qa = 0$ use $J_1(s) \sim s/2$. The first zero is at $qa = 3.8317$. Hence a disk of diameter D has its first dark ring at $\sin\theta = 1.22\lambda/D$. Two mutually incoherent identical points contribute the sum of their intensities. Rayleigh's conventional separation places each peak on the first zero of its neighbour. A coherent pair instead has an interference term that depends on the unknown relative phase. The same separation criterion cannot predict its measured trough without that phase.

A 5.0 mm pupil on a 100 mm focal-length lens at 550 nm gives a first-zero radius of 13.4 micrometres on the focal plane. A camera with 6.5 micrometre pixels spans the full bright core with about four pixels. Fitting the central pixel alone to find the image position would depend strongly on whether the spot lay on a pixel centre or a corner. A joint fit to integrated pixel intensities can estimate a position more precisely than the image width when enough photons are detected. That precision describes one isolated spot and does not directly give a two-source resolution.

Why a lens puts the spectrum close at hand

A thin positive lens adds the quadratic phase $-kr^2/(2f)$ to a field immediately in front of it. At the plane $z = f$, this cancels the source term of the Fresnel kernel exactly within paraxial theory. The focal intensity therefore samples a Fourier spectrum even for a tabletop lens standing much nearer than the free-space far-field distance. The physical exit pupil remains part of the amplitude. Closing it removes high transverse spatial frequencies in an imaging arrangement and broadens the spot of a point emitter.

For illumination with Gaussian amplitude $U \propto e^{-r^2/w^2}$, completing the transverse Fourier integrals gives another Gaussian, with focal-plane amplitude radius $w_0 = \lambda f/(\pi w)$ for a broad lens and an incident flat wavefront. If $w = 1.0$ mm, $f = 100$ mm and $\lambda = 633$ nm, $w_0 = 20.1$ micrometres. This is the $1/e$ field radius, also the $1/e^2$ intensity radius. The Rayleigh length $\pi w_0^2/\lambda$ is 2.02 mm. Moving a detector one Rayleigh length off the waist increases the radius by $\sqrt{2}$ and halves its on-axis intensity for fixed total beam power.

Use the ray-bundle experiment to locate the geometric focus before estimating its diffractive width

From a point image to transmitted object contrast

A microscope may resolve individual test objects yet deliver very little contrast for a periodic specimen at a similar length scale. The reason is that incoherent imaging acts linearly on intensity, while coherent imaging acts linearly on field amplitude. Start with mutually incoherent object points. Their expected image intensities add, so the image is the convolution of object intensity with a nonnegative intensity point-spread function $h = |h_A|^2$, where h_A is the coherent amplitude impulse response. The Fourier transform of h, normalised to one at zero spatial frequency, is the optical transfer function. Its magnitude describes attenuation of a sinusoidal intensity modulation.

For a focused unaberrated system the coherent amplitude response in Fourier space is the pupil function. The convolution theorem applied to $|h_A|^2$ shows that the incoherent transfer function is the pupil autocorrelation. For a clear circular pupil this amounts to asking how much two equal disks overlap as their

centres separate. The overlap consists of two circular sectors minus their triangular parts. With normalised frequency $s = f/f_c$ and incoherent cutoff $f_c = 2\,\mathrm{NA}/\lambda_0$ in object space, the result is

$$M(s) = \frac{2}{\pi}[\arccos s - s\sqrt{1-s^2}], \quad 0 \le s \le 1,$$

$$M(s) = 0 \quad (s > 1).$$

Here λ_0 is vacuum wavelength and $\mathrm{NA} = n \sin\theta_{\max}$. At $s = 0$ the disks coincide and the full mean intensity passes. At the cutoff their edges just touch and sinusoidal contrast vanishes. The coherent pupil transmits field spatial frequencies only through NA/λ_0. These two cutoffs answer different questions about different object and illumination statistics. A coherent phase grating viewed in exact focus can yield an almost uniform intensity image despite transmitting several field components.

For $\mathrm{NA} = 0.25$ at 550 nm, the incoherent cutoff is 909 cycles/mm. A sinusoidal target of 450 cycles/mm has $s = 0.495$ and $M \simeq 0.396$. An object with 0.60 sinusoidal intensity visibility is predicted to show about 0.238 before detector averaging. If the object-referred pixel width is 0.75 micrometres, rectangular pixel integration multiplies this contrast by $\mathrm{sinc}(\pi f p) = 0.823$; the recorded visibility is about 0.196. Nyquist frequency for this pixel spacing is 667 cycles/mm. Hence even some frequencies that the optics can transmit lie above the detector Nyquist limit and can alias back to lower apparent frequencies.

This prediction can be tested with an adjustable magnification while keeping entrance pupil and object fixed. Object-referred pixel width falls as magnification rises; a pixel-limited measurement should approach the optical transfer curve. Raising magnification also spreads light over more pixels, changing per-pixel count noise, so compare fitted contrasts with uncertainty. At frequencies where the ideal transfer is small, a high-contrast target and low read noise matter. Dividing by an arbitrarily small transfer function to sharpen the final image can turn noise into a convincing periodic pattern.

A pupil-phase error redistributes energy without absorbing it

Introduce a known defocus into a clear circular pupil. Let the phase relative to its centre vary as $\phi(r) = \Phi r^2/a^2$. The on-axis focal amplitude is proportional to the pupil average of $e^{i\phi}$, since all ideal contributions arrive in phase there. Integrate in polar coordinates and substitute $u = r^2/a^2$. Relative to the perfect on-axis intensity the result is

$$S = \left| \int_0^1 e^{i\Phi u} du \right|^2 = \left[\frac{\sin(\Phi/2)}{\Phi/2} \right]^2.$$

This special-case Strehl ratio assumes a uniform pupil and pure quadratic defocus. Removing a phase piston makes no change to intensity. Expanding for small Φ yields

$$S = 1 - \Phi^2/12 + \cdots.$$

The pupil-phase variance after removing the mean piston is $\Phi^2/12$, which explains the leading connection between reduced peak intensity and wavefront variance without assuming that every type of aberration has the same pattern.

A quarter of a wave of centre-to-edge optical path error gives $\Phi = \pi/2$ and $S = 0.811$. One half wave gives 0.405. In an ideal phase-only pupil the missing on-axis intensity appears elsewhere in the image; total

transmitted power over a sufficiently wide observation plane is unchanged. Contrast this with an absorbing pupil that removes a uniform fraction of intensity everywhere. Measuring only the brightest pixel confuses the two. Measure integrated power over increasing radii and fit a spatial profile as well as a peak.

To turn this into an axial calibration, derive the edge phase from the distance mismatch. For a paraxial marginal ray in air making angle approximately NA, an axial shift Δz introduces a relative phase magnitude near $k\,|\Delta z|\,\mathrm{NA}^2/2$. At $\mathrm{NA} = 0.10$, $\lambda = 500$ nm and $|\Delta z| = 25$ micrometres this is 1.57 rad, the quarter-wave case. With a larger numerical aperture use actual longitudinal wavevectors instead of the quadratic approximation and model polarisation where required. The formulas predict the same on-axis drop for positive and negative defocus. A single peak reading cannot give the sign of an axial misplacement; scanning through focus supplies that missing sign and can reveal an asymmetric aberration.

Distinguish aperture physics from quadrature error

Refine the aperture samples and vary their physical extent independently. A moving numerical null and a true diffraction minimum have different refinement signatures.

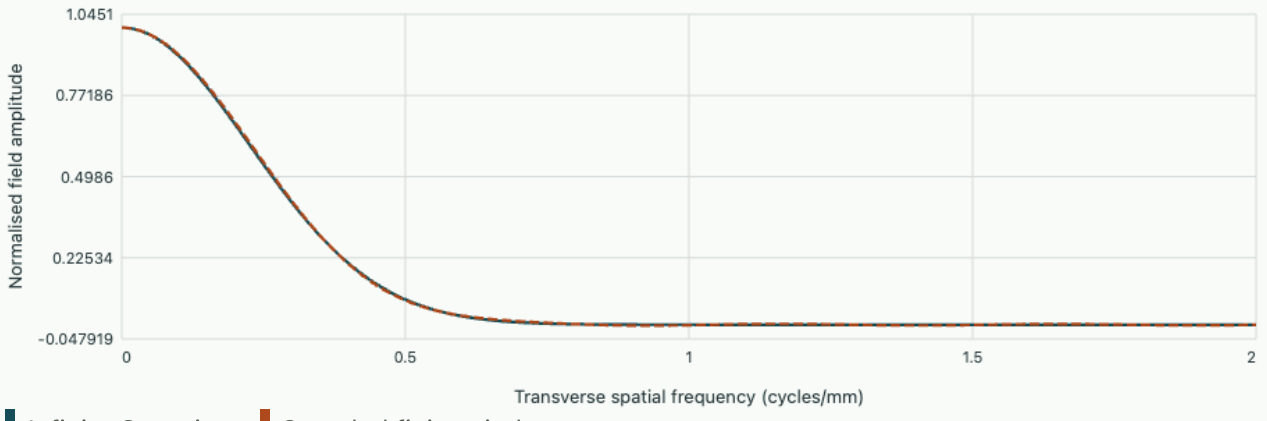
Open this experiment online

Try the calculation

Gaussian aperture radius w (mm) **1**

Window half-width / w **2**

Samples across window **30**



Sample spacing 0.13333 mm, Nyquist frequency 3.75 cycles/mm. Measured amplitude at 1 cycle/mm: -0.0014811.

Worked optical problem · determine how much light fits into one Airy disk

Two groups compare image quality using respectively the peak intensity and the energy contained inside the first dark ring. Derive the encircled-energy fraction for a uniform circular pupil and obtain its value at the first

ring. Would closing the aperture increase the useful energy fraction on a detector of fixed physical radius?

▼ **Integrate the radial intensity with its area measure**

Write $v = ka \sin\theta$, and use the small-angle focal-plane variable proportional to image radius. Focal intensity varies as $[2J_1(v)/v]^2$. The annular area element supplies a factor $v\,dv$. Use $J_0' = -J_1$ and $J_1' = J_0 - J_1/v$ to differentiate $J_0^2 + J_1^2$. The derivative is $-2J_1^2/v$, giving the fraction inside radius v as $1 - J_0(v)^2 - J_1(v)^2$. It tends to zero at the origin and to one at very large radius, which fixes the integration constants and normalisation without needing an absolute incident amplitude.

At $v = 3.8317$, $J_1 = 0$ and $J_0 \simeq -0.4028$. The encircled fraction is therefore 0.838. The remaining 16.2 percent is outside the central disk even in a perfect ideal image. If the physical detector radius corresponded to this value before closing the pupil, reducing the pupil radius decreases v at the detector edge and therefore decreases the collected fraction. It also ordinarily reduces the total incident power transmitted by the pupil.

Peak-normalised plots rescale each pupil's image to one at its centre, which hides both effects. For a fair power comparison record the total incident flux, the pupil transmission and the energy on the fixed detector. A phase aberration or obscuration can leave the nominal first-zero position a poor boundary for comparing different pupils. The same physical encircled radius offers a reproducible metric in those cases, provided the beam centre and background are estimated independently.

A phase ramp moves a diffraction pattern without changing its shape

Illuminate a thin aperture with a tilted plane wave in the paraxial regime. At the aperture plane this multiplies its complex amplitude by a linear phase ramp. The Fourier shift theorem then moves the entire focal-plane amplitude by a transverse amount $f \sin\theta_i$ along the tilt direction. Its intensity pattern retains its shape as long as the illumination envelope fills the pupil uniformly and the lens has enough acceptance. Tilting a narrow 532 nm beam by one milliradian in front of a 100 mm lens moves the pattern by about one tenth of a millimetre.

This provides a practical separation of phase tilt from aperture narrowing. A tilt moves both first zeros in the same direction, whereas changing the width of a centred slit moves the zeros apart or together. A position fit using one zero alone cannot separate these perturbations. Jointly fitting both zeros yields a mean position sensitive to tilt and a separation sensitive to width. A scanned detector with unequal position calibration on the two sides can mix the two observables.

For a spatially incoherent angular source, integrate intensities of the shifted patterns over the angular source distribution. An extended source then blurs zeros that would be sharp under one coherent angle. A coherent superposition of the same angular components instead requires adding the shifted amplitudes and can produce additional fringes. The illumination statistics determine which calculation describes the recorded image.

Tutorial · Distinguish a pupil defect from pixel integration

In a slit focal pattern the central lobe has one-sided first-zero distance 40 micrometres. An analyst samples precisely at that first zero using a centred pixel 8 micrometres wide and finds positive intensity. Estimate the finite-pixel fraction and propose a measurement that would expose an additional scattered background.

▼ **Show the pixel calculation and a discriminating test**

Let $x = x_0 + s$, where $|s|$ is much smaller than x_0. The sinc amplitude near its zero is approximately $-s/x_0$, giving intensity $I/I_{\mathrm{peak}} \simeq s^2/x_0^2$. Averaging across a pixel of width p gives $p^2/(12x_0^2)$, about 3.33×10^{-3}. This estimate uses a flat response within each pixel and a sample centred on the true minimum. A two-micrometre centring error contributes about another 0.0025 at leading order.

Use magnification that doubles the first-zero distance in pixel units while keeping pupil and wavelength fixed. Pixel leakage at a centred zero should then fall by four relative to the peak. A background proportional to local scattered light may respond differently and a dark offset can be estimated from opaque-frame counts. Calibrate differing exposure and throughput, subtract dark counts prior to normalisation and fit both flanks to locate the true minimum.

Tutorial · Design a propagation-grid check

A 512-point transverse grid with 2 micrometre spacing represents an aperture at 600 nm. What angle can the grid represent? Could its periodically repeated boundary contaminate a nominal 0.08 rad bundle after 10 mm propagation, if the beam spreads from the centre? Explain two refinements that test distinct errors.

▼ **Show the angular and spatial budgets**

Nyquist gives $|k_x| \leq \pi/\Delta x$, hence

$$|\sin\theta| \leq \lambda/(2\Delta x) = 0.15.$$

The 0.08 rad direction is resolved in principle, although modes near Nyquist would need a separate resolution study. Its lateral displacement over 10 mm is about 0.802 mm, larger than the 0.512 mm half-window even for a beam initially confined to its centre. A periodic FFT can therefore wrap light around into the opposite side.

Double the number of points with spacing fixed. This doubles the window while keeping the represented angular limit fixed. Then halve the spacing while doubling the number again to leave this enlarged window fixed and test transverse discretisation independently. Compare conserved flux, central intensity and the location of the first minimum, retaining absolute normalisation. Flux conservation in a periodic window can hold perfectly even when light has wrapped across its boundary. A plot of that single diagnostic would miss the contamination.

MIT 2.71 Optics · Fourier optics, apertures and propagation

Measure the phase speed of a numerical electromagnetic wave

Numerical field laboratory

Start with a lossless Maxwell wave, build a second-order time update and compare its actual travel time with what the continuum wave equation predicts.

A pulse can arrive late without losing energy

In a uniform lossless one-dimensional material take electric field along y, magnetic field along z and propagation along x. Maxwell's curl laws reduce to

$$\epsilon E_t = -H_x,$$
$$\mu H_t = -E_x.$$

Differentiate the first in time and substitute the second to obtain

$$E_{tt} = c^2 E_{xx},$$
$$c = 1/\sqrt{\mu\epsilon}.$$

A real laboratory signal is a combination of spatial frequencies. Different errors in travel time for different components can distort its shape even when the scheme does not attenuate any resolved Fourier mode.

Use periodic boundaries so that reflection from an end cannot be confused with this effect. Scale distance by one wavelength and time by one exact period. The periodic field $E(x,t) = \sin[2\pi(x-t)]$ has unit amplitude and returns to its initial shape after every integer period. A probe comparing just its envelope would miss a discrepancy in phase.

Derive the discrete dispersion relation

Take points separated by h and times separated by Δt. Centred second differences give

$$u_j^{n+1} = 2u_j^n - u_j^{n-1} + C^2(u_{j+1}^n - 2u_j^n + u_{j-1}^n),$$

with $C = c\Delta t/h$. Insert the complex trial field $e^{i(j\theta - n\Omega)}$. The time pair contributes $2\cos\Omega - 2$, the spatial pair $2\cos\theta - 2$. The physical positive-frequency branch obeys

$$\sin(\Omega/2) = C\sin(\theta/2),$$
$$\frac{c_{\text{phase,grid}}}{c} = \frac{\Omega}{C\theta}.$$

If $0 < C \le 1$ a real Ω exists for every grid phase, and each Fourier mode has constant discrete amplitude in exact arithmetic when started on one travelling branch. For $C > 1$ sufficiently rapid spatial oscillations have no real frequency. One time factor then grows. Testing stability on a single very smooth wave can miss this unstable part of the supported spectrum.

For long waves expand $\sin$ and $\arcsin$:

$$\Omega = C\theta + C(C^2-1)\theta^3/24 + O(\theta^5).$$

The relative phase-speed error is $(C^2 - 1)\theta^2/24$ at leading order. Here increasing C towards one improves wave propagation. This special cancellation depends on a uniform one-dimensional grid, constant wave speed and the selected difference operators.

Initialise a clean numerical experiment

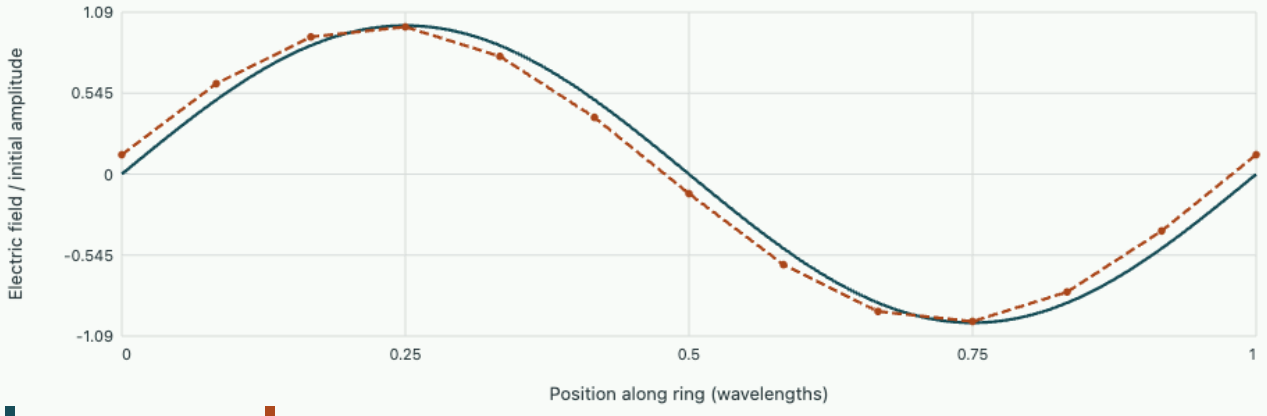
A second-order recurrence needs two time levels. Taking $u_j^0 = \sin(j\theta)$ and $u_j^{-1} = \sin(j\theta + \Omega)$ excites a single right-going numerical mode. Using the exact continuum phase $C\theta$ at the second level instead introduces a small mixture of right- and left-travelling grid modes. That is a valid test of initialisation, but it mixes errors if the quantity of interest is solely the numerical dispersion relation. State in your notes how the second level was produced.

Open the live experiment

Try the calculation

Points per wavelength **12**

Courant number **0.8**



75 steps, 5 physical periods. Phase speed ratio 0.99582. Accumulated phase lag 7.5244°.

The browser actually advances the nodal difference equations. It then reports the dispersion prediction as an independent check. The time is chosen as an integer number of numerical steps close to five exact periods and is stated with the result. At twelve points per wavelength and $C = 0.8$ there are fifteen steps per period. Both exact and discrete solutions have the same amplitude to roundoff. At the end of five periods examine how far their zero crossings have separated.

From a phase test to a pulse test

For a narrow packet, envelope motion is set by group velocity $d\omega/dk$. Differentiate the discrete relation at fixed $h, \Delta t$. The result is $v_g/c = \cos(\theta/2)/\sqrt{1 - C^2\sin^2(\theta/2)}$ on the positive branch for resolved nonendpoint modes. At $C = 0.8$ a mode close to the grid Nyquist limit has small group speed. A wave carrier

and its envelope can show different speed errors. In a real dielectric wavelength dependence may also come from the material itself; a grid-refinement test with unchanged material law separates these contributions.

As an independent benchmark, set $C = 1$. Then $\Omega = \theta$ for $0 \leq \theta \leq \pi$, and the right-going recurrence translates the samples exactly by one grid cell at each time step. The lab should align both curves even on a surprisingly coarse grid for this particular smooth periodic mode. This checks the sign and indexing of propagation without implying exact accuracy for arbitrary geometry and material interfaces.

Tutorial

A signal must accumulate less than 1 degree of phase error over ten periods. Estimate a required points-per-wavelength count for $C = 0.5$ using the leading error.

▼ **Work through the solution**

The total continuum phase advance is 20π rad. Requiring $20\pi(1 - C^2)(2\pi/N)^2/24 < \pi/180$ gives $N^2 > 450\pi^2$, or $N \gtrsim 67$. Evaluate the exact dispersion relation to check the chosen integer.

Why is measuring only the field at its crest a poor detector of small phase lag?

▼ **Work through the solution**

The derivative of a sinusoid vanishes at its crest. A small phase shift changes that value only quadratically; at a zero crossing the change is linear. A phase fit using the entire ring is more robust than either single sample.

The numerical envelope changes after a wall is introduced. What additional tests would separate dispersion from reflection errors?

▼ **Work through the solution**

First validate travel time on a uniform periodic domain. Next use a single known interface with an analytical reflection coefficient. Move the observation plane and refine the mesh separately. Compare reflected phase and amplitude with an energy-flux balance so a boundary error cannot masquerade as bulk dispersion.

Separate diffraction from a finite observation window

Numerical field laboratory

Recover the far-field angular spectrum of a smooth aperture, then isolate the errors from truncating and sampling the aperture field.

Begin with a transform that can be checked analytically

Consider a long slit system whose transmitted scalar field varies along one transverse coordinate x. At a distant observation point in direction θ, a small aperture element accumulates an approximate relative phase $-kx\sin\theta$. The Fraunhofer amplitude is proportional to $\int A(x)e^{-2\pi i f x}\,dx$ with spatial frequency $f=\sin\theta/\lambda$. The approximation neglects the quadratic phase across the aperture; either a suitable lens or sufficient propagation distance is needed. A numerical Fourier transform cannot cure violation of this optical assumption.

Take the ideal aperture field $A(x)=e^{-x^2/w^2}$ and normalise far-field amplitude by its on-axis value. Integrating by parts or completing the square gives

$$F(f)/F(0)=e^{-(\pi w f)^2}.$$

One derivation avoids any complex contour: differentiate F with respect to f and replace $xA=-(w^2/2)A'$. Integrating by parts gives

$$F'=-2\pi^2 w^2 f F.$$

Solve this first-order equation with the measured $F(0)$. Thus a wider beam on the aperture narrows the far-field distribution. Squaring the amplitude gives the intensity measured by an ordinary detector.

At $w=1$ mm, the $1/e$ amplitude spatial frequency is $1/(\pi w)\simeq 0.3183$ cycles/mm. For 500 nm light this is a small angle of 1.59×10^{-4} radians. An instrument calibrated in radians per millimetre rather than cycles per millimetre would need a factor 2π; stating the Fourier convention prevents that frequent scale discrepancy.

What the instrument has really sampled

A scanner covers $-W\le x\le W$ and gives one amplitude at each of N equally spaced cell midpoints. The numerical transform is the weighted sum $\Delta x\sum_j A(x_j)e^{-2\pi i f x_j}$, where $\Delta x=2W/N$. By symmetry its imaginary part vanishes for a perfectly centred real Gaussian; the program sums the cosine projection. It normalises by the numerical on-axis amplitude, allowing us to compare shapes on a common vertical scale.

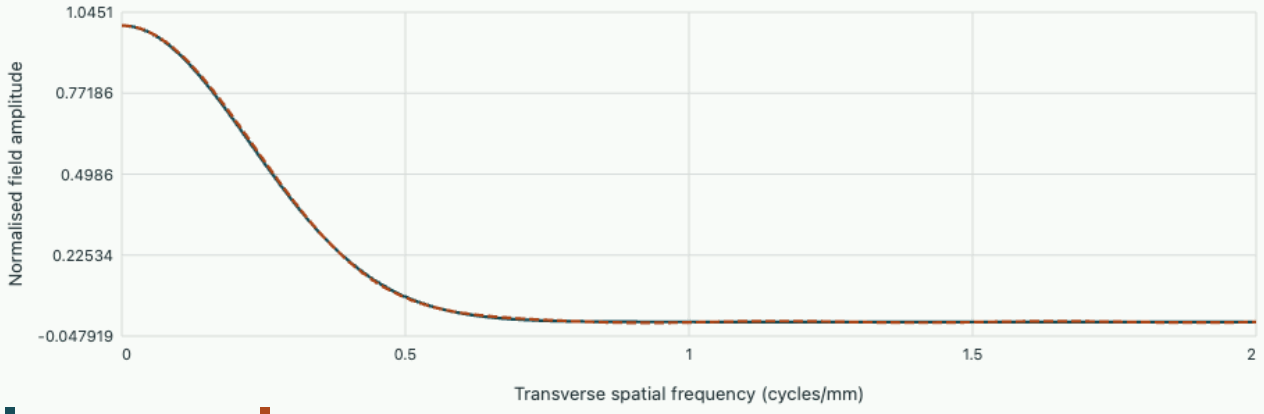
Two separate approximations have entered. Truncating the aperture multiplies the field by a rectangle. In frequency space this convolves the Gaussian spectrum with a sinc and can produce signed sidelobes. Sampling at regular intervals makes the spectrum repeat up to a phase convention, with alias copies separated by $1/\Delta x$. Increasing W at fixed N reduces truncation but worsens sampling. Widening the field of view and adding more detector pixels are not interchangeable changes.

Design a convergence test with two axes

Open the live experiment

Try the calculation

Gaussian aperture radius w (mm)	1
Window half-width / w	2
Samples across window	30



Sample spacing 0.13333 mm, Nyquist frequency 3.75 cycles/mm. Measured amplitude at 1 cycle/mm: -0.0014811.

Keep $w = 1$ mm and select 40 samples. Compare $W = w$ and $W = 3w$. The aperture field at the boundary falls from e^{-1} to e^{-9}, so an apparent sidelobe associated with the abrupt window should decrease strongly. Next fix $W = 3w$ and lower the number of samples. The displayed Nyquist frequency falls as Δx grows. When a plotted point lies above it, a smooth-looking value is an aliased sample of the periodic discrete spectrum.

At $W = 2w$, a centred Gaussian retains the on-axis amplitude fraction $\operatorname{erf}(2) \simeq 0.9953$ if integrated exactly. The normalised plot hides this 0.47% loss at zero frequency. If absolute power calibration matters, repeat the experiment without normalisation and independently integrate $|A|^2$ across the aperture. A shape comparison and a throughput comparison are different measurements, both meaningful.

Recover a size from a far-field measurement

Suppose a detector measures the intensity falling to one half at $f_{1/2}$. Since $I/I_0 = e^{-2(\pi w f)^2}$, the inferred w is $\sqrt{\ln 2/2}/(\pi f_{1/2})$. For $f_{1/2} = 0.1874$ cycles/mm the radius is about 1 mm. If the instrument actually reported amplitude half-width the corresponding formula would have $\sqrt{\ln 2}$ in the numerator; the radius would differ by a factor $\sqrt{2}$. The raw measured quantity and its threshold should accompany any quoted fitted radius.

A positive intensity cannot distinguish a negative sidelobe amplitude from a positive amplitude of the same magnitude. To observe the sign change shown by the finite-window amplitude curve, interfere the field with a

phase reference. Conversely, when comparing with a camera image, square the signed numerical amplitude once and apply the detector pixel average. Comparing a real-amplitude plot directly with a recorded intensity array is an avoidable model mismatch.

Tutorial

How could one distinguish an intrinsic physical sidelobe from ringing produced by the scanner window without a known reference aperture?

▼ **Work through the solution**

Increase the scanned width while keeping Δx fixed by adding samples. Window artefacts should change as the omitted tails shrink. Repeat with a different smooth weighting and report its effect on resolution. An intrinsic feature that survives these tests can then be compared against an independent physical model.

At $W = 3$ mm and $N = 24$, what spatial frequencies can be represented without aliasing, and what does adding zero padding do?

▼ **Work through the solution**

$\Delta x = 0.25$ mm, so the Nyquist limit is 2 cycles/mm. Zero padding supplies more displayed frequency samples of the same discrete transform. It adds no measured aperture detail and cannot resolve aliasing already present in the original samples.

A beam shifted by x_0 within a sufficiently wide window has the same far-field intensity. What information changes?

▼ **Work through the solution**

Substituting $x' = x - x_0$ in the Fourier integral multiplies its amplitude by $e^{-2\pi i f x_0}$. The phase ramp records the displacement; the intensity is invariant. A truncated window may break this invariance by clipping different amounts of field.

Estimate a cavity linewidth from a finite, biased record

Numerical field laboratory

Connect the time and frequency descriptions of one damped mode, then diagnose why a weak detector background systematically biases the inferred quality factor.

Know whether the detector reads field or energy

After a drive switches off, let a single cavity mode have complex analytic amplitude $a(t) = a_0 e^{-\kappa t/2} e^{-i\omega_0 t}$ for $t \geq 0$. The energy is proportional to $|a|^2$, giving $U = U_0 e^{-\kappa t}$. Its instantaneous decay rate $-U'/U = \kappa$ implies

$$Q = \omega_0 U / P_{\mathrm{loss}} = \omega_0 / \kappa.$$

A phase-sensitive receiver measuring field amplitude sees a time constant twice as long as the energy lifetime. Before fitting the data, ask how the measurement transducer converts the field into a reported number.

Fourier transforming the causal decaying amplitude gives an integral proportional to $1/[\kappa/2 - i(\omega - \omega_0)]$, up to the time convention. Its squared magnitude has half-power points at $|\omega - \omega_0| = \kappa/2$. Therefore the full power linewidth in hertz is $\Delta f = \kappa/(2\pi)$. A measurement of either the energy decay time $\tau_U = 1/\kappa$ or the spectral width can test the other without having to calibrate the absolute detector gain.

At $f_0 = 2.0$ GHz and $\tau_U = 0.50\ \mu$s, we obtain $\kappa = 2.0 \times 10^6$ s⁻¹, $\Delta f \simeq 318.3$ kHz and $Q \simeq 6283$. A field receiver would report an amplitude time constant of 1.0 µs. Substituting that time directly into $Q = 2\pi f_0 \tau_U$ would double the apparent quality.

Derive the two-point fit with a detector offset

Let a power detector report $S(t) = Ae^{-\kappa t} + B$, with a constant background $B > 0$. In the absence of background two samples separated by Δt give

$$\widehat{\kappa} = \log[S(t)/S(t+\Delta t)]/\Delta t.$$

With background, the logarithmic derivative is

$$-\frac{d}{dt}\log S = \kappa \frac{Ae^{-\kappa t}}{Ae^{-\kappa t} + B}.$$

It starts below the true rate and tends to zero in the tail. A straight line fitted to late log data can therefore look impressively precise while overestimating the lifetime. Increasing the acquisition duration can worsen this bias if the entire record is forced into a straight-line model.

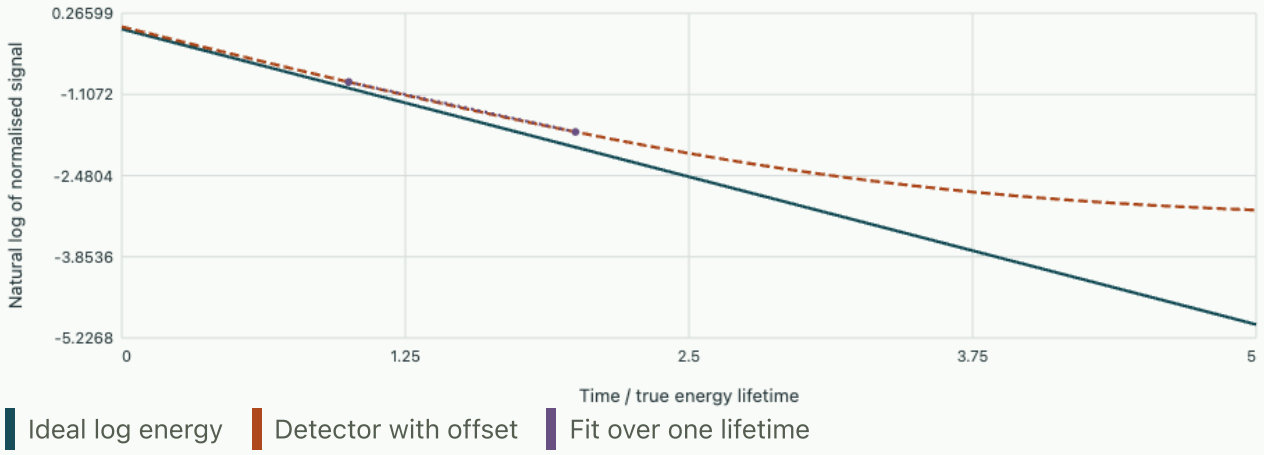
In the experiment set $b = B/A = 0.04$. A two-point fit from one true lifetime to two has dimensionless rate $\log[(e^{-1} + 0.04)/(e^{-2} + 0.04)]$, approximately 0.844. Its apparent lifetime is about 18.5% high. The source, walls and coupling have not changed. Only the interpretation of detector output has changed. Move the start of the fit towards three lifetimes and inspect the increasing gap from the ideal log line.

Open the live experiment

Try the calculation

Additive detector offset / initial energy **0.04**

Start of fitting window (lifetimes) **1**



Two-point inferred lifetime 1.1845 times the true lifetime. The same fractional bias transfers to Q for a known mode frequency.

Design a second independent inference

With three equally spaced, noiseless readings S_0, S_1, S_2 of one exponential plus a constant, subtract consecutive values before taking a logarithm. If $a = e^{-\kappa \Delta t}$, the differences are $A'(1-a)$ and $A'a(1-a)$, giving $a = (S_1 - S_2)/(S_0 - S_1)$. This cancels the constant background. At an interval of one lifetime, a first signal of unit energy and background 0.04 gives approximately $1.04, 0.407879, 0.175335$. Their ratio of differences recovers e^{-1}. In a real noise floor two nearly equal readings make this division unreliable. Use multiple windows or fit the raw signal with a measured noise model.

Finite spectral resolution creates a separate challenge. Multiplying the measured field by a rectangular window of duration T convolves its spectrum with a sinc-like response of width of order $1/T$. For the 318 kHz linewidth in the example, $T = 1$ µs gives frequency bins spaced by 1 MHz in a basic Fourier transform. Zero padding can interpolate its curve, but cannot restore the unmeasured late-time signal. A parametric fit may extract a sub-bin frequency or decay rate by using its model assumptions; its uncertainty then depends on signal-to-noise ratio and on whether the one-mode theory is adequate.

Check the single-mode model before refining its fit

If two nearby modes reach the same detector, their fields add before squaring. The result contains an interference term oscillating at their frequency difference on top of a decaying envelope. A scalar positive background cannot explain that alternating residual. Test by narrowing the excitation band, moving the pickup to a symmetry node of one mode, or examining a complex receiver's quadratures. A model with two exponentials has more degrees of freedom and can fit noise as well as signal; require consistency across independent records.

An honest linewidth report specifies the measured quantity, acquisition window, background treatment, coupling state, reference frequency and uncertainty. A fit to stored energy measures total loss. Inferring internal absorption alone additionally requires a known port coupling or a measurement that changes it separately, as in the course's cavity project.

Tutorial

A detector background is independently measured but subtracted too aggressively. What feature can appear in the logarithmic tail?

▼ **Work through the solution**

The corrected signal can approach zero and cross below it. The logarithm then becomes undefined. Even before that point its apparent local slope can become arbitrarily large. Inspect the raw signed residual and its uncertainty instead of dropping nonpositive points without accounting for selection bias.

A second experiment doubles the peak stored energy while the absolute detector offset is fixed. What should happen to the shape of the normalised log tail?

▼ **Work through the solution**

$b = B/A$ halves. Background contamination becomes important later by about $\tau_U \ln 2$. A real change in the physical lifetime need not follow that shift when the response remains linear.

Can a passive resonator's intrinsic losses be determined from a single ring-down measured at one port?

▼ **Work through the solution**

The observed decay is the sum of internal and external rates. A single accurately measured lifetime identifies their sum. A calibrated change in coupling or an independent power budget is needed to separate them.

Reconstructing a partially polarised electromagnetic wave

Electrodynamics

A detector averages enormous numbers of optical cycles. Its horizontal and vertical readings may be equal for a circularly polarised wave, a diagonal linearly polarised wave, or unpolarised light. Which additional intensities distinguish these preparations?

Start from what a polarising analyser measures

Take a quasi-monochromatic beam propagating in a fixed direction, resolved in transverse axes x, y. Use complex field amplitudes with time factor $e^{-i\omega t}$. Absorb common power and impedance factors into their units. The time- or ensemble-averaged coherency matrix $J = \langle \boldsymbol{E}\boldsymbol{E}^{\dagger} \rangle$ has four real degrees of freedom because it is Hermitian. An analyser selecting the normalised complex state $\boldsymbol{a}$ passes amplitude $\boldsymbol{a}^{\dagger}\boldsymbol{E}$ and measures $I_a = \boldsymbol{a}^{\dagger} J \boldsymbol{a}$. Thus every analyser projects the same matrix onto a distinct polarisation.

Choose $S_0 = J_{xx} + J_{yy}$, $S_1 = J_{xx} - J_{yy}$,

$$S_2 = 2\,\mathrm{Re}\,J_{xy},$$

$$S_3 = -2\,\mathrm{Im}\,J_{xy}.$$

For this convention the state $(1, i)/\sqrt{2}$ has positive S_3. Naming it by its explicit Jones vector avoids a handedness ambiguity between observer viewing directions. Inverting those definitions gives

$$J = \frac{1}{2}\begin{pmatrix} S_0 + S_1 & S_2 - iS_3 \\ S_2 + iS_3 & S_0 - S_1 \end{pmatrix}.$$

Horizontal and vertical linear analysers therefore give $(S_0 \pm S_1)/2$. The $+45^{\circ}, -45^{\circ}$ pair gives $(S_0 \pm S_2)/2$. The circular pair $(1, \pm i)/\sqrt{2}$ gives $(S_0 \pm S_3)/2$. Every pair sum recovers total intensity, providing an internal calibration check. Subtracting the complementary measurements cancels a common additive detector offset in each Stokes difference, though S_0 still needs a dark correction.

Derive the physically allowed domain

Because $I_a \geq 0$ for every analyser, both eigenvalues of J are nonnegative. Solve its quadratic characteristic equation using the trace S_0 and determinant $(S_0^2 - S_1^2 - S_2^2 - S_3^2)/4$. The eigenvalues are $(S_0 \pm |\boldsymbol{S}|)/2$, where $\boldsymbol{S} = (S_1, S_2, S_3)$. Consequently $P = |\boldsymbol{S}|/S_0$ lies between zero and one for a beam with nonzero intensity. A single fully polarised Jones state gives rank one and $P = 1$. An incoherent mixture of two orthogonal states of equal intensity has $J = (S_0/2)I$ and $P = 0$.

Suppose a corrected data set gives $I_x, I_y = 70, 30$; $I_{+45}, I_{-45} = 65, 35$; $I_{+}, I_{-} = 60, 40$ in the same units. The pair sums agree at 100. Reconstructed parameters are $(100, 40, 30, 20)$, giving $P = \sqrt{2900}/100 = 0.539$. The eigenvalues are 76.93 and 23.07. Subtracting the smaller times the identity leaves a rank-one contribution of intensity 53.85 and an unpolarised remainder of intensity 46.15. This decomposition describes the observed second moments; several microscopic preparations can produce the same J.

Predict a rotating analyser and identify its blind spot

For $\boldsymbol{a} = (\cos\alpha, \sin\alpha)$, expanding the projection and using double-angle identities gives

$$I_\alpha = [S_0 + S_1 \cos 2\alpha + S_2 \sin 2\alpha]/2.$$

A linear analyser alone is blind to S_3, even when sampled at hundreds of angles. In the example the extrema are 50 ± 25, with the maximum at $\alpha = \frac{1}{2}\,\mathrm{atan}\,2(30, 40) = 18.43^\circ$. The linear fringe visibility 0.50 is slightly below the total degree 0.539. Inserting a quarter-wave plate with Jones matrix $\mathrm{diag}(1, -i)$ before the same analyser transforms S_2 into S_3 in the intensity formula. One can now obtain the previously hidden signed correlation from readings at 45° and 135°.

The experiment prepares an adjustable polarised fraction with azimuth ψ and ellipticity angle χ, mixed with an unpolarised background. Its normalised Stokes vector is $P(\cos 2\chi \cos 2\psi, \cos 2\chi \sin 2\psi, \sin 2\chi)$. Make the unassisted intensity scan flat by choosing circular polarisation, then insert the fixed retarder virtually and predict the recovered difference. Setting $P = 0$ makes both scans flat. This comparison shows which measurement discriminates the two cases.

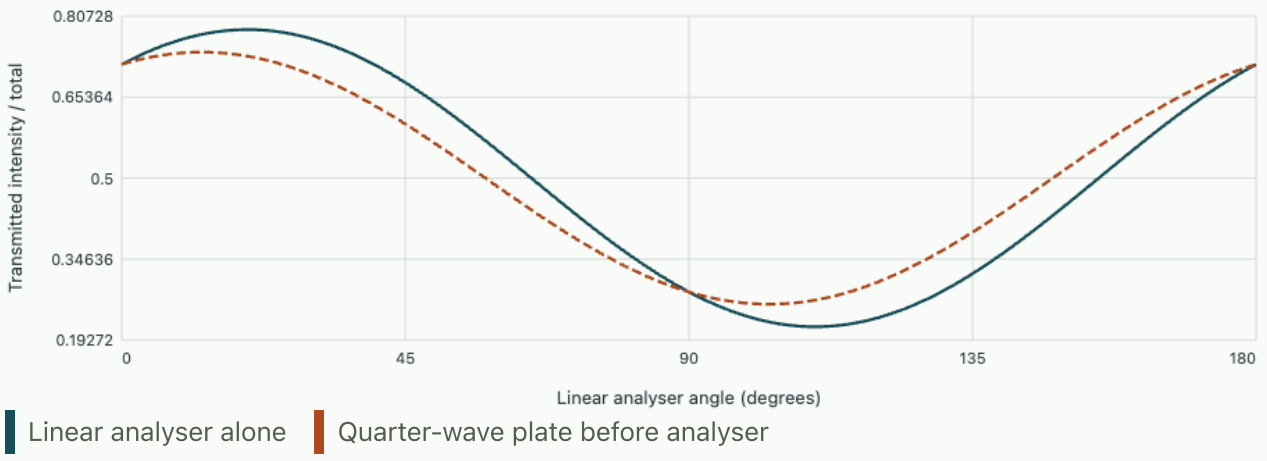
Open this experiment online

Try the calculation

Polarised fraction P **0.6**

Azimuth ψ (degrees) **20**

Ellipticity χ (degrees) **10**



Measured complementary pairs (horizontal, diagonal, circular): 0.716 / 0.284 · 0.6812 / 0.3188 · 0.6026 / 0.3974. Reconstructed $(S_1,S_2,S_3)/S_0$ = (0.4319, 0.3624, 0.2052); degree 0.6.

Tutorial on instrument design

A linear scan alone has $I_{\max} = 75, I_{\min} = 25$. Give the range of total degrees of polarisation allowed and the maximum possible magnitude of S_3.

▼ **Work through the solution**

The linear magnitude is 50 and $S_0 = 100$. Thus $0.5 \leq P \leq 1$, and the positivity bound gives

$$|S_3| \leq \sqrt{100^2 - 50^2} = 86.60.$$

A circular analyser measurement fixes the remaining magnitude and sign.

An ideal beam contains 60 units of pure horizontal light and 40 units of unpolarised light. Compare its Stokes vector with a coherent superposition of horizontal and vertical components with intensities 80 and 20 and zero relative phase.

▼ **Work through the solution**

The incoherent mixture has $(100, 60, 0, 0)$ and $P = 0.6$. The coherent preparation has $(100, 60, 80, 0)$ since $2\sqrt{80 \times 20} = 80$, and $P = 1$. Their horizontal and vertical readings coincide. A diagonal analyser gives 50 for the mixture and 90 for the coherent preparation.

All three measured component differences equal 70 at measured total 100. Can a physical beam have these parameters?

▼ **Work through the solution**

They imply $|\boldsymbol{S}| = 70\sqrt{3} = 121.2 > S_0$ and a negative matrix eigenvalue. Check dark subtraction, gain calibration, alignment and whether the source changed between readings. A physical fit with stated uncertainties is required before interpreting the data as an optical state.

Further reading · ETH Zürich, Stenflo on coherency and Stokes parameters.

Two polarisations moving through a uniaxial crystal

Electrodynamics

A crystal can send two pulses in different directions even when their wavefronts have the same normal. The distinction emerges by applying the divergence condition to electric displacement and measuring energy flow with the Poynting vector.

A field transverse to the wavevector need not be the electric field

Let a transparent, nonmagnetic uniaxial medium have $\boldsymbol{\epsilon}/\epsilon_0 = \mathrm{diag}(n_o^2, n_o^2, n_e^2)$. The optic axis is z; dispersion over the bandwidth under study is negligible. For a harmonic plane wave with wavevector in the xz plane, write $\mathbf{k} = (k_x, 0, k_z)$. Source-free Maxwell equations require $\mathbf{k}\cdot\mathbf{D} = 0$, $\mathbf{k}\cdot\mathbf{B} = 0$,

$$\mathbf{k}\times\mathbf{E} = \omega\mathbf{B},$$

$$\mathbf{k}\times\mathbf{H} = -\omega\mathbf{D}.$$

Eliminating the magnetic field gives

$$\mathbf{k}\times(\mathbf{k}\times\mathbf{E}) + k_0^2(\boldsymbol{\epsilon}/\epsilon_0)\mathbf{E} = 0,$$

$$k_0 = \omega/c.$$

An electric field along y satisfies this system with $k^2 = k_0^2 n_o^2$. This ordinary mode has transverse $\mathbf{E}, \mathbf{D}$ and energy travels parallel to $\mathbf{k}$. A second mode has $\mathbf{E}$ in the xz plane; $\mathbf{D}$ is transverse but generally not parallel to $\mathbf{E}$.

Find the phase index and the direction of energy separately

The remaining two Cartesian equations form a matrix with diagonal $k_0^2 n_o^2 - k_z^2,\ k_0^2 n_e^2 - k_x^2$ and off-diagonal elements $k_x k_z$. Its determinant vanishes for nontrivial fields. Expanding cancels the fourth-order cross term and gives

$$\frac{k_x^2}{n_e^2} + \frac{k_z^2}{n_o^2} = k_0^2,$$

$$\frac{1}{n_{\mathrm{ph}}(\theta)^2} = \frac{\sin^2\theta}{n_e^2} + \frac{\cos^2\theta}{n_o^2},$$

where θ is the angle from the optic axis to the wavevector. Along the axis both modes have the same index; perpendicular to it the extraordinary phase index is n_e. For the latter mode choose transverse $\mathbf{D} \propto (\cos\theta, 0, -\sin\theta)$. Applying the inverse permittivity tensor gives

$$\mathbf{E} \propto (\cos\theta/n_o^2, 0, -\sin\theta/n_e^2).$$

The magnetic field is along y, so the energy flow angle ψ from z satisfies

$$\tan\psi = -\frac{E_z}{E_x} = \frac{n_o^2}{n_e^2}\tan\theta.$$

The group velocity obtained by differentiating $\omega(\mathbf{k}) = c\sqrt{k_x^2/n_e^2 + k_z^2/n_o^2}$ has the same direction in this nondispersive model. This independent construction is normal to the constant-frequency ellipse. At an interface it is the tangential component of $\mathbf{k}$, fixed by phase matching, that selects the transmitted wavevector. The energy path can then diverge from the corresponding wave normal.

A crystal plate as a beam displacer and as a retarder

A normally incident beam with an oblique optic axis

Let $n_o = 1.50,\ n_e = 1.60$ and let the plate normal make 45° to the optic axis. At normal incidence the phase normal is the plate normal for both modes. For the extraordinary wave $\psi = \arctan(2.25/2.56) = 41.3^\circ$. Relative to its wave normal the energy travels about -3.7° toward the optic axis. Across a 10.0 mm thick plate the lateral displacement relative to the ordinary beam is approximately 0.65 mm. This describes propagation inside a parallel plate and neglects edge clipping.

A plate that turns linear input into circular output

Propagate perpendicular to the optic axis so both eigenmodes travel in the same direction. Their accumulated phase difference is $\Delta = k_0(n_e - n_o)d$. Linear incident polarisation at 45° to the eigenaxes has equal amplitudes. When $|\Delta| = \pi/2$ modulo 2π, those orthogonal components have a quarter-period phase difference and form circular polarisation, assuming equal transmitted amplitudes. At 600 nm with $n_e - n_o = 0.10$, the smallest such thickness is 1.50 µm. A larger multi-order plate can give the same retardance at this wavelength but responds more strongly to bandwidth and thickness errors.

Recover retardance from a pair of polarisers

Place the crystal axes at 45° to the input polariser and use a second polariser crossed with the first. Express the output field in the two crystal axes, then project onto the crossed analyser; its amplitude is proportional to $(e^{i\Delta} - 1)/2$. Thus the transmitted fraction is $\sin^2(\Delta/2)$. An arbitrary phase offset common to both modes disappears. A single intensity reading determines neither the sign of Δ nor its full order. Varying thickness or scanning wavelength can resolve the branch; the material dispersion then enters the interpretation.

Predict the transverse displacement of a crystal plate

Choose principal indices 1.50 and 1.60 and a 45° wave normal. The extraordinary beam is displaced by about −0.65 mm through a 10 mm plate; the ordinary beam follows the normal. Swap the principal indices and compare the sign and magnitude of walk-off.

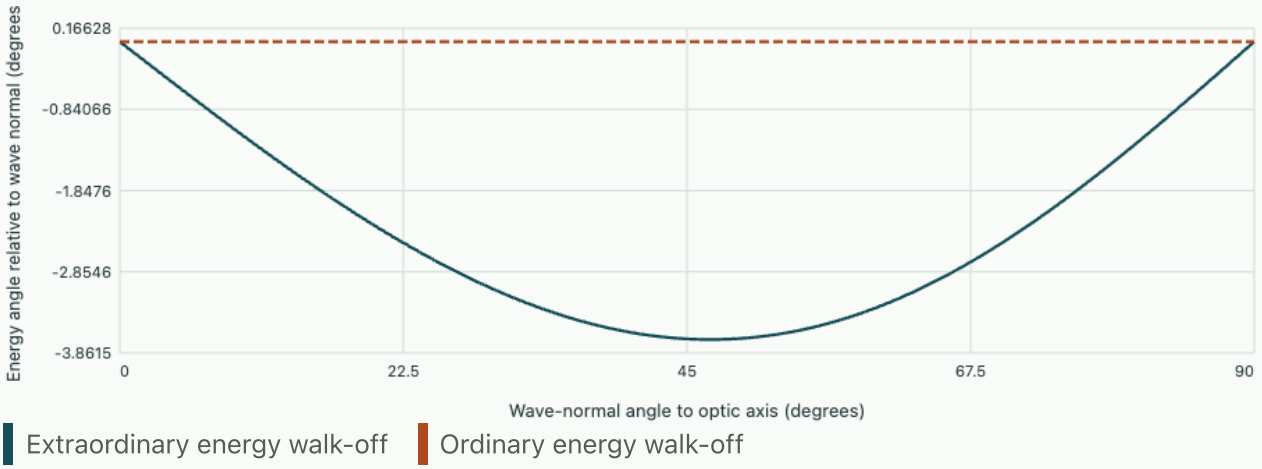
Open this experiment online

Try the calculation

Ordinary principal index 1.5

Extraordinary principal index 1.6

Probe wave-normal angle (degrees) 45



At the chosen angle the extraordinary phase index is 1.548. A 10 mm plate normal to the wavevector displaces its energy by -0.6445 mm relative to the ordinary ray.

Problem sheet · use energy and phase observations

Infer birefringence from a walk-off angle

At $\theta = 45^\circ$, an extraordinary energy ray makes $\psi = 40^\circ$ to the optic axis. Assuming the same model, recover n_e/n_o. Is the medium positive or negative uniaxial?

▼ **Work through the solution**

The ratio is

$$\sqrt{\tan 45^\circ / \tan 40^\circ} = 1.092.$$

The extraordinary principal index exceeds the ordinary one, a positive uniaxial material. A measurement referenced to the plate normal must first be converted to the angle from the optic axis used in the equation.

Calculate the contrast loss caused by angular misalignment

An ideal half-wave plate has its eigenaxis at angle α to the input polariser. Keep the analyser crossed and find the intensity transmitted when the angle is 40° instead of 45°.

▼ **Work through the solution**

Projection into the crystal axes and back out gives

$$I/I_0 = \sin^2(2\alpha)\sin^2(\Delta/2).$$

At half-wave retardance the factor from phase is unity, and the misaligned transmission is

$$\sin^2 80^\circ = 0.970.$$

The reduction alone cannot distinguish angular error from a small error in retardance.

Further reading

- Griffiths, 5th ed., §§4.3–4.4 and 9.1.4 · constitutive laws and polarisation
- Fließbach, chapter 33 · wave solutions in media

What a momentum eigenstate means in a finite experiment

Quantum states, dynamics and measurement

A detector never counts a perfectly sharp momentum eigenket. Derive continuum normalisation from a finite box, change the spectral measure from momentum to energy and predict the full evolution of a normalisable Gaussian packet.

Follow a normalised state as its allowed box grows

Consider a particle moving in one spatial dimension. A periodic box of length L is a mathematical model for an observation made far from any reflecting wall. The inner product is the integral of $\psi^*(x)\chi(x)$ over the box. The momentum operator is $-i\hbar\, d/dx$ with periodic boundary conditions. Integration by parts cancels its boundary form for periodic differentiable functions. Its normalised modes $L^{-1/2}e^{ik_n x}$ have $k_n = 2\pi n/L$ and momenta $p_n = \hbar k_n$. Their overlap is one if the integers coincide and zero otherwise. Increasing L reduces the spacing $\Delta p = 2\pi\hbar/L$ but also reduces the amplitude of each individual box mode at a fixed position. If one simply replaces the Kronecker delta by a Dirac delta while carrying over $L^{-1/2}$, the units and probabilities will fail.

Let c_n be the unit-normalised amplitudes with $\sum_n |c_n|^2 = 1$. For a smooth normalisable packet the probability in an interval dp has a limit $|\phi(p)|^2 dp$. Thus $c_n \simeq \sqrt{\Delta p}\,\phi(p_n)$. Substituting in the synthesis sum and using $L\Delta p = 2\pi\hbar$ produces

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} e^{ipx/\hbar}\phi(p)\, dp,$$

$$\phi(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} e^{-ipx/\hbar}\psi(x)\, dx.$$

The distributions $\langle x|p\rangle = (2\pi\hbar)^{-1/2}e^{ipx/\hbar}$ have $\langle p|p'\rangle = \delta(p-p')$. The delta acts under integration; $\delta(0)$ has no finite interpretation as the probability for a detector. A physical probability is $\int_{p_1}^{p_2} |\phi(p)|^2 dp$. Changing to wave number gives

$$|\phi_k(k)|^2 = \hbar|\phi(\hbar k)|^2.$$

A graph labelled density in k with a vertical axis copied directly from the momentum plot loses the Jacobian and in general ceases to normalise to one.

Change from momentum bins to energy bins

For a free particle $E = p^2/(2m)$. Each positive energy has two distinct momenta $p_\pm = \pm\sqrt{2mE}$. A detector measuring energy but not direction sums over these mutually orthogonal alternatives. Apply the ordinary change-of-variables rule on each monotone half-line:

$$w(E) = \frac{m}{\sqrt{2mE}}\left[|\phi(+\sqrt{2mE})|^2 + |\phi(-\sqrt{2mE})|^2\right], \qquad E > 0.$$

The factor $m/|p|$ is $1/|dE/dp|$, with units of momentum per energy. Measuring a finite energy interval around zero can have a density diverging as $E^{-1/2}$ if $\phi(0) \neq 0$, yet the integrated probability is finite because $\int_0^{\delta} E^{-1/2} dE = 2\sqrt{\delta}$. The same effect appears in a box as relatively dense low-energy modes. Standing-wave box boundary conditions change the counting of travelling directions; include degeneracy once and use the appropriate mode spacing, rather than doubling both.

On an infinite line it is useful to introduce $|E, +\rangle = \sqrt{m/|p|}\,|p\rangle$ at $p > 0$, and a corresponding negative branch. In the overlap,

$$\delta(p(E) - p(E')) = \delta(E - E')/|dp/dE|;$$

the two square-root factors give $\delta(E - E')$. The resolution of identity contains $\sum_{\pm} \int_0^{\infty} dE\, |E, \pm\rangle\langle E, \pm|$. Measuring energy while coherently manipulating the two degenerate directions requires keeping both amplitudes. Adding their probabilities is sufficient only for an instrument whose outcome identifies energy without mixing the directions before detection.

Integrate the free evolution of a Gaussian

Prepare a minimum-uncertainty wave packet centred at x_0 with average momentum p_0 and initial position standard deviation s:

$$\psi(x, 0) = (2\pi s^2)^{-1/4} \exp[-(x - x_0)^2/(4s^2) + i p_0 (x - x_0)/\hbar].$$

Its squared modulus is a Gaussian with variance s^2, which explains the factor four in the amplitude exponent. Complete the square in the Fourier integral to find a momentum probability Gaussian centred at p_0 of variance $\hbar^2/(4s^2)$. The initial state has zero symmetrised position–momentum covariance. Multiplying by a chirp factor $\exp[i\kappa(x - x_0)^2]$ leaves its initial position density unchanged but changes momentum variance and covariance. A snapshot of one density does not determine its later free spread.

Each momentum component evolves by $e^{-ip^2 t/(2m\hbar)}$. Split $p = p_0 + q$. The constant term in p^2 produces a common phase; the linear term shifts the envelope by $p_0 t/m$; the quadratic term changes the width. Carrying out the remaining Gaussian integral yields

$$|\psi(x, t)|^2 = \frac{\exp[-(x - x_0 - p_0 t/m)^2/(2 s_t^2)]}{\sqrt{2\pi}\, s_t},$$

$$s_t^2 = s^2 \left[1 + \left(\frac{\hbar t}{2 m s^2}\right)^2\right].$$

An independent route uses the Heisenberg operator $x(t) = x(0) + p(0) t/m$. Its variance is $s^2 + (t/m)\langle\{\Delta x, \Delta p\}\rangle + (t/m)^2 (\Delta p)^2$. Inserting zero covariance and the Gaussian momentum width reproduces s_t^2. For a chirped packet the covariance term can be negative; momentary narrowing during free evolution is consistent with uncertainty because momentum width and the phase correlation must be included. The density integral stays one, so spreading lowers the height of the peak without removing particles.

Take a model atom of mass 1.44×10^{-25} kg prepared with $s = 1.00\,\mu\mathrm{m}$. Its spread time $t_s = 2ms^2/\hbar = 2.73$ ms. At $t = t_s$ the standard deviation is 1.41 µm. With mean velocity 2.00 mm/s its mean

has advanced 5.46 µm. At long flight times $s_t \simeq t\Delta p/m$. Measuring position after a known long flight therefore maps the momentum distribution for this free packet. At $0.5t_s$ the neglected initial position variance is still dominant, so the same inference would substantially overestimate Δp.

A detector window and a Fourier grid do different jobs

Suppose the position-sensitive detector has a Gaussian point-spread function of width d. The measured density is a convolution of the true density with that response if detections are independent and gain is uniform. Gaussian variances then add:

$$s_{\mathrm{obs}}^2 = s_t^2 + d^2.$$

This broadening belongs to the observation model; it does not increase the particle's momentum spread before the measurement. For $d = 1.20\,\mu\mathrm{m}$ at $t = t_s$, the observed standard deviation becomes $1.85\,\mu\mathrm{m}$. A fit to a single image with unknown d cannot disentangle the two contributions. A sequence of flight times constrains the coefficient of t^2 independently of a fixed instrument width.

A discrete Fourier propagation represents a periodic interval sampled at spacing Δx. Its wave-number spacing is $2\pi/L$, and the highest resolved wave number is order $\pi/\Delta x$. Enlarging L at fixed Δx samples the momentum profile more finely and delays periodic wraparound. Reducing Δx at fixed L raises the bandwidth without moving the periodic boundary. Either operation alone can leave the other error intact. A norm-preserving FFT propagator can therefore return a perfectly normalised wave packet that has wrapped through an unphysical boundary or whose carrier momentum has aliased. Compare the phase, group velocity and width against independent formulas, then refine the two scales separately.

Why boundary conditions are part of a momentum observable

The expression $p = -i\hbar\, d/dx$ does not by itself specify a complete operator. On a finite interval $[0, L]$, integrate two sufficiently smooth square-integrable functions by parts:

$$\langle f, pg\rangle - \langle pf, g\rangle = -i\hbar[f^*(x)g(x)]_0^L.$$

A symmetric domain makes this boundary form vanish for every pair in the domain. A self-adjoint operator additionally has the same domain as its adjoint. This distinction matters for a physical generator of unitary translations and for assigning sharp spectral outcomes. Imposing a real or imaginary matrix representation in a finite basis can conceal boundary conditions that become essential as the basis is enlarged.

For a ring allow quasi-periodicity $g(L) = e^{i\theta}g(0)$ with fixed θ common to all functions in the domain and square-integrable first derivative. Then the boundary form vanishes. Conversely, for an adjoint candidate f the requirement that it vanish for all possible $g(0)$ forces $f(L) = e^{i\theta}f(0)$, so the same condition reappears. Solving $pg = pg$ yields the momenta $p_n = \hbar(2\pi n + \theta)/L$. Thus boxes with precisely the same local differential expression have momentum spectra shifted by an externally fixed phase. In a physically realised ring such a phase can encode a magnetic flux through a gauge-transformed boundary condition; a gauge consistent with that interpretation must also be used in the Hamiltonian and physical momentum.

Compare with forcing $g(0) = g(L) = 0$ as in a hard-wall well. The boundary form vanishes for all f, g satisfying both zero conditions, so the derivative is symmetric there. But the adjoint boundary form vanishes

for arbitrary endpoint values of f when g vanishes at both ends. Consequently this two-zero derivative domain is smaller than its adjoint domain. There is no ordinary rigid translation that carries every state of a hard-wall interval back into that same interval without assigning what happens at its boundary. Meanwhile the kinetic energy $-\hbar^2 d^2/(2m\,dx^2)$ with suitable Dirichlet domain is self-adjoint and has a well-defined discrete spectrum. The symbol $p^2/(2m)$ in a Dirichlet well does not turn its first derivative with the same two endpoint conditions into a self-adjoint sharp momentum observable.

A release-and-detect experiment makes this distinction concrete. Prepare the first Dirichlet state $\sqrt{2/L}\sin(\pi x/L)$ supported in $0<x<L$, then remove the walls suddenly and measure momentum on the full line. Its momentum wavefunction is the Fourier transform of the spatial function extended by zero outside. Multiplying the exponential and sine and integrating gives a continuous amplitude proportional to $[1+e^{-ipL/\hbar}]/[(\pi/L)^2-(p/\hbar)^2]$. The apparent singularities at $p/\hbar=\pm\pi/L$ are removable because the numerator also vanishes. It is a mistake to predict two delta peaks by replacing the standing wave with the sum of two infinite plane waves. The finite support multiplies that sum by a spatial window and broadens both momentum components.

The mean momentum for this real function is zero. Its second moment after release is $\hbar^2\int|\psi'|^2dx=\pi^2\hbar^2/L^2$, consistent with the original kinetic energy. Thus a continuous released momentum distribution can have an rms momentum whose square exactly recovers a discrete energy of the original bounded system. This is a useful quantitative example where the equality of moments across two measurement procedures does not imply equality of their full spectral probability distributions. Adding a short laboratory impulse to every particle after release should translate the whole distribution by the corresponding momentum kick, preserving its width. That response provides an independent check on a momentum detector's sign and scale.

Derive the free propagator with a convergent limiting prescription

Partitioning the free unitary into momentum eigenstates produces its position kernel

$$K(x,t;x',0)=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}\exp\left[\frac{i}{\hbar}\left(p(x-x')-\frac{p^2t}{2m}\right)\right]dp.$$

It is oscillatory rather than absolutely integrable. Define it by first damping $e^{-\varepsilon p^2}$, evaluating a convergent Gaussian and then taking $\varepsilon\downarrow 0$ on smooth normalisable test packets. Completing the square in the exponent gives, for $t>0$,

$$K=\sqrt{\frac{m}{2\pi i\hbar t}}\exp\left[\frac{im(x-x')^2}{2\hbar t}\right].$$

The phase convention of the square root for positive time follows continuously from the regularised Gaussian. The units are inverse length so $\psi(x,t)=\int K(x,t;x',0)\psi(x',0)\,dx'$ has the same units as the initial wavefunction. As $t\to 0^+$, the kernel does not approach a finite spike pointwise. Its increasingly fast oscillations across x' cancel except near $x'=x$, recovering the initial state in the appropriate strong or distributional sense. For a constant initial amplitude one may recover unity formally, but such an amplitude on the full line is itself not a normalised test state.

Two checks involve no wave packet fitting. Inserting two kernels for t_1, t_2 and integrating over an intermediate position gives the kernel for $t_1 + t_2$, either by another Gaussian integration or, more directly, by composing the momentum phase factors. Complex conjugation and reversing the time arguments yield the inverse evolution. Neither statement holds for an arbitrary choice of magnitude prefactor assigned merely to obtain the correct phase. In Fourier numerics the corresponding checks are unit norm after repeated evolution and recovery of the original sampled complex state after a forward followed by backward propagation. These are necessary algorithm checks, even though a periodic or aliased computational model can still satisfy them exactly.

This kernel also clarifies the long-time momentum mapping. Factor $e^{imx^2/(2\hbar t)}$ from K and identify the Fourier transform of $e^{imx'^2/(2\hbar t)}\psi(x', 0)$ evaluated at $p = mx/t$. If the occupied initial region is narrow enough that $mx'^2/(2\hbar t)$ varies little across it about the selected origin, the extra phase can be neglected and

$$|\psi(x,t)|^2 \simeq \frac{m}{t}\left|\phi\Big(\frac{mx}{t}\Big)\right|^2.$$

The factor m/t is the Jacobian dp/dx. When the initial packet is displaced, use a suitable shifted origin or account for the induced linear phase explicitly. This derivation applies well beyond Gaussians and gives a scale for testing the approximation before interpreting a flight image as a momentum histogram. The quadratic residual phase provides the matter-wave analogue of the near-field correction in diffraction.

Distinguish a detector crossing from a snapshot in space

A detector counts particles passing through an aperture over time. Its signal is related to flux, whereas a camera image of identical preparations at one time probes a position distribution. For a one-dimensional free wavefunction the continuity equation is $\partial_t|\psi|^2 + \partial_x j = 0$ with $j = (\hbar/m)\,\mathrm{Im}(\psi^*\partial_x\psi)$. Integrating over the region $x < X$ gives the change in probability to the left of a fixed plane. If there is no flux at negative infinity, the net integrated current through X equals the reduction in that probability. A positive-reading counter with absorption is a more detailed instrument and cannot in general be identified with a signed current in the presence of reflected waves or interference.

For a minimum-uncertainty Gaussian initially at x_0, the exact phase in the freely expanding envelope has a quadratic term. Differentiating the phase gives the local velocity entering current,

$$\frac{j(x,t)}{|\psi(x,t)|^2} = \frac{p_0}{m} + \frac{t}{t_s^2 + t^2}\left(x - x_0 - \frac{p_0 t}{m}\right).$$

At the centre this equals the group velocity p_0/m, but ahead of the centre it is larger and behind it smaller. A single constant velocity multiplying density is inadequate for a significantly spreading packet. For the earlier atom at $t = t_s$, one current-standard-deviation distance ahead corresponds to $1.41\ \mu\mathrm{m}$ and the additional speed is about 0.259 mm/s. The trailing side gives the opposite correction. Direct differentiation of its Gaussian density verifies that this current satisfies continuity. A code reproducing s_t but evolving the wrong complex phase may fail this current test; judging an entire propagation by the changing width alone would miss the error.

A useful inverse question is whether two time-of-flight images suffice to recover an unknown complex wavefunction. In the long-time limit the image yields $|\phi(p)|^2$, which contains no direct momentum-dependent

phase. Translation of the initial wave packet changes $\phi(p)$ by a linear phase, leaving the momentum density unchanged. Complex conjugate or chirped preparations can also share particular intensity records. Additional calibrated evolution, known reference interference or structural assumptions can provide phase information. Even an exact free propagator only defines the forward map; its invertibility on complex amplitudes does not imply intensity-only data identify every possible input. The experiment must specify which class of preparations and which measurement settings render the desired parameters estimable.

Track a continuous superposition through focus

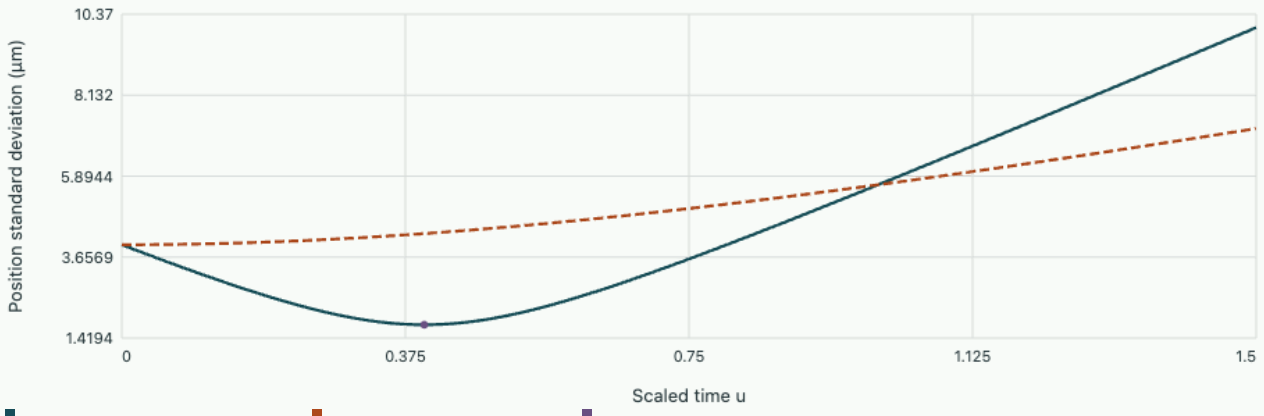
Change a position-momentum correlation and compare the full packet width with an initially unchirped packet. A narrow width at one time does not determine its subsequent spread.

Open this experiment online

Try the calculation

Initial phase curvature b

-2



For u ≥ 0 the minimum width is 1.789 µm at u = 0.4.

Plan an FFT comparison using physical position and momentum resolution

A grid propagator can track the same packet analytically solved earlier, then carry over its verified conventions to less elementary potentials. A common trap is to regard the number of points N as a single accuracy parameter. Define a periodic interval of physical length $L = N\Delta x$. Discrete Fourier modes have $k_j = 2\pi j/L$, with negative wave numbers represented by negative indices after reordering. Their spacing is $\Delta k = 2\pi/L$; the largest distinct magnitude is close to $\pi/\Delta x$. A narrow structure in momentum requires large L. A high carrier momentum requires small Δx. These conditions can fail independently even when the norm stays exactly one.

If $\psi_l = \psi(x_l)$ is a sampled wavefunction in inverse square-root metres and the forward transform sums without $1/N$, the approximated continuum k-amplitude is

$$\phi_k(k_j) \simeq \Delta x \sum_l \psi_l e^{-ik_j x_l}/\sqrt{2\pi}.$$

The phase includes the physical coordinate origin adopted for the grid. Parseval then gives

$$\Delta k \sum_j |\phi_k(k_j)|^2 = \Delta x \sum_l |\psi_l|^2.$$

A library returning a unitary discrete transform instead assigns a factor $1/\sqrt{N}$ to both directions; multiplying its output by $\sqrt{\Delta x/\Delta k}$ converts sample units. Plotting a library output directly against physical momentum without these scale factors is especially misleading when comparing two meshes. Total discrete ℓ^2 norm and continuum probability density are different normalisations of the same samples.

For the atom of mass 1.44×10^{-25} kg, $s = 1.00$ µm and $v_0 = 2.00$ mm/s,

$$k_0 = mv_0/\hbar = 2.73\,\mu\mathrm{m}^{-1},$$

$$\Delta k_{\mathrm{rms}} = 1/(2s) = 0.50\,\mu\mathrm{m}^{-1}.$$

A box 80 µm wide gives spacing $0.0785\,\mu\mathrm{m}^{-1}$, about 6.4 modes per rms momentum width. With $N = 256$, grid spacing is 0.3125 µm and Nyquist wavenumber is $10.05\,\mu\mathrm{m}^{-1}$, exceeding $k_0 + 5\Delta k_{\mathrm{rms}} = 5.23\,\mu\mathrm{m}^{-1}$. This comfortably samples the initial carrier distribution. It is still necessary to check that the envelope remains inside the periodic box at the desired final time.

At $t = 10$ ms the packet centre has moved 20 µm and $s_t \simeq 3.80$ µm. Starting at the centre of an interval spanning -40 to 40 µm, its five-standard-deviation front approaches 39 µm and begins testing the boundary margin. At 20 ms the centre alone reaches the periodic boundary. A perfectly unitary calculation will show it reappearing on the opposite side. This is correct for a ring and incorrect for a free-space flight image. Enlarging N while keeping the same L improves the Nyquist limit but does nothing to postpone this crossing. An open-boundary absorber can delay reflected artefacts but attenuates norm; a lost norm in that model should be compared with counted escape rather than naively enforced to zero.

Infer displacement and speed from a wrapped packet without mistaking periodicity for reflection

A code on a periodic box of length 80 µm reports a sharp packet peak at $+35$ µm in one frame and -36 µm in the next frame 4.50 ms later. Find a consistent uniform drift and identify the ambiguity if no speed prior is available.

▼ Derive and check the result

The smallest forward displacement is $+9$ µm after adding one full box length to the second position. It corresponds to 2.00 mm/s and agrees with the earlier preparation. Direct subtraction without unwrapping would give -71 µm, or -15.8 mm/s, an apparent reversal. All displacements $9 + 80n$ µm for integer n yield the same wrapped pair, so the frames alone cannot identify the number of turns. A momentum measurement or a bound on speed relative to frame spacing resolves the branch. For an extended packet crossing a seam, use a periodic circular moment such as $\langle e^{2\pi i x/L} \rangle$ so an ordinary spatial mean does not spuriously jump

toward the box centre. If that circular moment is nearly zero, even its angular centre is ill-defined and a full density comparison is preferable.

Tutorial · design a momentum measurement

Distinguish an unresolved line from a finite-width packet

A periodic calculation of length $L = 100\,\mu\mathrm{m}$ contains mode n. A detector bins momentum in intervals $1.0 \times 10^{-29}\,\mathrm{kg\,m/s}$. How many box-mode spacings fit into a bin? What changes if the box length doubles at fixed physical packet?

▼ **Work through the solution**

Here $\Delta p = h/L = 6.626 \times 10^{-30}\,\mathrm{kg\,m/s}$, so the detector bin spans about 1.51 spacings. The precise number of grid points falling within a bin depends on its endpoints. Doubling L gives roughly 3.02 spacings per bin. Each corresponding coefficient squared decreases approximately in proportion to Δp, while the number included increases. The total probability should approach a stable integral on a progressively finer grid. If a graph compares $|c_n|^2$ across the boxes, divide by the corresponding Δp before calling the ordinate a continuum density.

Calculate a threshold energy interval without evaluating a delta at zero

A packet's momentum density is nearly constant w_0 over $-p_c < p < p_c$. Obtain the probability for $0 < E < E_c = p_c^2/(2m)$ by two different integrations. What changes when only positive-momentum particles reach the counter?

▼ **Work through the solution**

Direct integration in momentum gives $2p_c w_0$. In energy the two branches contribute $2mw_0/\sqrt{2mE}$. Its integral gives $4mw_0\sqrt{E_c}/\sqrt{2m} = 2p_c w_0$ as required. A positive-direction selector halves this particular probability when applied to the original ensemble. If the entire subsequent experiment is renormalised to the selected particles, divide again by the total positive-momentum probability. Those two reported probabilities correspond to different conditioning and should be labelled explicitly.

Infer the packet width from two images

The model atom has an initial measured width 1.56 µm and a width 1.85 µm after 2.73 ms. Assume a time-independent Gaussian instrument response and a freely evolving unchirped minimum-uncertainty state. Infer s and d, and identify an additional check.

▼ **Work through the solution**

Subtract squared widths to remove the detector:

$$s_{\mathrm{obs}}(t)^2 - s_{\mathrm{obs}}(0)^2 = \hbar^2 t^2/(4m^2 s^2).$$

The observed difference is about $0.99\,\mu\mathrm{m}^2$, giving $s \simeq 1.0\,\mu\mathrm{m}$ and $d = \sqrt{s_{\mathrm{obs}}(0)^2 - s^2} \simeq 1.2\,\mu\mathrm{m}$, within rounding. A third flight time tests the predicted quadratic in t. A term linear in time would point to a nonzero initial correlation or to measurement drift. Independent calibration of the point-spread function distinguishes those interpretations from a nonminimum momentum width.

Further reading

- Fließbach · Eigenwerte und Eigenfunktionen
- TUM · Theoretische Physik 3, representations and momentum states
- Course investigation · probability density and current

From a finite pulse to a continuum transition rate

Quantum states, dynamics and measurement

Switch a weak perturbation on for a measured duration. Its Fourier transform determines the amplitude of a discrete transition. Summing over a dense set of final states reveals where a time-independent transition rate comes from and where it fails.

Remove the known free evolution before approximating

Let $H_0|n\rangle = E_n|n\rangle$ and $H(t) = H_0 + V(t)$. Expand $|\psi(t)\rangle = \sum_n c_n(t) e^{-iE_n t/\hbar}|n\rangle$. Substituting into the time-dependent Schrödinger equation cancels the free phase derivatives and leaves an exact equation $i\hbar \dot{c}_f = \sum_n e^{i(E_f - E_n)t/\hbar} V_{fn}(t) c_n(t)$. Suppose the initial state is $|i\rangle$ and the perturbation is appreciable only between 0 and T. Iterating the integral equation once, replacing $c_n(t)$ on the right by δ_{ni}, yields

$$c_f^{(1)}(T) = -\frac{i}{\hbar} \int_0^T e^{i\omega_{fi} t} V_{fi}(t)\, dt,$$

$$\omega_{fi} = (E_f - E_i)/\hbar.$$

For $f \neq i$ the leading transition probability is $|c_f^{(1)}|^2$. It is second order in the perturbation even though the amplitude is first order. Conservation of total probability to this order requires a second-order correction to the initial amplitude. Simply appending all positive final probabilities while keeping the initial probability exactly one would fail that test. The approximation is controlled when total probability removed from the initial state is small over the observation duration and higher-order pathways are negligible.

Let the resonant component of a drive have matrix element $V_{fi}(t) = v e^{-i\omega t}$ for $0 < t < T$. For a real cosine perturbation the two exponentials each carry half the real peak amplitude, so v refers to just the near-resonant part. Define $\Delta = \omega_{fi} - \omega$. Direct integration gives

$$P_f^{(1)} = \frac{|v|^2 T^2}{\hbar^2} \operatorname{sinc}^2(\Delta T/2),$$

$$\operatorname{sinc} x = \sin x / x.$$

At resonance the amplitude grows linearly in time while its phase remains coherent. A detuned contribution winds around in the complex plane and largely cancels. The first zeros occur at ordinary detuning $\pm 1/T$ in hertz. The full width at half height is approximately $0.886/T$ in ordinary frequency since $\sin x/x = 1/\sqrt{2}$ at $x \simeq 1.3916$. Doubling duration quadruples the line-centre probability for one final state but halves the linewidth, a distinction central to interpreting a continuum rate.

Sum over final states before taking a long-time limit

A large box provides a discrete set of final levels with spacing δE; at fixed finite time many can lie inside the pulse's energy bandwidth. Let $\rho(E)$ be the density of the accessible final states per joule, including degeneracy exactly once. Replace the sum by an integral with $E_* = E_i + \hbar\omega$:

$$P_{\text{out}} \simeq \frac{T^2}{\hbar^2} \int dE\, \rho(E) |v(E)|^2 \operatorname{sinc}^2 \left[\frac{(E - E_*)T}{2\hbar} \right].$$

If $\rho|v|^2$ changes little across the dominant width and the line is well away from thresholds, take it outside the integral. Substitute $u = (E - E_*)T/(2\hbar)$, and use $\int_{-\infty}^{\infty} (\sin u/u)^2 du = \pi$. The area of the squared Fourier transform grows linearly, giving

$$\Gamma = \frac{P_{\mathrm{out}}}{T} \simeq \frac{2\pi}{\hbar} |v(E_*)|^2 \rho(E_*).$$

The identity for the sinc area also follows by Parseval from the energy in the rectangular time window. This is Fermi's golden rule with ρ per energy. If a supplied density is per unit angular frequency,

$$\rho_\omega = \hbar \rho_E;$$

failing to change that measure introduces a factor of $\hbar$. The units are a strong check: energy squared times states per energy divided by action is inverse seconds.

Several limits have been taken in a definite order. The perturbation lasts long enough to resolve the scale on which $\rho|v|^2$ varies but short enough that many levels still fit in its bandwidth. A window of times must exist between a reservoir correlation time and a recurrence time of order $2\pi\hbar/\delta E$. In addition $\Gamma T \ll 1$ for the first-order depletion argument. Taking $T \to \infty$ at fixed finite box instead resolves individual levels and can produce coherent oscillations or recurrences. Extending the constant rate to a full exponential decay uses a further memoryless approximation and renewed weak transitions of the remaining population. A linear short-time fit by itself does not establish exponential dynamics at every time.

Keep matrix elements and density normalisation consistent

For illustration choose $|v| = 5.00 \times 10^{-6}$ eV and $\rho = 20.0\ \mathrm{eV}^{-1}$ for box-normalised accessible modes dense enough on the pulse scale, giving $\Gamma \simeq 4.77 \times 10^{6}\ \mathrm{s}^{-1}$ as a formal product. But the implied mean level spacing $1/\rho = 0.050$ eV corresponds to a recurrence scale of only 8.27×10^{-14} s. Using a nanosecond pulse with this literal finite density violates the continuum criterion, irrespective of how small ΓT looks. This numerical mismatch is a useful trap: reproducing a golden-rule number to many digits does not ensure the system supports irreversible decay in the chosen time range.

Now enlarge the quantisation volume by 10^6 for continuum-normalised physical coupling. The number of accessible states per energy grows by 10^6, while each squared matrix element shrinks by 10^6 for a localised initial state and individually unit-normalised extended final states. The same product gives the same Γ, but spacing is now 5.0×10^{-8} eV and the recurrence scale 82.7 ns. At $T = 2.0$ ns, $\Gamma T = 0.00954$ and roughly $h/(T\delta E) \simeq 41$ spacings fit up to the first positive zero. Whether density and coupling are locally smooth on that bandwidth remains an additional condition. The scaling removes explicit box dependence from a physical rate while clarifying why direct sums at finite volume can require larger spaces.

Use phase space to predict the shape of a continuum edge

For free relative motion in three dimensions, count single scalar internal channels in a cubic periodic volume $\mathcal{V}$. One state occupies volume $(2\pi)^3/\mathcal{V}$ in wave-vector space. The shell between $k, k + dk$ thus contains $\mathcal{V} 4\pi k^2 dk/(2\pi)^3$. From $E = \hbar^2 k^2/(2m)$ it follows that $\rho(E) = \mathcal{V} m k/(2\pi^2\hbar^2)$, which vanishes like $\sqrt{E}$ at threshold. In one dimension the corresponding free density diverges like $1/\sqrt{E}$; the geometrical shell

measure is different. A threshold suppression observed in one spectrum could come from available states, the coupling matrix element, or both.

If a local transition couples chiefly to a final orbital partial wave l, its regular low-k free radial factor near the source scales as k^l when the potential is sufficiently short ranged and no threshold resonance changes the law. Squaring and multiplying by the three-dimensional density gives

$$\Gamma_l \propto k^{2l+1} \propto E^{l+1/2}.$$

For a p-wave channel doubling the excess energy increases this ideal rate by $2^{3/2} = 2.83$. Long-range Coulomb fields alter the final-state radial behaviour and can invalidate the short-range threshold power. Very close to any edge a finite pulse straddles energies where the density changes rapidly or vanishes. Replace the local golden-rule product by the finite-window integral to model the measured onset.

Compare pulse shapes at equal area and equal energy

The rectangular pulse is useful because its Fourier transform is elementary. Its abrupt start and end also generate sidelobes that can excite nearby unwanted transitions. For a general narrowband drive write its resonant matrix element $v(t)e^{-i\omega t}$. First order gives

$$P_f = |\int v(t)e^{i\Delta t}dt|^2/\hbar^2.$$

At zero detuning only the complex pulse area $\int v(t)dt$ matters. After summing over a flat continuum, Parseval gives $\int P_f(\Delta)d\Delta = (2\pi/\hbar^2)\int |v(t)|^2 dt$, so it is the temporal integral of squared coupling that controls total spectral weight. Two pulses with the same area can have different continuum probabilities; two with the same integrated intensity can have different resonant peak heights.

For a Gaussian envelope $v(t) = v_0 e^{-t^2/(2\tau^2)}$ on the full line, complete the square in its Fourier integral. The amplitude is $v_0\sqrt{2\pi}\tau e^{-\Delta^2\tau^2/2}$, so the transition profile has no sidelobes and has full half-height angular width $2\sqrt{\ln 2}/\tau$. The time intensity is $|v_0|^2 e^{-t^2/\tau^2}$, with temporal full half-height width $2\sqrt{\ln 2}\tau$. Multiplying this by the ordinary-frequency full width gives

$$2\ln 2/\pi = 0.441.$$

Replacing an amplitude width by an intensity width changes that product by $\sqrt{2}$; specifying the plotted quantity is necessary when comparing a spectrum with temporal photodetector data.

Take a Gaussian with $\tau = 0.50$ ms. Its spectral intensity full width is 530 Hz and pulse intensity full temporal width 0.833 ms. An unwanted line 1.0 kHz away receives a relative first-order probability $e^{-(2\pi 1000\tau)^2} \simeq 5.17 \times 10^{-5}$. A nominally same-duration rectangular pulse may place an ideal zero near some precisely chosen detuning, but neighbouring frequencies generally see sidelobes. Which pulse gives smaller unwanted excitation depends on what quantity is fixed, where the extra lines lie and how much uncertainty there is in their positions. A shaped envelope can suppress spectral leakage without improving every possible resolution metric.

A pair of pulses illustrates how coherently adding amplitudes differs from broadening a single envelope. Let identical, short pulses of area A be separated by start-time difference T_s, and let the second have a

controlled drive phase ϕ. At a detuning small compared with each pulse bandwidth, the amplitude is proportional to $A[1 + e^{i(\Delta T_s + \phi)}]$. Its squared magnitude has fringes $2|A|^2[1 + \cos(\Delta T_s + \phi)]$ inside the broader single-pulse line. Their ordinary-frequency spacing $1/T_s$ can be much smaller than the inverse individual pulse duration. Randomising the relative phase removes fringes while keeping the two single-pulse energies unchanged. The observed fine spectral structure is therefore a measurement of coherent time separation as well as energy difference.

Derive depletion and the energy shift from the same kernel

The first-order rate gives initial population loss in a useful time window. To see which extra assumptions allow exponential decay, retain a single initial discrete level coupled to many initially empty final levels with time-independent couplings v_f in a rotating frame. Let $c_i(0) = 1,\ c_f(0) = 0$. In the interaction picture $\dot{c}_f = -(i/\hbar) v_f e^{i\omega_{fi} t} c_i(t)$. Integrate this equation exactly in terms of the unknown c_i and insert the result in its partner equation for $\dot{c}_i$. Assuming final levels do not couple directly to one another gives

$$\dot{c}_i(t) = -\int_0^t K(t - t') c_i(t')\, dt',$$

$$K(\tau) = \hbar^{-2} \sum_f |v_f|^2 e^{-i\omega_{fi}\tau}.$$

No irreversibility has been assumed in this elimination. The previous discrete amplitudes are encoded in a memory kernel and can feed back into the initial level. At very short times $c_i(t) \simeq 1 - K(0)t^2/2$ when the second moment is finite, yielding a quadratic survival loss. At long times a sharply structured or bounded continuum can produce nonexponential decay. A finite number of exactly known final levels evolves unitarily under the full Hamiltonian and can show recurrences; treating the memory as identically zero after any arbitrarily short interval would discard this possibility.

When the spectral coupling varies slowly over the line and extends across a sufficiently broad band, phase cancellation makes $K(\tau)$ decay over a correlation time τ_c. For τ_c short relative to changes in c_i, write $c_i(t - \tau) \approx c_i(t)$ inside its memory integral after the initial transient, and extend its upper limit to infinity. The regulated integral $\int_0^\infty e^{-i\omega\tau - \eta\tau} d\tau = (\eta + i\omega)^{-1}$ tends to $\pi\delta(\omega) - i\,\mathcal{P}(1/\omega)$ as a distribution. Its real part gives

$$\dot{c}_i = -\Gamma c_i/2 + \cdots,$$

with the same golden-rule population rate Γ. The imaginary part gives a coherent energy shift proportional to $\mathcal{P}\sum_f |v_f|^2/(E_i - E_f)$, requiring the physical bandwidth and a consistent bare energy. This shows how an absorptive linewidth and a dispersive shift can emerge from different parts of the same elimination, without claiming a perfectly flat infinite continuum supplies a physically meaningful absolute shift.

For a symmetric model with constant spectral strength inside $-W < \omega_{fi} < W$, the kernel is proportional to $\sin(W\tau)/\tau$. Narrowing the band increases its memory time and can delay the onset of an approximately constant rate. Moving the discrete level close to a band edge also removes the cancellation between positive and negative detunings responsible for a vanishing symmetric shift. An experiment that varies the carrier across such an edge should fit the full finite-window spectrum before inferring a time-independent linewidth or level shift. A numerically cheap sum over known discrete final energies provides a check: double the

spectral cutoff, halve level spacing and track the early quadratic regime separately from any intermediate exponential segment.

Ask whether frequent projective tests inhibit a transition

For an initial pure state with finite energy variance under a time-independent H, expand the survival amplitude $A(t) = \langle i|e^{-iHt/\hbar}|i\rangle$ to second order. The linear term $-it\langle H\rangle/\hbar$ is a phase. Squaring gives

$$P_i(t) = 1 - (\Delta H)^2 t^2/\hbar^2 + O(t^3),$$

with still higher-order control depending on available moments. Define $t_Z = \hbar/\Delta H$. In the model with one initial level coupled only off-diagonally to the final levels, $(\Delta H)^2 = \sum_f |v_f|^2$, agreeing with the memory-kernel expansion. If the needed moments diverge in an idealised continuum, this short-time expansion requires a physical high-frequency regularisation.

Now make ideal instantaneous projective tests of whether the system is still in $|i\rangle$ every $\tau = T/N$, selecting only histories that have passed every test. Conditional on survival, each projection prepares exactly the same initial state. The joint survival probability is $P_i(\tau)^N \simeq [1 - (T/Nt_Z)^2]^N \to 1$ as $N \to \infty$ for fixed total duration T. This is a mathematically precise limiting model of the quantum Zeno effect. It is not obtained by simply looking at a stored record more often after a single final measurement; the successive physical interventions reset the conditional state.

With a resonantly coupled two-state Hamiltonian $H = v\sigma_x$, the survival against a single interval is $\cos^2(v\tau/\hbar)$. Choose $vT/\hbar = \pi/2$; unobserved survival at T is zero. For ten equally spaced perfect tests the joint survival is

$$\cos^{20}(\pi/20) = 0.7805.$$

If the instrument has 2% independent loss per successful test, the recorded joint fraction falls to

$$0.7805(0.98)^{10} = 0.638.$$

A finite-duration detector may also change H or broaden spectral coupling rather than implement an instantaneous projection. The ideal repeated-measurement result gives a comparison baseline, while predicted observed counts must include the physical instrument and conditioning.

Derive the sweep-rate scale for a transition through resonance

For a chirped weak drive let its instantaneous detuning from one discrete transition change linearly, $\Delta(t) = \dot{\Delta}\, t$, crossing zero at the middle of a long pulse. The phase accumulated in the first-order integral is $\int^t \Delta(t')dt' = \dot{\Delta}\, t^2/2$, not $\Delta(t)t$. For approximately constant coupling v around the crossing and sufficiently far endpoints the oscillatory Gaussian integral gives $|c_f^{(1)}|^2 \simeq 2\pi|v|^2/(\hbar^2|\dot{\Delta}|)$. A slower passage accumulates a larger resonant amplitude, but this weak-coupling formula cannot exceed probability one. At large transfer the full two-level avoided-crossing dynamics replaces it.

For $|v|/\hbar = 2\pi \times 200$ Hz and an ordinary-frequency sweep $d(\Delta/2\pi)/dt = 10^8$ Hz/s, the estimate is 0.0158. Changing only the sign of the sweep gives the same first-order probability in this symmetric idealisation, while the phase of the transition amplitude changes. Coherently driving a second passage can convert that phase difference into an observable interference pattern. Missing endpoints, asymmetric pulse shaping and additional nearby lines can spoil the simple symmetry. Comparing both sweep directions and checking that the result scales as inverse speed over a low-transfer range are concrete tests of the stationary-phase approximation instead of assuming every apparent line crossing signifies complete adiabatic transfer.

Separate a transition frequency from coherence loss

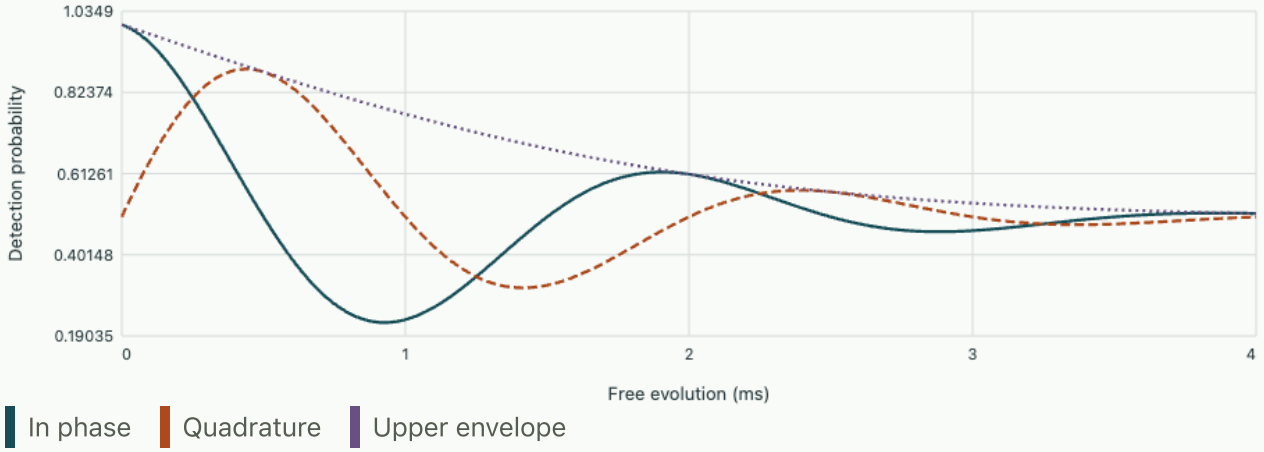
Vary detuning and shot-to-shot spread while reading both pulse quadratures. Determine whether a change of population comes from a shifted carrier or a fading phase.

Open this experiment online

Try the calculation

Mean cyclic detuning (Hz) **500**

Shot-to-shot rms spread (Hz) **80**



Contrast after 1 ms = 0.53455. The two probabilities at 0.50 ms are 0.5 and 0.87729.

Derive the spectral distinction between a phase-coherent pulse and stationary noise

A fluctuating perturbation is often characterised by a power spectral density rather than a prescribed pulse phase. Suppose $V_{fi}(t) = d_{fi}F(t)$ with F a real classical, zero-mean stationary random process and the coupling weak enough for first-order amplitudes in each realisation. Averaging the modulus squared produces a double time integral involving the autocorrelation $C_F(\tau) = \langle F(t+\tau)F(t)\rangle$:

$$\langle P_f(T)\rangle = \frac{|d_{fi}|^2}{\hbar^2}\int_{-T}^{T}(T-|\tau|)C_F(\tau)e^{i\omega_{fi}\tau}\,d\tau.$$

The factor $T - |\tau|$ counts the length of pairs of times in $[0, T]$ separated by τ. If T is long compared with the correlation time and depletion remains weak, divide by T to find rate $|d_{fi}|^2 S_F(\omega_{fi})/\hbar^2$, where $S_F(\omega) = \int_{-\infty}^{\infty} C_F(\tau) e^{i\omega\tau} d\tau$ is the two-sided angular-frequency spectrum. No sum over a continuum of final quantum states was involved here. A continuum of drive frequencies supplies a different physical route to an approximately linear transition probability of one resolved final state.

For exponentially correlated noise $C_F(\tau) = F_{\mathrm{rms}}^2 e^{-|\tau|/\tau_c}$, integration gives

$$S_F(\omega) = 2F_{\mathrm{rms}}^2 \tau_c / [1 + \omega^2 \tau_c^2].$$

Fix F_{rms} and vary correlation time. At a nonzero fixed transition frequency, extremely rapid fluctuations have little integrated spectral weight there because their fixed variance is spread over a very broad band. Very slow fluctuations concentrate their weight near zero frequency. Differentiating the spectral expression gives a maximum excitation rate at $\tau_c = 1/|\omega_{fi}|$. This nonmonotonic dependence cannot be inferred just by quoting the field's total rms amplitude. Unit conventions are also critical: changing to ordinary-frequency spectral densities and a one-sided instrument display requires the corresponding Jacobian and factor of two.

A real classical stationary process has an even spectrum $S_F(-\omega) = S_F(\omega)$, so this model drives upward and downward transitions equally per available source population when matrix elements are conjugate. A quantum environment at thermal equilibrium can have unequal positive and negative frequency unsymmetrised spectra. Its absorption and emission probabilities encode detailed balance. Thus an observed difference between calibrated upward and downward rates cannot generally be explained by merely drawing samples from a real classical noise spectrum of the same symmetrised power. Distinguishing a driven classical source from an equilibrium quantum bath requires precise statements about which correlations enter the transition probability.

Infer a fluctuation timescale from a rate ratio

Two transitions with equal coupling strengths see the same exponentially correlated classical field at ordinary frequencies f and $2f$. Their measured weak transition rates obey $\Gamma(2f)/\Gamma(f) = 0.40$. Infer the dimensionless correlation frequency and, for $f = 50.0$ kHz, the correlation time. What feature of the acquisition would invalidate this inference?

▼ **Derive and check the result**

Let $x = (2\pi f \tau_c)^2$. The ratio is $(1 + x)/(1 + 4x) = 0.40$, giving $0.60 = 0.60x$ and $x = 1$. The correlation time is

$$1/(2\pi f) = 3.18\,\mu\mathrm{s}.$$

At the lower frequency the noise spectrum lies at half its zero-frequency value; the higher line samples one fifth of that value. If the exposure time is comparable to or shorter than τ_c, the triangular window in the double integral still influences both rates and the infinite-time Lorentzian ratio is inadequate. Likewise unequal matrix elements or a finite observation band must be corrected separately. Measuring a third transition at another frequency supplies an independent test of the exponential-correlation model rather than simply determining its fitted timescale.

Tutorial · choose a pulse duration and check the perturbation

Measure a frequency width without confusing two line shapes

A rectangular pulse lasts 4.00 ms with resonant coupling $|v|/\hbar = 25.0\,\mathrm{s}^{-1}$. Find the line-centre probability, separation of the first zeros and full half-height width in hertz.

▼ **Work through the solution**

At resonance first-order probability is

$$(25 \times 0.004)^2 = 0.0100.$$

The first zeros are 250 Hz on either side, giving a 500 Hz separation. The full half-height width is

$$0.886/0.004 = 221.5\,\mathrm{Hz}.$$

A Lorentzian fit with its width interpreted as a spontaneous lifetime would mistake pulse broadening for decay. Doubling T while keeping coupling fixed is a diagnostic: the first-order peak would rise fourfold and the width halve. At large peak probability use the full two-level dynamics rather than extrapolating the quadratic indefinitely.

Verify the breakdown against exact two-level dynamics

Assume an isolated resonantly driven two-state system with rotating-frame Hamiltonian off-diagonal v and zero detuning. What replaces the first-order expression when $T = \pi\hbar/(2|v|)$? At which small pulse area is the fractional error approximately one percent?

▼ **Work through the solution**

The resonant Hamiltonian squares to $|v|^2 I$, so the exact probability is $\sin^2(|v|T/\hbar)$. At the stated pulse time it is unity while the first-order result is $(\pi/2)^2 = 2.47$, impossible as a probability. For pulse area $a = |v|T/\hbar$,

$$\sin^2 a = a^2 - a^4/3 + \cdots.$$

The relative discrepancy is approximately $a^2/3$, reaching 1% at $a \simeq 0.173$. The pulse can already make a measurable few-percent transition while first-order theory remains reasonably accurate. A dense continuum requires checking the total across all levels as well.

Infer an accessible density from the observed time dependence

At long-enough but weak-depletion times an ensemble loses 2.0% population in 5.0 ns. The squared box matrix element is calibrated as $2.5 \times 10^{-17}\,\mathrm{eV}^2$. Infer ρ, estimate the recurrence scale and test whether this interpretation is internally consistent.

▼ **Work through the solution**

The slope estimate is $4.0 \times 10^{6}\ \mathrm{s}^{-1}$. With $\hbar = 6.5821 \times 10^{-16}\,\mathrm{eV\,s}$ the inferred $\rho = \Gamma\hbar/(2\pi|v|^{2}) = 1.68 \times 10^{7}\,\mathrm{eV}^{-1}$. The corresponding spacing is 5.97×10^{-8} eV and recurrence scale $h/\delta E \simeq 69$ ns. A 5 ns observation fits within that scale, with about 14 spacings to the first zero. Two percent is compatible with weak depletion. Smoothness of the density and a single accessible channel still require independent spectral checks. Repeating the measurement at several durations reveals whether the inferred slope is stable.

Further reading

- Fließbach · Näherungsmethoden
- MIT 8.322 · time-dependent perturbation theory and scattering
- Course investigation · driven transitions and finite pulse spectra

Which part of a Bell correlation survives imperfect state preparation?

Computational quantum experiments

A source emits two spin-half systems in the singlet state. Two observers select measurement directions separately and record outcomes $+1$ or -1. Construct a Bell–CHSH statistic from four measurable correlations and determine how state preparation, angular alignment and finite sample size affect the prediction.

Relate a counted coincidence to an operator

Denote unit measurement directions by $\mathbf{a}, \mathbf{b}$. Outcome projectors are $(1 + s\,\boldsymbol{\sigma}\cdot\mathbf{a})/2$ and $(1 + t\,\boldsymbol{\sigma}\cdot\mathbf{b})/2$, for $s, t = \pm 1$. The singlet has zero individual spin expectation and tensor $\langle \sigma_i \otimes \sigma_j \rangle = -\delta_{ij}$. The diagonal entries follow because measurements along any common axis are exactly opposite and the state is rotationally invariant. Its joint probability is

$$p(s,t|\mathbf{a},\mathbf{b}) = \tfrac{1}{4}(1 - st\,\mathbf{a}\cdot\mathbf{b}),$$

$$E(\mathbf{a},\mathbf{b}) = -\mathbf{a}\cdot\mathbf{b}.$$

Summing over either outcome leaves one half for the other observer, independent of the remote angle. Covariance can change while the individual counting rates remain constant. At an angle of 60 degrees the probabilities of the two equal-outcome combinations sum to one quarter, and of opposite outcomes to three quarters. Their signed difference gives –0.5. The lab models spin-axis angles; linear polarisation analysers give a doubled angle in their effective two-state representation.

Derive a bound and choose a set of directions

If a trial had four predetermined local binary responses A, A', B, B', its combination $AB + AB' + A'B - A'B' = A(B + B') + A'(B - B')$ would have magnitude two, since exactly one of the parentheses vanishes. Averaging over a common distribution of such assignments preserves the bound. The quantum prediction combines expectation values measured on different subensembles with independently chosen setting pairs. Use coplanar angles $a = 0,\ a' = \pi/2,\ b = \theta,\ b' = -\theta$.

$$S = E(a,b) + E(a,b') + E(a',b) - E(a',b') = -2(\cos\theta + \sin\theta).$$

At $\theta = 45^\circ$ its magnitude is $2\sqrt{2}$. The derivative vanishes there, making small symmetric angular errors second order. If a fraction $1 - v$ of pairs is completely unpolarised, the density operator is $v|\Psi^-\rangle\langle\Psi^-| + (1 - v)I/4$. Every correlator gains a factor v. The best attainable magnitude exceeds two only when $v > 1/\sqrt{2}$. This is a statement about this measurement family and this noise model. A pair source and a scalar contrast alone do not define an experimental protocol.

Assess one set of finite counts

At $v = 0.80$, $\theta = 45^\circ$ the expected magnitude is about 2.263. Each trial product is a binary variable with variance $1 - E^2$. For N independent pairs per setting, the sample correlation variance is $(1 - E^2)/N$. Assuming independent subensembles, variance of the signed sum is the sum of the four variances. Here

every squared correlation is 0.32. With 1000 trials per setting the approximate standard error of the statistic is

$$\sqrt{4(1-0.32)/1000} = 0.0522.$$

The excess 0.263 is roughly five standard errors under this idealised independent-trial model. It is not a substitute for a finite-sample hypothesis test that accounts for an actual protocol's setting choices, losses, timing and possible memory.

Change settings and preparation independently

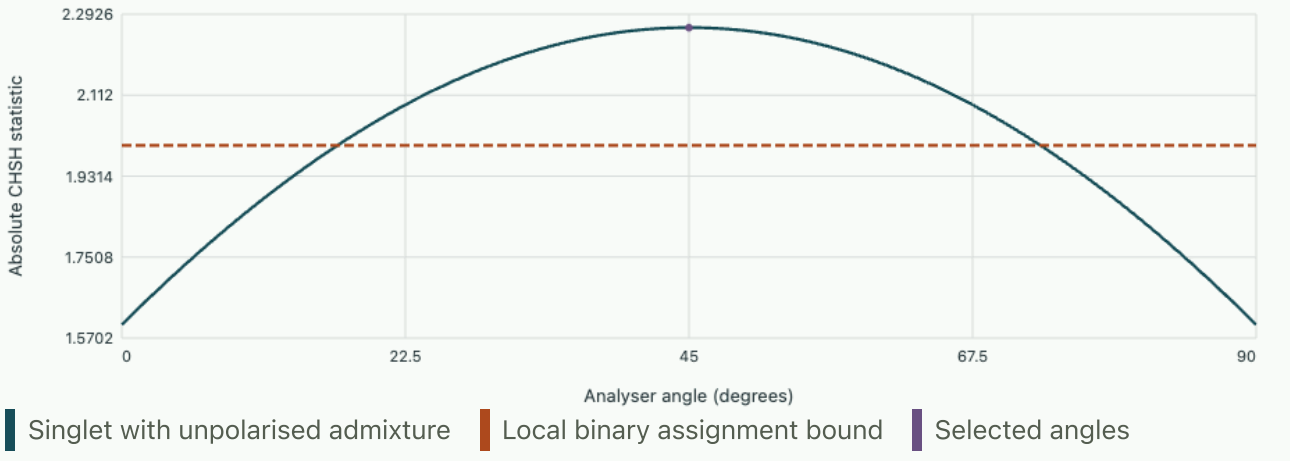
Inspect the predicted absolute statistic and the local assignment bound. Find the angles for which the chosen source can exceed two. Then vary visibility at a fixed angle. Small individual spin polarisation drifts could alter the marginal counts; unequal channel efficiencies could change a coincidence-selected sample. Monitoring all four joint count tables and both single-observer marginal distributions allows more checks than retaining only their final sum.

Open this experiment online

Try the calculation

Preparation visibility **0.8**

Symmetric analyser angle (degrees) **45**



Predicted |S| = 2.2627. With these angles the visibility required to reach two is 0.70711.

Tutorial

At visibility 0.70, can increasing the number of trials eventually make the expected maximum exceed two?

▼ **Work through the solution**

No. The largest mean value is $2\sqrt{2}(0.70) = 1.980$. More trials narrow the sampling distribution around that value. Changing the preparation or measurement family addresses a different question.

What is the predicted fractional loss at an angle δ away from the optimum, and how large is it at five degrees?

▼ **Work through the solution**

Write $\cos\theta + \sin\theta = \sqrt{2}\cos(\theta - \pi/4)$. The fractional loss is $1 - \cos\delta$, about 0.00381 at five degrees. Asymmetric offsets on different analyser axes need not enter the same way.

A paper gives one large value of $|S|$ but discards all pairs for which one detector was silent. Can the ideal bound derived above automatically be applied to these retained counts?

▼ **Work through the solution**

The derivation averaged all setting pairs over a common underlying ensemble. Setting- or state-dependent retention can alter those ensembles. One needs the trial definition, data handling and efficiency assumptions to interpret the reported statistic.

For state preparation and independent measurement on tensor factors, revisit the composite-state chapter.

Recover a transition frequency from two separated pulses

Computational quantum experiments

A two-level system is driven by two short pulses with a free evolution interval between them. Determine why a population measurement carries phase information. Use a scan at several delays to separate a constant detuning, intrinsic coherence loss and trial-to-trial frequency variation.

Follow the amplitudes through both pulses

Use basis $|g\rangle, |e\rangle$. Let the first pulse prepare $(|g\rangle + |e\rangle)/\sqrt{2}$ in the frame of the applied oscillator. A detuning Δ in radians per second accumulates a relative phase $\Delta\tau$ in free time τ. Immediately before the second pulse each state has population one half; measuring population here cannot reveal detuning. Choose a second pulse equivalent to projecting onto $(|g\rangle + e^{-i\phi}|e\rangle)/\sqrt{2}$. Then one output population is

$$P(\tau, \phi) = \tfrac{1}{2}[1 + \cos(\Delta\tau - \phi)].$$

The sign of the phase follows our definitions of detuning and the reference pulse. Shifting the reference phase makes it possible to check that sign. A measurement using only $\phi = 0$ cannot distinguish Δ from $-\Delta$. Coherence is the off-diagonal density-matrix element that the second pulse converts into an observable population difference.

Derive two envelopes with different origins

A homogeneous decay time T_2 multiplies the off-diagonal element by $e^{-\tau/T_2}$. If instead the detuning of each experimental shot is constant during that shot and drawn from a Gaussian distribution of mean $\bar{\Delta}$ and standard deviation σ_Δ, average the complex phase, not its absolute value. Completing the square in the Gaussian integral gives

$$\langle e^{i\Delta\tau} \rangle = e^{i\bar{\Delta}\tau} e^{-\sigma_\Delta^2 \tau^2/2}.$$

With both independent mechanisms the measured contrast is $C(\tau) = \exp[-\tau/T_2 - \sigma_\Delta^2\tau^2/2]$. One mechanism gives an initial logarithmic slope, the other adds curvature. A single low contrast measured at just one delay cannot establish which one caused it.

Infer parameters from quadratures and a frequency scan

Record the $\phi = 0, \pi/2$ outputs, then form $X = 2P_0 - 1 = C\cos(\bar{\Delta}\tau)$ and $Y = 2P_{\pi/2} - 1 = C\sin(\bar{\Delta}\tau)$. Their length estimates contrast and their angle estimates accumulated phase modulo 2π. Suppose at $0.50\,\mathrm{ms}$ the readings are 0.500 and 0.850. Then $X = 0,\ Y = 0.700$, and the phase is $\pi/2$ modulo 2π. With an independent bound $|\bar{\Delta}/2\pi| < 1\,\mathrm{kHz}$, the implied cyclic detuning is $+500\,\mathrm{Hz}$. Longer delays resolve fine frequency differences at the cost of phase ambiguity and declining contrast. Use a shorter-delay record to unwrap a longer one.

At a contrast of 0.700 the relation $-\ln(0.700) = \tau/T_2 + \sigma_\Delta^2\tau^2/2$ supplies one equation for two unknowns. Several delays, with independent error bars, are needed to identify the linear and quadratic terms. At very low

signal the length $\sqrt{X^2 + Y^2}$ has a positive noise bias. Fitting the two directly measured quadratures can be preferable to fitting the length.

Explore two mechanisms

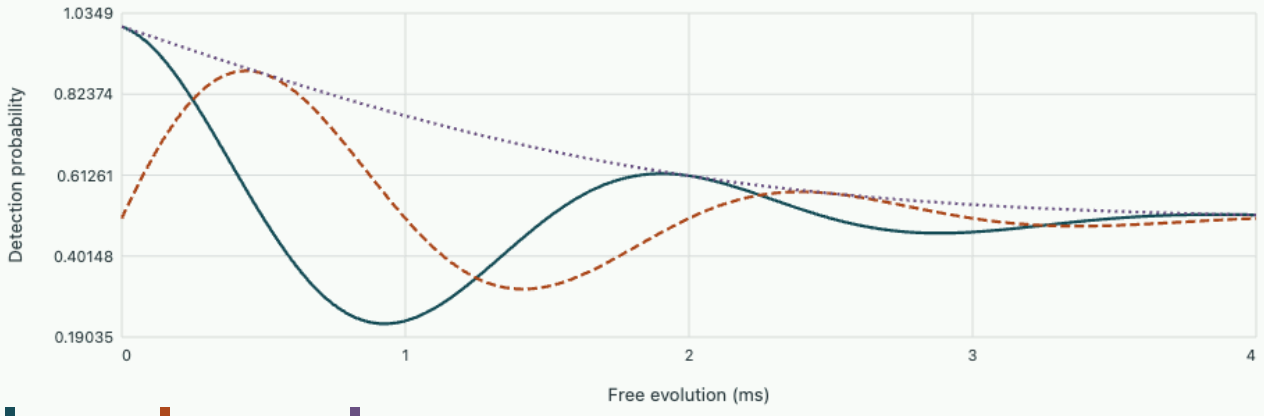
The graph holds the homogeneous T_2 at 2 ms. It lets you set signed cyclic detuning and shot-to-shot cyclic frequency spread. Compare both quadrature populations with the upper contrast envelope. First reverse the sign of the detuning, then increase the spread without changing the mean. Which effect interchanges peaks of one quadrature and which damps both?

Open this experiment online

Try the calculation

Mean cyclic detuning (Hz) **500**

Shot-to-shot rms spread (Hz) **80**



Contrast after 1 ms = 0.53455. The two probabilities at 0.50 ms are 0.5 and 0.87729.

Tutorial

What is the fringe spacing in a scan of cyclic detuning at fixed delay of 0.80 ms?

▼ **Work through the solution**

One full phase cycle needs $\delta f \, \tau = 1$. The spacing is

$$1/0.00080 = 1250\,\mathrm{Hz}.$$

The angular-frequency spacing is 2π times larger.

A fixed delay gives

$$P_0 = P_{\pi/2} = 0.5.$$

Does that show a phase of 45 degrees?

▼ **Work through the solution**

Both centred quadratures are zero, so the contrast is unresolved and the angle is undefined. A phase of 45 degrees with finite contrast gives equal, nonzero centred quadratures.

Explain how a refocusing pulse halfway through the evolution helps identify quasistatic detuning noise.

▼ **Work through the solution**

The pulse exchanges the phase-accumulating states, reversing phase winding for a detuning that stays constant through both halves. Random variations during the interval need not cancel. Recovering contrast in the refocused sequence supports a slowly varying contribution to the original envelope.

Compare a continuous drive in the transition chapter with the role of the second pulse here.

Mechanical and thermal effects of light

Calculate how a beam can translate a bead, couple resonances or exchange energy with a moving boundary. Test each proposal against a measured signal.

1. Optical trapping and Brownian calibration of a dielectric bead
2. Coupled optical and mechanical resonances
3. Thermal radiance, étendue and detector calibration
4. Moving-mirror Doppler metrology
5. Where momentum waits between source and receiver
6. Why a small transparent sphere scatters light
7. From wavefronts to rays, focusing and caustics

Optical trapping and Brownian calibration of a dielectric bead

Electromechanics, optics and measurement

An image of a bead remaining near a focus does not establish the strength of its trap. Derive the force pulling a small dielectric sphere into a bright region, compare it with the forward force from scattering and learn what a motion record can tell you about stiffness, drag and temperature. The model here uses an inert dielectric sphere in a homogeneous Newtonian liquid.

From two dielectric boundaries to the energy of an induced dipole

A sphere of radius a and relative permittivity ϵ_b sits in a liquid of relative permittivity ϵ_m. Assuming the sphere is small compared with a wavelength in the liquid, solve its local electric problem in an almost uniform applied field E. Rotational symmetry about the field admits an interior potential proportional to $r\cos\theta$. The exterior consists of the imposed potential $-Er\cos\theta$ and an induced dipolar term proportional to $\cos\theta/r^2$. Equality of potential and of normal displacement at the surface gives the local-field and dipole factors

$$E_{\mathrm{in}} = \frac{3\epsilon_m}{\epsilon_b + 2\epsilon_m} E,$$

$$p = \alpha E,$$

$$\alpha = 4\pi\epsilon_0\epsilon_m a^3 \frac{\epsilon_b - \epsilon_m}{\epsilon_b + 2\epsilon_m}.$$

The field doing the polarising work is the external field in the host at the particle position before inserting it. To turn that work into a potential, build up the dipole from zero rather than assuming a permanent dipole moment. Quasistatic assembly gives

$$-\int_0^E \alpha E'\, dE' = -\alpha E^2/2.$$

For a harmonic peak field average the square over an optical period. In a nearly lossless liquid of refractive index n_m, the peak field is related to travelling-wave intensity by $I = n_m\epsilon_0 c|E|^2/2$. The resulting conservative force is

$$U = -\frac{\mathrm{Re}\,\alpha}{4}|E|^2 = -\frac{\mathrm{Re}\,\alpha}{2n_m\epsilon_0 c} I,$$

$$\mathbf{F}_{\mathrm{grad}} = -\nabla U.$$

The two factors of two have different origins: induced-dipole assembly and time averaging of a sinusoid. Using a permanent-dipole potential with a peak field would miss both. A bead of higher refractive index than the liquid has positive real polarizability and is attracted to increasing intensity. A lower-index inclusion reverses this local tendency. For large particles, strong absorption or rapidly structured fields, a dipole with real static polarizability no longer specifies the force; a field or multipole calculation of the complete momentum balance is required.

Investigate the small-sphere dipole response and its dependence on particle size

Compute a transverse stiffness and test how local it is

Assume the beam near its focal plane has intensity $I(x) = I_0 e^{-2x^2/w^2}$, with power P, waist w defined at $1/e^2$ intensity and $I_0 = 2P/(\pi w^2)$. Put $U_0 = \mathrm{Re}\,\alpha I_0/(2n_m\epsilon_0 c) > 0$. The potential measured above its minimum is $U_0[1 - e^{-2x^2/w^2}]$. Expanding to second order and equating its leading term to $\kappa_x x^2/2$ gives

$$\kappa_x = 4U_0/w^2.$$

Stiffness therefore scales linearly with power, as the third power of sphere radius and as the inverse fourth power of waist at fixed indices. Each scaling applies within the stated approximations.

For a 40 nm radius bead with $n_b = 1.59$, take $n_m = 1.33$, $w = 0.65$ micrometres, vacuum wavelength 1064 nm, $P = 90$ mW and $T = 298$ K. The index contrast factor is 0.1252 and the waist intensity is 1.36×10^{11} watts per square metre. The predicted depth 3.03×10^{-20} J is 7.36 times the thermal energy. The lateral stiffness is 2.87×10^{-7} N/m. A displacement of 50 nm then requires approximately 14.3 fN of transverse balancing force. At a displacement of half a waist, the exact restoring force is the harmonic estimate multiplied by $e^{-1/2} = 0.607$; a Hookean calibration cannot be extrapolated that far.

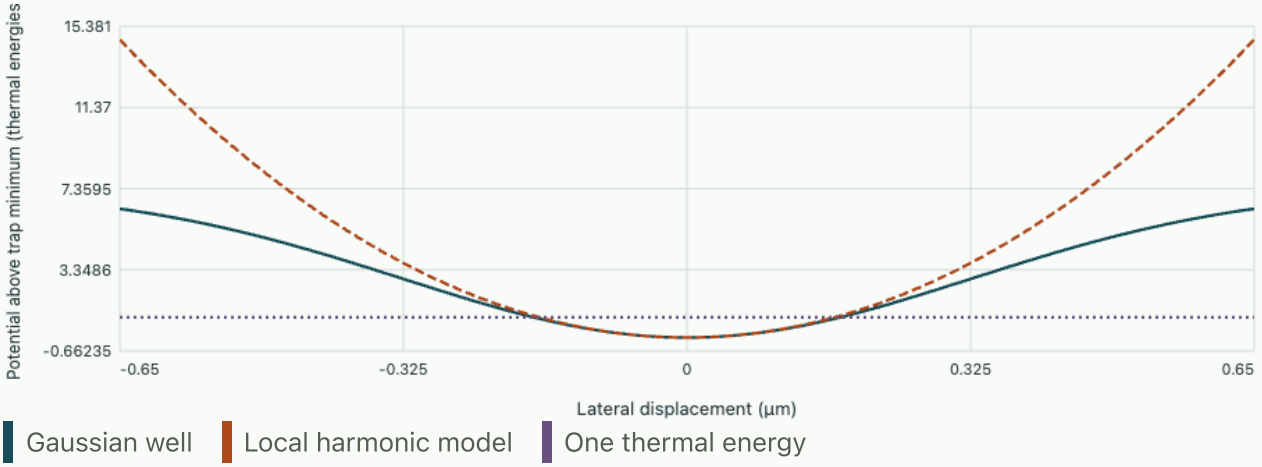
Open the live experiment

Try the calculation

Bead radius (nm) **40**

Optical power (mW) **90**

Beam waist (µm) **0.65**



The well depth is 7.36 kBT. The local model gives an rms displacement of 119.8 nm and a corner frequency of 60.5 Hz at 298 K.

Use the two curves to identify the interval in which the local quadratic energy is within ten percent of the Gaussian prediction. First lower power at fixed waist. Both curves shrink together in energy units while the harmonic thermal displacement grows. Next increase bead radius at fixed power and watch the well deepen, but also evaluate the size parameter $2\pi n_m a/\lambda_0$. It is 0.314 at the default and grows linearly with radius. Stronger trapping predicted by this simplified model does not establish the accuracy of the dipole approximation at arbitrary sizes.

Compare axial restoring force with radiation pressure

An induced dipole reradiates a portion of the incident field. In a uniform host use the Rayleigh scattering cross section $\sigma_s = (8\pi/3)k_m^4 a^6 |(\epsilon_b - \epsilon_m)/(\epsilon_b + 2\epsilon_m)|^2$. In a common travelling-wave momentum-flux model for a bead in a stationary nondispersive liquid the axial scattering force is approximately $F_s = n_m I \sigma_s / c$; absorption would add a further contribution. This estimate uses the total scattering cross section because the leading dipole pattern has vanishing mean forward cosine. For a larger forward-scattering sphere a transport cross section weights the deflected momentum instead.

In the default case the cross section is roughly 2.05×10^{-18} square metres. The forward force at the waist is about 1.23 fN. A Gaussian axial profile has $I(0,z) = I_0/[1 + (z/z_R)^2]$ with $z_R = \pi n_m w^2/\lambda_0 = 1.66$ micrometres. Its local axial stiffness $2U_0/z_R^2 = 2.20 \times 10^{-8}$ N/m is much smaller than its transverse stiffness. Balancing the constant forward force with that local stiffness predicts an equilibrium about 56 nm downstream of the waist. The small displacement relative to the Rayleigh length supports the linearisation.

Both the axial stiffness and forward scattering force scale with power in this simple model. A power sweep can thus increase confinement while barely shifting the equilibrium offset, a useful discriminating prediction. At fixed focus the offset grows rapidly with bead size because the scattering force scales with the sixth power and the gradient force with the third power. A stable lateral position alone does not establish stable three-dimensional capture. Check the axial restoring slope at a force zero and verify that thermal escape and fluid flow are negligible over the acquisition interval.

Extract stiffness from fluctuations or from a relaxation time

Near a stable point in one Cartesian direction, let x denote displacement from the mean and ζ the viscous drag. At sufficiently low frequencies the inertial term is negligible and the overdamped Langevin equation is

$$\zeta \dot{x} = -\kappa x + \xi.$$

At equilibrium the fluctuating force has zero mean and white covariance $2k_B T \zeta \delta(t - t')$ over the frequencies considered. The deterministic relaxation time is

$$\tau = \zeta/\kappa.$$

Multiplying the equation at positive delay by $x(0)$ and averaging removes the uncorrelated later noise and gives the decaying covariance. Equipartition sets its initial value.

$$C(t) = \frac{k_B T}{\kappa} e^{-|t|/\tau},$$

$$f_c = \frac{\kappa}{2\pi\zeta},$$

$$S_x^{(1)}(f) = \frac{4k_B T \zeta}{\kappa^2 + (2\pi f \zeta)^2}, \quad f \geq 0.$$

The stated spectrum is one-sided in ordinary frequency. Its integral from zero to infinity equals $C(0)$; doubling another time would double the variance. In an unbounded fluid with viscosity 1.00 mPa s, Stokes drag is $6\pi\eta a = 7.54 \times 10^{-10}$ N s/m. The default predicts $f_c = 60.5$ Hz and rms displacement 120 nm. The latter is roughly 0.18 waist, so the outer part of a long thermal trajectory may sample appreciable

anharmonicity. A finite-depth well in an unbounded bath does not have a globally normalisable trapped equilibrium. These local equilibrium predictions apply to a sufficiently long-lived, locally sampled trapped state.

If a voltage detector gives $V = gx$, the spectral corner and normalised autocorrelation time do not depend on the displacement gain g. They determine only κ/ζ. Equipartition using a calibrated spatial variance gives κ independently of drag, provided mechanical fluctuations are in equilibrium and sufficiently resolved. If the viscosity or nearby-wall drag is uncertain, combining a known stage oscillation with the Brownian spectrum can determine further calibration quantities. Keep a stage-driven coherent peak out of the noise integral used for equipartition.

Account for exposure, sample size and unobserved motion

A camera typically reports the position averaged during an exposure of duration t_e. Its variance is the double integral of the covariance over the exposure interval divided by t_e^2. Evaluating the exponential local model and introducing $u = t_e/\tau$ gives a multiplicative factor $2(u - 1 + e^{-u})/u^2$. At $u = 1$ the reported physical-motion variance is only 0.736 of the instantaneous variance. Interpreting it as the full variance would overestimate stiffness by approximately 36 percent. Independent pixel-localisation noise adds additional variance with the opposite sign of bias.

Estimate that localisation noise with an effectively immobilised reference bead at a comparable brightness. Test the averaging correction by changing exposure duration at constant sampling frequency where possible. A camera that reports overlapping exposures has correlated measurement errors and needs a different covariance model. A long series of closely spaced frames also contains fewer independent thermal configurations than its frame count suggests. Compare estimates across several blocks substantially longer than the relaxation time; residual low-frequency drift can dominate even when high-frequency position precision looks excellent.

A separate check uses a known uniform liquid flow. At a steady small displacement x_s, force balance gives

$$\kappa x_s = \zeta u_{\mathrm{flow}}.$$

For the default drag and stiffness a flow of 10 micrometres per second gives an offset of 26.3 nm. Repeat with reversed flow and use half the displacement difference to cancel an unknown zero-force position. Near a wall, the local flow at the bead and wall-corrected drag may differ from a stage velocity and the unbounded Stokes value. Those effects alter the mechanical calibration even if the optical field is unchanged.

A known stage oscillation calibrates position in a fluid

Brownian fluctuations alone determine the product of several unknown calibration parameters. A weak sinusoidal translation of the sample cell adds a known mechanical signal at a chosen frequency. Work in laboratory coordinates with a fixed optical trap and with fluid moving uniformly at velocity $u(t)$ near the bead, well away from walls in the idealised model. The drag is now $-\zeta(\dot{x} - u)$. Superposing a stage displacement $A\cos\omega t$ on thermal fluctuations gives a deterministic response whose complex amplitude satisfies $(\kappa - i\omega\zeta)\widetilde{x} = -i\omega\zeta A$ in the negative-exponential convention. Therefore its amplitude ratio is

$$\frac{|\widetilde{x}|}{A} = \frac{\omega\tau}{\sqrt{1+\omega^2\tau^2}}.$$

At the trap corner the bead displacement amplitude is the known stage amplitude divided by the square root of two. A calibrated 80.0 nm stage oscillation there produces a predicted bead amplitude of 56.6 nm. If a position-sensitive receiver records a coherent peak amplitude of 0.170 V, its displacement gain is 3.01 mV/nm. At very low drive frequencies the fluid velocity is small for fixed stage amplitude and the trap holds the bead nearly stationary. At high enough frequency the bead follows the fluid displacement within the overdamped model, but finite hydrodynamic, optical and electronic bandwidths eventually matter.

After determining the gain, convert the surrounding voltage spectrum into spatial units and use the area of the stochastic fluctuation spectrum to find stiffness. Exclude the calibrated line and account for the finite spectral averaging window; a narrow drive that leaks into neighbouring bins can inflate the inferred thermal variance. Dividing measured stiffness by the independently determined corner rate then gives drag, which can be compared with the known viscosity and sphere radius. This avoids assuming free-space Stokes drag in the first step of a near-wall measurement.

There are two limiting systematic effects that a good fit to a single drive frequency cannot resolve. A frequency-dependent stage calibration changes the actual fluid displacement, and motion of the optics relative to the cell generates a direct detector background even with the bead immobilised. Repeat the drive above and below the corner, compare coherent phase as well as amplitude and acquire a trapped and immobilised reference. Correct conversion should explain the entire complex response with one spatial gain and one corner time. A single calibration number obtained at one frequency can hide a phase lag in the stage controller or in the optical position channel.

One last cross-check reverses the stage drive phase by half a period. The coherent bead response must reverse, while thermal background power remains unchanged. A component that fails this reversal could come from ambient vibration unrelated to the stage. Averaging the two traces with and without a sign change separates these contributions before either is used in an optical-force calibration.

Tutorial · Separate a unit-conversion error from excess noise

In a single-axis trap at 298 K a calibrated detector reports a variance of 2500 square nanometres after removal of localisation noise and exposure correction. An independent spectral fit gives a corner of 175 Hz. Infer stiffness, drag and the equivalent unbounded hydrodynamic radius for a liquid of viscosity 1.00 mPa s. Compare against a nominal 80 nm radius bead.

▼ Infer the three quantities and test their consistency

A nanometre squared is 10^{-18} square metres, so the variance is 2.50×10^{-15} square metres. Equipartition gives 1.65×10^{-6} N/m. Dividing by $2\pi f_c$ gives drag 1.50×10^{-9} N s/m and radius $\zeta/(6\pi\eta) = 79.5$ nm. The agreement supports the local overdamped description under the assumed temperature and viscosity. Forgetting the factor of 2π would predict a radius roughly six times larger and could be mistaken for a huge near-wall correction.

Now imagine that additional uncorrelated read noise had been left in the variance. This raises apparent variance, lowers the equipartition stiffness and would lower the inferred drag at the same fitted corner.

Nearby-wall drag for a given radius instead normally raises hydrodynamic resistance for parallel translation. Their effects on the inferred radius thus differ in sign under the stated assumptions. Varying height and independently calibrating the measurement noise are appropriate diagnostics before attributing a discrepancy to bead dimensions.

Tutorial · Test a force-law inference with power and concentration controls

At fixed bead radius a trap is operated at two powers differing by a factor of two. Measured corner frequencies increase from 42 to 84 Hz, but rms displacement falls only from 110 to 87 nm. What does the harmonic model with fixed drag and temperature predict? Estimate a power-independent additive position-noise variance that could reconcile the two readings.

▼ **Use two variances to separate scale and offset**

The doubled corner gives doubled stiffness if drag is unchanged. Physical thermal variance should halve. The predicted rms at higher power would be 77.8 nm if the lower-power value were entirely physical. Let the low-power measured variance be $v + b$ and the high-power one $v/2 + b$, where b is a constant additive noise variance. Solving from 12,100 and 7,569 square nanometres gives $v = 9062$ and $b = 3038$ square nanometres, or approximately 55.1 nm rms noise.

This is a large inferred detector noise and demands an independent immobilised reference check. Power-dependent heating of the liquid changes temperature and viscosity, so an unchanged product inferred from only two quantities cannot prove that heating is absent. Test several powers, reverse their order to reveal slow drift and vary bead concentration to exclude overlapping particles in the tracking window. The inferred additive term should remain constant across those controls if it comes from the specified detector model.

Tolić-Nørrelykke and colleagues · position and force calibration using thermal and driven measurements

Coupled optical and mechanical resonances

Electromechanics, optics and measurement

A vibrating mirror can imprint its position on an optical field, while the stored light pushes back on the mirror. An apparent pair of peaks can indicate two hybrid modes, interference at a readout port or two unrelated resonances. Starting from energy in a moving cavity, derive the coupled response and learn which measurements establish a coupling strength and its uncertainty.

From a moving boundary to a force per photon

Take a nearly empty Fabry–Pérot cavity of optical length L with one moving end. A longitudinal resonance has $\omega_c = m\pi c/L$ for integer m, neglecting frequency-dependent mirror phase. For a small extension x, $\delta\omega_c = (\partial\omega_c/\partial x)x$ with signed slope $G_x = -\omega_c/L$. At fixed photon number n the optical energy is $n\hbar\omega_c(x)$. Its derivative gives the radiation-pressure force on the coordinate. For this mirror $F = -n\hbar G_x > 0$ pushes outward. The derivative must be at fixed photon number. Differentiating a driven steady-state photon number as though it were a closed-system energy would incorrectly include work exchanged with the pump.

For a 10 mm cavity resonant near 1064 nm the frequency slope magnitude in ordinary frequency is about 28.2 MHz per nanometre. A static shift of 10 pm therefore changes the optical resonance by approximately 282 kHz. One photon exerts a mean force $\hbar\omega_c/L \simeq 1.87 \times 10^{-17}$ N. An equivalent classical derivation counts a momentum change twice per normal reflection for each relevant round trip. The two descriptions yield the same mean force. The mirror displacement is the properly normalised mode coordinate; in a flexural device the optically sampled position and the total physical mass are generally different from centre-of-mass displacement and mass.

Introduce a single mechanical eigenmode with effective mass m_{eff}, angular frequency Ω_m and displacement scale $x_{\mathrm{zpf}} = \sqrt{\hbar/(2m_{\mathrm{eff}}\Omega_m)}$. The single-photon coupling magnitude is $g_0 = |G_x|x_{\mathrm{zpf}}$. This normalisation is convenient even when the observed vibrations are entirely classical. For $m_{\mathrm{eff}} = 10$ ng and $\Omega_m/(2\pi) = 1.0$ MHz, $x_{\mathrm{zpf}} = 9.16 \times 10^{-16}$ m and $g_0/(2\pi) = 25.8$ Hz for the stated optical geometry. The small displacement scale explains why a coherent pump is normally used to make the interaction measurable.

How a laser couples frequencies that are otherwise far apart

An optical carrier oscillates at hundreds of terahertz, whereas the mirror may vibrate at only megahertz. Direct comparison of these two bare frequencies would never suggest a resonant exchange. A strong coherent pump supplies a phase reference and converts a mechanical displacement into optical sidebands. Move into a frame rotating at the pump frequency and write the cavity amplitude as its large steady part plus a small perturbation. The leading frequency-shift interaction is proportional to photon number times displacement. Expanding the intensity around the large pump keeps terms linear in both the probe field and displacement. The resulting coupling is $G = g_0\sqrt{n_c}$, where n_c is mean intracavity pump photon number.

Tune the pump about one mechanical frequency below the optical cavity resonance. In the resolved-sideband and weak-fluctuation limit a near-resonant process moves one excitation between the mechanical mode and an upper optical sideband. The rapidly rotating opposite-sideband terms average out. Now the two coupled frequencies are an optical sideband offset and the mechanical frequency, measured relative to a common rotating reference. The large carrier frequency has already been accounted for by the pump. The pump phase can be absorbed into the mode definitions, so G may be chosen positive for this isolated pair.

If instead the pump is placed near the upper sideband, the surviving interaction creates or annihilates excitations in pairs. The passive exchange equations below cease to apply and strong pumping can produce amplification or instability. A bare two-oscillator machine driven mechanically near a shared natural frequency often has analogous exchange equations without a carrier pump. Identical matrices can describe different energy supplies; identify the drive and rotating reference before assigning a literal photon-to-phonon energy conversion.

Derive complex poles and the effect of unequal damping

Let a, b be small energy-normalised optical and mechanical complex amplitudes. In a frame where the uncoupled optical sideband has zero frequency, let the mechanical detuning be δ. Let κ, Γ denote energy linewidths in angular frequency, so amplitude decay rates are half these values. With an optical probe s at offset ω, the linear system is

$$\begin{pmatrix} \kappa/2 - i\omega & iG \\ iG & \Gamma/2 - i(\omega - \delta) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} s \\ 0 \end{pmatrix}.$$

Zeros of the determinant give autonomous complex mode frequencies. In the lossless case the eigenvalues are $\delta/2 \pm \sqrt{G^2 + (\delta/2)^2}$. Far off resonance each branch approaches a bare mode. At zero detuning the two eigenvectors have equal optical and mechanical weights and the splitting is $2G$. The eigenvectors rotate continuously through the avoided crossing, an important discriminator from two uncoupled lines simply passing through one another.

At zero detuning with losses, completing the square in the determinant gives

$$\omega_{\pm} = -i(\kappa + \Gamma)/4 \; \pm \sqrt{G^2 - (\kappa - \Gamma)^2/16}.$$

The real parts separate once $G > |\kappa - \Gamma|/4$. The shared imaginary part above that threshold corresponds to a hybrid energy linewidth $(\kappa + \Gamma)/2$. Resolving two intensity features on a given instrument is a stronger, readout-dependent condition. The dimensionless cooperativity $C = 4G^2/(\kappa\Gamma)$ is yet another measure, comparing interaction against the product of dissipation scales. A value greater than one does not automatically prove two resolved normal-mode poles in a device with very unequal linewidths.

Use frequency numbers divided by 2π consistently from here for examples. With optical linewidth 240 kHz, mechanical linewidth 40 kHz and coupling 160 kHz, the real pole separation is

$$2\sqrt{160^2 - 50^2} = 304\,\mathrm{kHz}.$$

Both hybrid energy linewidths are 140 kHz, and cooperativity is 10.7. If coupling falls to 45 kHz, cooperativity is still about 0.844 and no real zero-detuning pole separation appears. At 50 kHz the square root reaches zero. Near that threshold even a small linewidth or detuning calibration error has a large effect on a coupling estimate inferred from a single splitting.

Compute what an optical or mechanical receiver actually sees

Invert the matrix. With $D = (\kappa/2 - i\omega)[\Gamma/2 - i(\omega - \delta)] + G^2$, the small amplitudes are $a = s[\Gamma/2 - i(\omega - \delta)]/D$ and $b = -iGs/D$. The poles are shared. Their numerators are different, so the two channels can show different apparent maxima. A port in reflection also has a directly reflected field that interferes with the leaking intracavity amplitude. Its intensity can display a narrow dip or asymmetric feature without a matching resonance-height pattern inside the cavity. A legitimate forward model must describe the measured port.

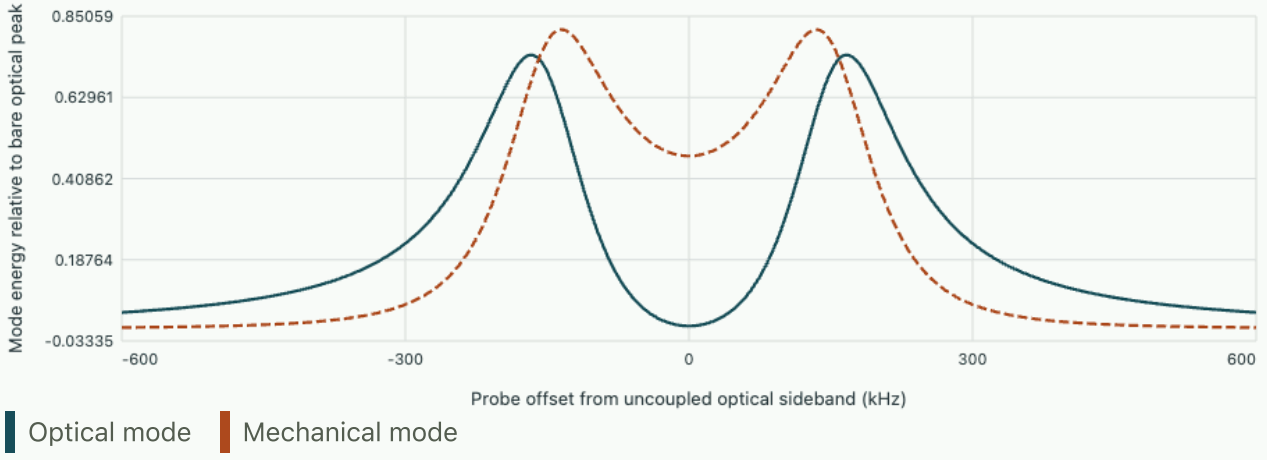
Open the live experiment

Try the calculation

Coupling G/(2π) (kHz) **160**

Optical linewidth κ/(2π) (kHz) **240**

Mechanical detuning (kHz) **0**



At zero bare detuning the real pole separation is 304.0 kHz. For the selected coupling and linewidths the cooperativity is 10.67. The mechanical linewidth is fixed at 40 kHz.

The graph plots the internal squared amplitudes relative to the bare on-resonance optical probe energy. Set the mechanical detuning to zero and reduce the coupling through 50 kHz. Compare the optical dip with the predicted real pole separation. Then set a substantial positive detuning and follow the transfer of mechanical response onto one branch. As a control set the coupling exactly to zero. The optical curve recovers the bare cavity Lorentzian and the mechanical probe-induced response vanishes. A thermally excited mechanical background, absent from this driven calculation, could remain in a real detector.

At zero detuning and zero probe offset,

$$a/a_{\mathrm{bare}} = 1/(1 + C).$$

For the default parameters the internal optical probe intensity at the centre is only $1/(1 + 10.67)^2 = 0.00735$ of the bare peak. This deep feature alone still does not establish the pole spacing. An incoherent background, a frequency-dependent detector transfer or coherent feedthrough can

alter peak and dip positions. Fit complex quadratures across a detuning series when available, and verify bare linewidths by independent ringdown or far-detuned spectra.

Check how energy linewidth and amplitude ringdown are related for one isolated resonance

Watch energy exchange in time and identify the linewidth convention

In an ideal equal-linewidth system starting with optical perturbation $a(0) = a_0$, $b(0) = 0$ at zero detuning, change to symmetric and antisymmetric modes. Recombination gives

$$a(t) = a_0 e^{-\kappa t/2} \cos Gt,$$

$$b(t) = -i a_0 e^{-\kappa t/2} \sin Gt.$$

Their normalised total energy decays as $e^{-\kappa t}$, while energy moves back and forth between optical sideband and vibration. The first full transfer occurs at $t = \pi/(2G)$ if the rotating-wave approximation holds. For $G/(2\pi) = 100$ kHz, this is 2.50 microseconds. With a shared energy linewidth of 20 kHz in ordinary frequency the surviving energy fraction then is

$$e^{-2\pi(20\ \mathrm{kHz})(2.50\ \mu\mathrm{s})} = 0.730.$$

A gated excitation shorter than the exchange period but long enough to avoid unwanted modes can prepare an approximately optical initial perturbation. Independent detection of an optical revival and a delayed mechanical response tests coherent exchange more strongly than the presence of two peaks alone. Comparing times of maxima directly with one half the reciprocal measured optical linewidth misidentifies the exchange physics. Check the decay of the sum of appropriately calibrated energies and the relative phase of the exchange. Electrical or optical feedthrough can produce a prompt apparent response with no mechanical build-up.

Infer a coupling with propagated measurement errors

When the two hybrid pole frequencies have been obtained from a justified fit and zero bare detuning is established, the coupling follows from the measured real separation S:

$$G = \sqrt{S^2/4 + (\kappa - \Gamma)^2/16}.$$

All quantities here can be expressed in the same ordinary-frequency units. If $S = 300 \pm 8$ kHz, $\kappa = 240 \pm 10$ kHz and $\Gamma = 40 \pm 4$ kHz, the estimate is 158.1 kHz. Independent small errors give

$$\partial G/\partial S = S/(4G) = 0.474,$$

$$\partial G/\partial \kappa = -\partial G/\partial \Gamma = (\kappa - \Gamma)/(16G) = 0.0791.$$

Combining in quadrature gives about 3.9 kHz uncertainty. A common frequency calibration can correlate all three inputs, and a fitted detuning and splitting from one scan may also be correlated. Use their covariance rather than multiplying independent error bars blindly.

A second inference varies intracavity pump power. Since $G^2 = g_0^2 n_c$, plot the squared fitted coupling against independently calibrated n_c. A nonzero intercept could indicate an additional direct coupling, baseline fitting bias or a mode incorrectly assigned at low power. The linewidths themselves can vary with pump power because of heating or backaction, so fit or measure them across the scan. Incident optical power is not identical to stored photon number; cavity coupling, detuning and loss determine their relation.

A measured frequency shift and linewidth change reveal optical feedback

The same instrument can be operated below the resolved splitting regime and read out a dispersive shift of the mechanical response. Algebraically eliminate the optical amplitude in the two-mode equations while driving the mechanical mode weakly. Its inverse susceptibility acquires a self-energy contribution proportional to the optical susceptibility. Let x be probe frequency measured from the bare optical sideband and use the same linewidth convention as above. In the mechanical denominator the additional term is

$$\Sigma(x) = \frac{G^2}{\kappa/2 - ix} = \frac{G^2(\kappa/2 + ix)}{(\kappa/2)^2 + x^2}.$$

Its real part adds to the amplitude damping, while its imaginary part shifts the resonant condition. At very weak coupling one can evaluate it at the bare mechanical frequency $x = \delta$, because variation across the narrow mechanical peak is small. Near an avoided crossing or when the mechanical peak samples a large change in optical susceptibility, solve the coupled poles directly; simply adding a linewidth and shift then neglects the very hybridisation being tested.

In ordinary-frequency units choose $G = 10$ kHz, optical linewidth 200 kHz, mechanical linewidth 2 kHz and detuning $+100$ kHz. At that offset the optical amplitude half-width is 100 kHz, giving $\mathrm{Re}\,\Sigma = 0.50$ kHz and $\mathrm{Im}\,\Sigma = 0.50$ kHz. The full mechanical linewidth increases from 2.0 to approximately 3.0 kHz. Its centre moves upwards by about 0.50 kHz with the adopted matrix signs. Moving to minus 100 kHz reverses the shift and preserves the added damping in this symmetric model. At zero detuning the predicted shift vanishes but the added linewidth is largest.

A scan on both sides is more discriminating than measuring broadening alone. A background rise in mechanical temperature changes integrated thermal noise and may change intrinsic loss, but need not reproduce this odd shift and even broadening pattern. Measure intrinsic linewidth far from the optical response and subtract that baseline. An unknown constant frequency offset can be removed with the two symmetric detunings, but an optical power dependence of cavity frequency can move the actual symmetry centre. Monitor the optical resonance during the scan.

The effective susceptibility describes classical feedback through the delayed optical field. It does not by itself specify the final mechanical temperature. A bath occupation or noise spectrum is also required to convert a damping change into an energy distribution. A driven response peak that broadens and lowers in amplitude is evidence for a changed susceptibility; thermal cooling additionally requires a calibrated measurement of the fluctuating mechanical displacement and its optical transduction.

Tutorial · A high cooperativity without resolved modes

A device has ordinary optical linewidth 1.00 MHz, mechanical linewidth 1.0 kHz and pump-enhanced coupling 30 kHz. Calculate cooperativity, compare the coupling with the real-splitting threshold and estimate the relative intracavity probe intensity at the central frequency. What observation would distinguish a narrow interference window from two separated poles?

▼ **Evaluate the three criteria independently**

In kHz units the cooperativity is

$$4(30)^2/(1000 \cdot 1) = 3.6.$$

The real-splitting threshold is $999/4 = 249.75$ kHz, far above the actual coupling. Yet the zero-detuning centre intensity is $1/(1+3.6)^2 = 0.0473$ relative to the bare optical peak. A narrow, deep structured feature is possible with no separated real pole frequencies in this model. The linewidth contrast permits mechanical response to interfere strongly within a broad optical susceptibility.

Acquire both amplitude and phase while scanning the absolute mechanical detuning and compare a coupled complex-response model with independently measured bare parameters. A pair of well-resolved poles should trace two consistent trajectories in both observables and can generate underdamped exchange in the time domain. A single optical intensity scan containing two flanks around a dip is insufficient to make that inference.

Tutorial · Diagnose an erroneous estimate of the single-photon coupling

An independent linewidth analysis gives pump-enhanced couplings 80 and 160 kHz at photon numbers 4.0×10^6 and 1.6×10^7. Estimate $g_0/(2\pi)$ and predict the coupling at 9.0×10^6 photons. A calibration revision doubles all inferred photon numbers while leaving the measured spectra unchanged. Which conclusions move?

▼ **Use coherent amplitude scaling and separate calibration from spectroscopy**

Divide the respective couplings in ordinary frequency by the square roots of their photon numbers. Both give 40 Hz and the intermediate prediction is 120 kHz. Doubling all inferred photon numbers lowers the inferred single-photon coupling by $1/\sqrt{2}$, to 28.3 Hz. The observed enhanced couplings, pole locations and cooperativities computed from measured linewidths stay the same. A prediction labelled by fixed physical photon number changes; a prediction at a repeatable laser setting can still be calibrated empirically.

To test the absolute calibration, compare the inferred single-photon value with the independent cavity frequency slope times the mechanical zero-point displacement. This requires an effective mass appropriate to the measured displacement and the actual cavity optical overlap. An incorrect choice of mechanical mode normalisation can mimic an error in stored photon number. Agreement across a power ratio tests the square-root scaling but cannot by itself determine the absolute optical or mechanical calibration.

Aspelmeyer, Kippenberg and Marquardt · cavity optomechanics, linearisation and strong coupling

Thermal radiance, étendue and detector calibration

Electromechanics, optics and measurement

A square millimetre of a heated object emits thermal radiation into an entire hemisphere. A detector sees only the rays accepted by its optics and only wavelengths to which it responds. Derive the equilibrium spectrum, follow it through an optical collection system and infer what temperature or emissivity a calibrated reading can actually determine.

Count optical states in a large equilibrium cavity

Take a cubic airless cavity of volume V in thermal equilibrium at temperature T. For its bulk mode density we can use periodic boundary conditions. Allowed wavevectors are separated by $2\pi/L$ in each coordinate. Count a spherical shell in wavevector space, divide by the volume per state and include two independent transverse polarisations. Substituting $k = 2\pi\nu/c$ gives

$$g(\nu)d\nu/V = 8\pi\nu^2 d\nu/c^3.$$

Using standing waves in an ideally reflecting large box gives the same leading volume density when directions and polarisations are counted consistently.

One oscillator mode of frequency ν can hold $m = 0, 1, 2, \ldots$ photons. Its thermal occupation has probabilities proportional to $e^{-mh\nu/k_BT}$. Summing the geometric series and taking its logarithmic derivative gives

$$\langle m\rangle = 1/(e^{h\nu/k_BT} - 1).$$

Multiplying by the density of modes and energy per photon gives spectral energy density u_ν. The zero-point contribution of a harmonic oscillator is not part of the thermal power exchanged between these black surfaces. At low frequency the result approaches an energy k_BT per mode; extending that classical limit to arbitrarily high frequency would make the integrated power diverge.

An isotropic distribution of energy travelling at speed c produces a normal flux through a small aperture weighted by the direction cosine. Per unit solid angle its radiance is $B_\nu = cu_\nu/(4\pi)$. Transform to equal wavelength intervals using $B_\lambda d\lambda = B_\nu|d\nu|$. The two useful spectral forms are

$$B_\nu = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/k_BT} - 1},$$

$$B_\lambda = \frac{2hc^2}{\lambda^5}\frac{1}{e^{hc/(\lambda k_BT)} - 1}.$$

The wavelength expression has units of watts per square metre per steradian per metre of wavelength. To express it per micrometre multiply by one millionth. At 400 K and 10 micrometres the result is 33.56 W per square metre per steradian per micrometre. At 295 K it is 9.14 in the same units. Equal wavelength and equal frequency bins carry different widths, so their distributions reach their maxima at different converted wavelengths. A spectrum's peak should always name its horizontal variable and spectral-density convention.

Distinguish emitter power from brightness in a chosen direction

For a black surface with angle-independent radiance into its outward hemisphere, integrate $\cos\theta\, d\Omega$. The result is π, not 2π. The emitted spectral power per unit area is πB_λ. Integrate over frequency using $x = h\nu/(k_B T)$. The remaining integral $\int_0^\infty x^3/(e^x - 1)\, dx = \pi^4/15$ follows by expanding $1/(e^x - 1)$ as a sum of exponentials and integrating term by term. This yields Stefan–Boltzmann exitance $M = \sigma T^4$ and identifies the constant $\sigma = 2\pi^5 k_B^4/(15 h^3 c^2)$.

A black patch of 1.0 square millimetre at 400 K emits about 1.45 mW into its entire outward hemisphere over all wavelengths. The same patch at 295 K emits about 0.429 mW. Replacing one by the other changes all-wavelength hemispherical power by about 1.02 mW. This is not the power a small infrared receiver sees through a narrow filter. The integrated emitter number is valuable for energy conservation, but by itself contains no receiver area, direction, passband or optical transmission.

The surface spectrum of a body in thermal equilibrium has directional spectral emissivity equal to its absorptivity in the corresponding reciprocal optical modes under the usual reciprocal-material conditions. An opaque diffuse grey surface with emissivity ϵ in an isotropically illuminated background can be modelled by $\epsilon B_\lambda(T) + (1 - \epsilon) B_\lambda(T_{\mathrm{bg}})$. It emits less than a black surface but also reflects background. A shiny object viewed in a warm room may consequently display a radiance temperature nearer the reflected scene than its own physical temperature. Specular surfaces require explicit directional knowledge of that scene.

Derive the geometric acceptance without counting the angle twice

An area element dA normal to the collection axis emits into $d\Omega$ at angle θ a spectral power $B_\lambda \cos\theta\, dA\, d\Omega$. For a uniformly filled cone of half-angle α and a uniform patch of area A, integrating gives $A\pi \sin^2\alpha$. This is its geometrical throughput in air when there is no vignetting. Optical étendue more generally uses $n^2 dA \cos\theta\, d\Omega$. A lossless passive optical relay in geometrical optics preserves this phase-space measure; correspondingly L/n^2 is conserved along a ray in transparent media under the usual stationary conditions.

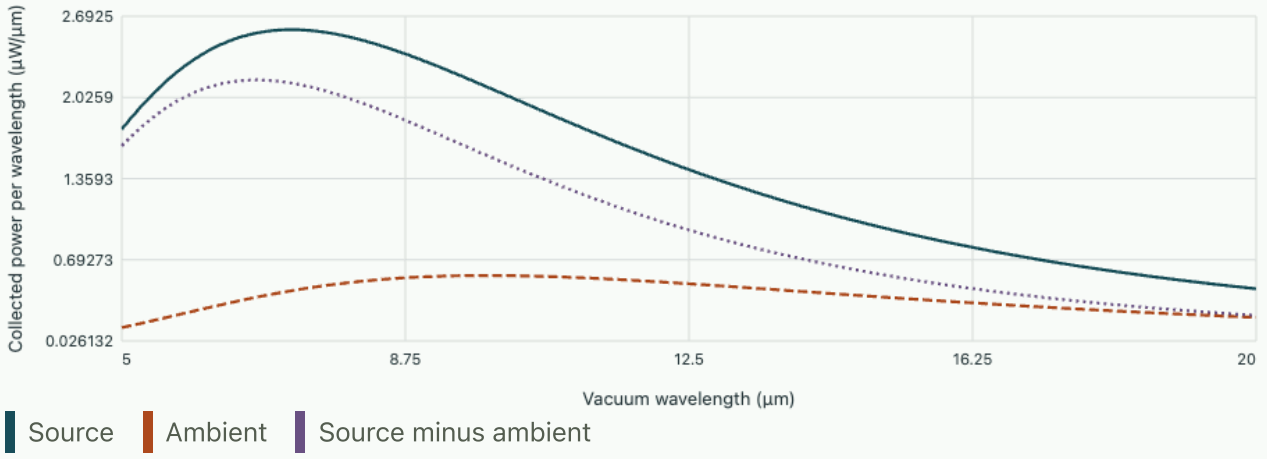
A direct lens example makes the restriction tangible. An image relay enlarging a uniformly bright patch by a factor of two in each transverse direction forms an image four times the area. In a lossless paraxial design filling the original pupil, each image point then occupies one quarter the solid angle. A detector covering the whole enlarged image and accepting that cone can receive the original power. Placing a fixed small sensor at the image plane intercepts less area instead. Higher magnification has not created additional radiance, even if it makes a small detail easier to resolve.

For the same 1.0 square millimetre patch seen within ten degrees, the projected acceptance is 9.47×10^{-8} square metre steradian. A wavelength-independent combined transmission and detector absorption of 0.65 leaves effective acceptance 6.16×10^{-8} square metre steradian. Integrating the Planck curve from 8 to 12 micrometres gives source and background band radiances 133.74 and 35.43 W per square metre per steradian. Their difference yields only 6.05 microwatts in the receiver. Keeping a single midpoint radiance times four micrometres happens to be fairly close here, but there is no fixed universal midpoint correction across temperatures or filters.

Open the live experiment

Try the calculation

Source temperature (K)	400
Ambient temperature (K)	295
Collection half-angle (°)	10



The net 8–12 µm signal is 6.053 µW for a 1 mm² area and 65% throughput. Ideal source peak wavelength is 7.24 µm.

Set source and ambient temperatures equal. The emitted and ambient curves remain finite while the differenced instrument reading vanishes. Increase the cone half-angle from ten to twenty degrees and predict the scale factor $\sin^2 20^\circ / \sin^2 10^\circ = 3.88$. The spectral shape should stay fixed if transmission is separable from angle. If real data shift spectral shape with aperture, inspect filter angle dependence, source nonuniformity and stray light. The integral in the readout is computed on the specified 8–12 micrometre band even though the graph shows wavelengths beyond it.

Convert received power to calibrated voltage or counts

A thermal detector may respond to absorbed power through a temperature rise and an electrical responsivity. A photon-counting detector measures absorbed photons, so its count rate contains a factor $\lambda/(hc)$ multiplying spectral power before integration. Two receivers with identical geometric acceptance and nominal quantum efficiency can respond differently to a changing source temperature if one output is proportional to energy and the other to photon number. Calibrate the actual electrical transfer and make the time interval explicit for accumulated counts.

For an ideal linear power detector let $V = V_0 + gP$, with offset V_0 and gain g. Two calibrated sources spanning the use range determine both. A black reference port at the same ambient temperature can also reject a common background in a chopped measurement. A filter, window or warm lens that absorbs thermal radiation also emits it. If these optics remain unchanged and stable, part of their own emission cancels in a reference difference. A reference traversing a different path generally leaves a residual input that is not captured by an electronic dark measurement.

Suppose effective source-minus-ambient powers of 2.00 and 8.00 microwatts yield 0.120 and 0.420 V. The fitted slope is 0.0500 V per microwatt with an offset of 0.020 V. A reading of 0.320 V implies 6.00 microwatts

after offset correction. Substituting the same reading directly into a single-point gain with zero assumed offset would bias the inferred temperature. Rechecking a third reference near six microwatts tests linearity, while repeated zero-input readings at realistic instrument temperature test offset drift.

Invert the radiance and propagate uncertainties

A narrowband blackbody radiance measurement can be inverted analytically by solving Planck's expression for T. In a broad passband use numerical quadrature across the calibrated spectral response and solve a scalar monotone equation when emissivity and background are known. The derivative at a single wavelength is particularly informative for estimating precision. With $x = hc/(\lambda k_B T)$,

$$\frac{\partial \log B_\lambda}{\partial \log T} = \frac{x}{1 - e^{-x}},$$

$$T = \frac{hc/(\lambda k_B)}{\log[1 + 2hc^2/(\lambda^5 B_\lambda)]}.$$

At 10 micrometres and 400 K the logarithmic sensitivity is 3.70. A one-percent relative radiance uncertainty therefore gives about 1.08 K temperature uncertainty if all other quantities are exact. An unknown five-percent emissivity error can dominate that estimate. A two-colour temperature ratio may reduce uncertainty from an emissivity that is constant across the two narrow bands, but cannot cancel an arbitrary spectral emissivity or a large reflected ambient component.

As an example consider a grey object at 400 K and ambient 295 K with emissivity 0.70. At ten micrometres its outgoing radiance is $0.70(33.56) + 0.30(9.14) = 26.24$ in the earlier spectral units. Inverting as a perfect blackbody gives a brightness temperature of about 375 K. The difference is a physical radiance–temperature degeneracy. Repeat at known ambient temperatures while keeping the object thermally stabilised to constrain the reflected fraction. If heating the background changes the actual object temperature, a second independent temperature sensor or a transient thermal model is necessary.

When detector heat capacity sets the measured signal

Consider a thermal receiver of heat capacity C_d coupled to its surroundings with thermal conductance G_d. For small excursions its temperature obeys $C_d\dot{\theta} + G_d\theta = P_{\mathrm{abs}}(t)$, where θ is the excess above its operating point. A voltage thermometer with responsivity r_T converts this into output. Its frequency-dependent power responsivity is $r_T/(G_d - i\omega C_d)$. The geometric and spectral calculations determine absorbed optical power, while this thermal equation determines how much of its modulation appears at the electrical output.

Take $C_d = 2.0$ microjoules per kelvin and $G_d = 0.10$ milliwatt per kelvin. The thermal time constant is 20 ms and ordinary corner frequency is 7.96 Hz. A sinusoidal absorbed-power amplitude of 6.0 microwatts would give a static temperature rise of 0.060 K. At 20 Hz the amplitude is reduced by $1/\sqrt{1 + (2\pi f\tau)^2} = 0.370$, leaving about 22.2 mK with a phase lag magnitude near 68 degrees. A receiver calibrated with a steady optical power is not automatically calibrated for this chopped experiment.

A square-wave chopper also changes the relation between time-domain peak differences and a lock-in's fundamental reading. For equal hot and reference dwell times, a square wave switching between powers P_1, P_0 has fundamental sinusoidal peak amplitude $2(P_1 - P_0)/\pi$, with different factors for rms and peak reports. The thermal transfer attenuates this component at the chopper frequency and the odd harmonics at

their respective frequencies. Substituting the full hot-minus-reference power directly into the first-harmonic response would overestimate the expected signal.

Measure a frequency sweep on a stable reference and test whether amplitude and phase share a single time constant. A geometrical error rescales the amplitude but ordinarily leaves this thermal phase unchanged. A second thermal mass or an electronic filter can introduce another pole. Lowering chop frequency increases thermal signal until the static limit but can expose low-frequency environmental drift. Choose the frequency by comparing the calibrated response with the measured noise spectrum.

Finally, a reference and an unknown source must illuminate the same active area if responsivity varies across the receiver. Imaging one source as a small spot and the other as a filled patch can create a calibration bias even when total spectral powers agree. Translating a reference spot across the detector or imaging the same aperture successively at both temperatures tests that spatial uniformity assumption.

If the sensor is used on a source much smaller than the calibrated reference aperture, record an image or a field-stop scan to determine its filling fraction. Otherwise even exact knowledge of filter transmission and thermodynamic temperature leaves an unknown geometrical scale factor. Aperture filling is part of the measurement equation, not an electronic gain setting.

Tutorial · Infer emissivity using two controlled backgrounds

An opaque diffuse sample at fixed temperature is observed at 10 micrometres against two isotropic backgrounds with calibrated spectral radiances 8.00 and 16.00 W per square metre per steradian per micrometre. The resulting outgoing readings are 25.00 and 27.00 in the same units. Infer emissivity and the sample's blackbody-equivalent radiance. Estimate the significance of independent reading errors of 0.10 units.

▼ **Separate the reflected and emitted contributions**

The increased outgoing radiance is 2.00 for an 8.00 increase of background. Thus reflectivity is 0.250 and emissivity 0.750. The first measurement gives intrinsic blackbody radiance $[25.00 - 0.25(8.00)]/0.75 = 30.67$. The second returns the same value. With 0.10 independent uncertainty on each outgoing reading and exact backgrounds, the slope uncertainty is $\sqrt{2}(0.10)/8 = 0.0177$, as is the absolute emissivity uncertainty. Larger background separation can improve this slope estimate if sample temperature stays fixed.

If the sample transmits radiation, the residual fraction does not all correspond to reflection. A controlled change behind the sample helps detect that additional path. A directional specular reflector exposed to a warm enclosure also does not necessarily see the same angular average used in the diffuse model. Without these checks, a formally precise fitted emissivity can describe the experimental surroundings instead of the material.

Tutorial · Design a numerical spectral-integration check

An engineer tabulates spectral radiance per micrometre on a wavelength mesh. Another rewrites the detector model on an equally spaced frequency mesh, obtaining a signal that differs by several percent after

independent mesh refinement. Explain what must transform, what stays dimensionless and how to compare a photon-counting channel with a power channel.

▼ **Transform the complete measurement integral**

The differential power $B_\lambda \, d\lambda$ must equal $B_\nu \, |d\nu|$, with $|d\lambda/d\nu| = c/\nu^2$. A dimensionless transmission curve becomes the same physical transmission at $\lambda = c/\nu$; it acquires no additional Jacobian. The integration bounds reverse when increasing wavelength is replaced by increasing frequency. A response specified per unit wavelength, or a calibrated gain integrating a particular finite bin, must instead be interpreted from how it was defined.

Evaluate a unit-transmission test using a very narrow band, then the full 8–12 micrometre band whose radiance values are given above. Refine each grid independently, treating sharp filter edges consistently and converting per-micrometre densities before using metres. For a photon counter divide each spectral power contribution by photon energy before summation. Replacing that energy by a single central-wavelength value is an approximation that should be checked against the full band and source temperature. Close agreement between two different meshes using the same incorrect density is not a valid units check.

NIST radiance-temperature calibrations · spectral radiance, sources and uncertainty

Moving-mirror Doppler metrology

Electromechanics, optics and measurement

A moving mirror changes optical phase, shifts the reflected frequency and exchanges energy and momentum with the light. In many displacement measurements the classical phase rule is extraordinarily accurate. Derive its exact inertial-frame counterpart, identify which time a detector samples and find the practical uncertainties that usually matter well before a relativistic correction does.

Match phases on a boundary that moves

In a vacuum laboratory let an incident plane wave travel along $+x$ with phase $kx - \omega t$. A perfectly reflecting plane mirror has position $X(t)$, receding for positive velocity. The reflected wave travels along $-x$ with phase $-k_r x - \omega_r t + \phi_0$. At the moving conducting surface the tangential electric fields must cancel in the mirror rest frame. For a constant mirror speed this requires equality of the two phase rates evaluated on its trajectory. The phase reversal at reflection supplies a constant offset and disappears on differentiation.

$$kv - \omega = -k_r v - \omega_r,$$

$$\frac{\omega_r}{\omega} = \frac{1-\beta}{1+\beta},$$

$$\beta = v/c.$$

Here $k = \omega/c$ and $k_r = \omega_r/c$ are positive magnitudes. A receding mirror redshifts the reflected light and an approaching mirror blueshifts it. Another derivation boosts into the mirror frame, where incident and reflected frequencies are equal, reverses the wavevector and boosts back. This yields a product of two Doppler factors, one for each transformation. Multiplying the first factor by its inverse without reversing the wavevector would predict no shift.

Define the signed beat $f_b = f_i - f_r$. The exact and first-order relations with vacuum wavelength λ_0 are

$$f_b = \frac{2v}{\lambda_0(1+\beta)},$$

$$f_b \simeq \frac{2v}{\lambda_0},$$

$$v = \frac{\lambda_0 f_b}{2 - \lambda_0 f_b/c}.$$

At 633 nm and $v = 2.00$ mm/s the beat is 6319.12 Hz. The relative correction to the first-order expression is only 6.67×10^{-12}. Frequency scale, geometry and air refractive index are ordinarily much larger uncertainties in an uncorrected bench experiment. The same exact expression makes a genuinely different prediction at $\beta = 0.10$: $f_r/f_i = 0.81818$, whereas the first-order frequency ratio is 0.80. This latter value is a theoretical benchmark for the formula rather than a required moving-bench experiment.

Connect reflection time with the detector's interference phase

Put an ideal stationary beamsplitter and detector at the origin, and let the source continuously illuminate the mirror along the measurement arm. A wave arriving back at laboratory time t was reflected at t_r satisfying $t = t_r + X(t_r)/c$. It left the beamsplitter at $t_0 = t_r - X(t_r)/c$. Combining these gives

$$t - t_0 = 2X(t_r)/c.$$

With a stable source and a fixed reference arm, the accumulated return phase relative to the reference is proportional to $2kX(t_r)$ up to a constant and an electronic quadrature sign.

$$\phi(t) = \frac{4\pi X(t_r)}{\lambda_0} + \phi_{\mathrm{off}},$$

$$\frac{1}{2\pi}\frac{d\phi}{dt} = \frac{2v(t_r)}{\lambda_0[1 + v(t_r)/c]}.$$

The denominator comes from differentiating the arrival-time condition. It is not obtained from differentiating a geometrical arm length at detector time while keeping the reflection time fixed. For very slow motion and a modest arm one can replace the retarded trajectory by the present one to high accuracy. At higher bandwidth a fixed path delay can still matter even if the velocity correction is tiny. A one-metre mirror distance gives roughly 3.34 ns from reflection event to detection.

A sinusoidally driven mirror with amplitude 100 nm and ordinary mechanical frequency 1.0 kHz has phase amplitude $4\pi A/\lambda_0 = 1.99$ rad at 633 nm. The peak speed is 0.628 mm/s and peak signed beat about 1.99 kHz. The instantaneous beat changes sign at each mechanical turning point. A frequency counter returning a single unsigned average misses that directional information. Demodulating the phase or fitting the sinusoidal motion to both quadratures can recover position, speed and reversal times.

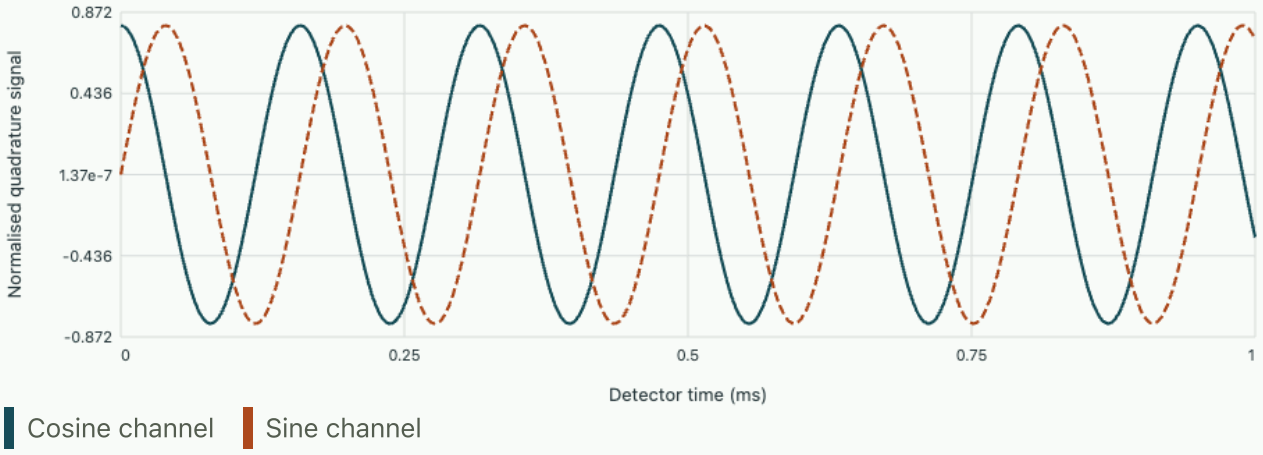
Infer displacement or signed speed without losing complete fringes

If stationary reference and returned fields overlap in polarisation, spectrum and space, their combined intensity contains a cross term proportional to $\cos\phi$. A single intensity channel cannot distinguish increasing from decreasing phase at an intensity extremum. Two calibrated quadrature outputs ideally read $X_s = V\cos\phi$, $Y_s = V\sin\phi$ after removal of offsets and gain normalisation. Recover wrapped phase with the two-argument arctangent. Unwrap successive phase differences only when the sampling interval and noise allow the branch to be identified.

Open the live experiment

Try the calculation

Mirror speed, positive receding (mm/s)	2
Vacuum wavelength (nm)	633
Fringe visibility	0.8



The signed beat is 6319.12 Hz. In 1 ms the mirror travels 2.00 µm and the unwrapped phase changes by 39.70 rad.

The signed velocity control reverses the direction traced by the quadratures. Hold its absolute value fixed and change sign. The cosine channel alone remains almost identical at laboratory speeds; the sine channel reverses the progression from its initial value. Reduce the visibility and keep in mind that the graph shows noise-free curves. With fixed electronic amplitude noise a smaller circle makes angular uncertainty larger, even though the exact underlying phase still advances at the same rate.

One full phase turn represents a physical mirror displacement of half the vacuum wavelength at normal incidence. A 20-fringe movement at 633 nm is 6.33 micrometres. If both endpoints retain phase residuals, add their signed difference to the complete turn count before converting to distance. A missed fringe biases displacement by 316.5 nm regardless of how accurately sub-fringe phases are fitted. For the default speed and wavelength the magnitude of frequency is 6.32 kHz. To unwrap a plain uniformly sampled phase with less than half a turn between adjacent samples one needs a sampling frequency above 12.64 kHz, and usually a margin for noise and acceleration.

Quadrature imbalance distorts the ideal circle into an ellipse shifted away from the voltage origin. Taking an arctangent of raw voltages then introduces a periodic position error. Record a slow multi-fringe calibration, fit channel offsets, scales and nonorthogonality, and test the recovered phase on held-out data with reversed motion. Calibration using only a tiny arc of the ellipse makes offset and scale poorly identifiable even if the residuals on that arc are small.

Close the energy and momentum balance at the mirror

One incident photon has energy hf_i and momentum $+hf_i/c$. On reflection it has energy hf_r and momentum $-hf_r/c$. The mirror acquires momentum $h(f_i + f_r)/c$ and energy $h(f_i - f_r)$. Their ratio is exactly v when the prescribed constant-speed reflection formula applies. A recoil-free moving boundary can exchange

mechanical work $v\,\Delta p$ while a drive maintains its velocity. For a freely recoiling low-mass mirror illuminated by a small number of photons the fixed-velocity approximation must be replaced by energy and momentum conservation for the joint system.

For an incident beam with power P_i crossing a stationary plane in the lab, the number of photons encountering the receding surface per laboratory time is $P_i(1-\beta)/(hf_i)$. Multiply by transferred momentum per encounter to get the force on that moving surface. The result simplifies to $F = (2P_i/c)(1-\beta)/(1+\beta)$. At rest, a 10 mW perfectly reflecting beam gives 66.7 pN. For a 2.0 mm/s receding mirror at that power, the corresponding mechanical work rate is only 1.33×10^{-13} W to leading order.

A detector fixed upstream sees reflected-wave intensity reduced by the square of the reflected frequency ratio for an ideal steady plane wave. Its received power change compared with a fixed-plane incident flux is not in general equal to the instantaneous work on the mirror. The region of field between the fixed reference plane and the moving surface grows or shrinks and stores energy. Instead compare powers crossing the moving boundary with their proper encounter rates, or account for the change of field energy in a specified control volume. For a finite pulse the integrated photon count is unchanged under perfect reflection and the total energy loss per photon is the Doppler shift already derived.

Estimate realistic errors before pursuing small corrections

At normal incidence in vacuum the first-order speed estimate is $\lambda_0 f_b/2$. For a beam at angle θ to the translation direction, with an appropriately arranged reflection/return geometry interrogating mirror-normal motion, the usual small-speed sensitivity along the normal includes $\cos\theta$. A tilt of two degrees gives a relative cosine factor 0.99939, a 610 parts-per-million reduction. A plane mirror tilted relative to a fixed coaxial receiver can additionally send the reflected beam outside the collection path, so specify how oblique reflected and reference fields are recombined. Blindly interpreting reduced fringe visibility as a velocity correction conflates these two effects.

In a homogeneous stationary transparent medium a quasistatic phase measurement measures optical path nL. If n changes during a displacement measurement, the phase derivative includes both $n\dot{L}$ and $L\dot{n}$ when the simple uniform-medium approximation holds. A refractive-index change of one part in a million over a 0.5 m arm imitates about 0.50 micrometres of path-length change. A short differential arm, a refractive-index reference or a controlled environment can separate it from actual motion. The vacuum relativistic formula cannot simply absorb an arbitrary time-varying medium by replacing c everywhere with c/n.

For a phase-difference velocity estimate over a record of duration T, the displacement noise from two independent endpoint phase errors of standard deviation s_ϕ is $\lambda_0\sqrt{2}s_\phi/(4\pi)$. At 633 nm, $s_\phi = 0.010$ rad and $T = 2$ ms, the resulting velocity uncertainty is about 0.356 micrometres per second. Fitting a slope over many samples can improve white-noise precision, but laser drift, mechanical acceleration and missed branches require their own models. Report the time interval over which a constant velocity is being claimed.

Recover a small vibration amplitude from optical sidebands

A phase-sensitive receiver is not the only way to recover periodic motion. Let a mirror vibrate at angular frequency Ω about a fixed mean, with displacement $A\cos\Omega t$ and speed sufficiently small for a first-order optical path description. The reflected analytic signal relative to its carrier contains $e^{im\cos\Omega t}$ with modulation

index $m = 4\pi A/\lambda_0$. Expand the exponential in its Fourier series. The power of sideband order j relative to the incident unmodulated carrier is $J_j(m)^2$, where J_j is a Bessel function. Energy conservation of the phase factor requires that all the sideband powers sum to one in this phase-only approximation.

For a 10.0 nm displacement amplitude at 633 nm the index is 0.1985. To leading small-index order, each first sideband carries $(m/2)^2 = 0.00985$ of the original carrier power. The central component loses approximately twice this amount. A receiver collecting only one sideband would infer the wrong vibration amplitude if it treated that signal as the sum of both. At larger amplitudes higher Bessel orders become important and a single sideband amplitude is no longer a monotone function of displacement. A multi-sideband fit or an independent coarse amplitude resolves the ambiguity.

With a balanced optical heterodyne reference offset far enough from zero frequency, both sidebands can be separated electronically from the carrier beat. Measuring their power ratio to the central beat cancels a common optical gain, but any electronic response slope between their frequencies needs calibration. Pure phase modulation predicts symmetric sideband powers. Asymmetry can arise from simultaneous amplitude modulation or an unequal frequency response rather than from a change in mirror speed. Measuring their complex phases distinguishes amplitude and phase modulation more directly.

A useful consistency check uses three independent measured quantities. Infer displacement amplitude from the first two sidebands, obtain the mechanical frequency from their separation and predict peak Doppler beat from the product of those two values. Then compare with a separate unwrapped quadrature phase trace. For a rigid sinusoidal mirror these should agree within timing and wavelength uncertainties. An interferometric signal containing two different moving surfaces can fit a few sideband powers yet fail the predicted time-dependent phase; changing the illuminated spot or one surface position tests that possibility. A spectrum and a time trace jointly constrain motion in a way that a single fringe visibility cannot.

Tutorial · Resolve a branch ambiguity in a vibrating mirror

A mirror returns a signed mean beat of 8.0 kHz while moving uniformly over a short time window. The vacuum wavelength is 632.8 nm. The detector samples quadrature at 10.0 kHz and a naive nearest-branch unwrap reports minus 2.0 kHz. Explain the discrepancy, infer both candidate speeds and design a measurement to select the correct one.

▼ **Track the sampled phase and use an independent constraint**

A true phase advance of 0.80 cycle per sample is indistinguishable from minus 0.20 cycle per sample when only consecutive wrapped angles are available. The corresponding physical speeds are plus 2.531 mm/s and minus 0.6328 mm/s. Both give exactly the same ideal complex samples at that sampling rate under the first-order laboratory formula. A smaller least-squares residual on one constant-speed branch cannot resolve the ambiguity without another observation.

Increase the sampling frequency above 16 kHz with margin or use an independently known direction and bounded speed range that excludes alternative aliases. A second wavelength also changes the beat while preserving physical velocity and can resolve the branch if the two ambiguous frequency ladders do not coincide over the admissible range. For accelerated motion verify the condition at the maximum instantaneous beat rather than at its signed average over a complete vibration period.

Tutorial · Can this beat measurement verify a relativistic correction?

A translation standard operates at 1.00 m/s and an interferometer at 500 nm records a nominal 4.00 MHz beat. Estimate the difference between exact and first-order vacuum predictions. A timebase has a fractional uncertainty 10^{-8}, and wavelength is uncertain by 2×10^{-8}. Assess the test and state one way to reject an offset without confusing it with the desired correction.

▼ Calculate the small difference without subtracting nearly equal optical frequencies

Use the exact beat formula or expand it directly. The correction is approximately $-2v^2/(\lambda_0 c) = -0.0133$ Hz for the receding mirror. The stated timebase contributes 0.040 Hz and the wavelength scale contributes 0.080 Hz to a 4 MHz prediction, even before including the velocity standard or alignment. This setup does not resolve the correction with those absolute uncertainties. Subtracting two floating-point optical frequencies near hundreds of terahertz is also a poor numerical route to such a small difference.

Reverse the velocity with equal magnitude. The leading signed beat is odd in velocity, while its leading relativistic correction is even. Adding the two signed beats cancels the ideal leading term and leaves about minus 0.0267 Hz. This reversal also preserves a fixed electronic frequency offset, which would add twice to the same sum. Take stationary-mirror reference readings and vary the speed magnitude; the wanted even term scales as speed squared. Different arm alignments or asymmetric velocity calibration during reversal can still imitate part of that effect.

Cambridge Electromagnetism · wave propagation, radiation momentum and Lorentz transformations

Where momentum waits between source and receiver

Electrodynamics

Derive the electromagnetic momentum balance, trace light pressure through a finite pulse, and account for the angular impulse when a magnetic flux decays.

Give a region a mechanical and a field account

Charges respond to Lorentz force density $\mathbf{f} = \rho\mathbf{E} + \mathbf{J} \times \mathbf{B}$. To predict the total force on enclosed matter, one can measure the fields on a surrounding vacuum surface. First eliminate the sources. In vacuum Gauss's law gives $\rho = \epsilon_0 \nabla \cdot \mathbf{E}$, and the Ampère–Maxwell law gives

$$\mathbf{J} = (\nabla \times \mathbf{B})/\mu_0 - \epsilon_0 \partial_t \mathbf{E}.$$

Introduce $\mathbf{g} = \epsilon_0 \mathbf{E} \times \mathbf{B} = \mathbf{S}/c^2$. Its time derivative supplies both $\epsilon_0 \partial_t \mathbf{E} \times \mathbf{B}$ and $\epsilon_0 \mathbf{E} \times \partial_t \mathbf{B}$. The first cancels the displacement-current term in the force; Faraday's law rewrites the second.

For example, the electric terms now read $\epsilon_0[\mathbf{E}(\nabla \cdot \mathbf{E}) - \mathbf{E} \times (\nabla \times \mathbf{E})]$. Component i is $\partial_j[\epsilon_0(E_i E_j - \delta_{ij} E^2/2)]$, with repeated Cartesian indices summed. The magnetic terms have the same form because $\nabla \cdot \mathbf{B} = 0$. Integration gives

$$T_{ij} = \epsilon_0(E_i E_j - \tfrac{1}{2} E^2 \delta_{ij}) + \frac{1}{\mu_0}(B_i B_j - \tfrac{1}{2} B^2 \delta_{ij}),$$

$$\mathbf{F}_{\text{matter}} = \oint T\mathbf{n}\, dA - \frac{d}{dt}\int \mathbf{g}\, dV.$$

The normal points out of the chosen volume. Positive T_{xx} denotes tension along x in this convention. For a wave travelling in $+x$ with transverse fields, $T_{xx} = -u$, where u is energy density. On an inlet whose outward normal is $-\hat{\mathbf{x}}$, the normal traction therefore points in $+x$.

Measure the time before light reaches a detector

A finite plane pulse in vacuum has $|\mathbf{S}| = cu$. Its total momentum is U/c along the propagation direction. Surround a freely absorbing target and a length of initially empty space upstream with a vacuum control surface. While a pulse enters but has yet to reach the target, the inlet stress integral is nonzero. The electromagnetic momentum inside grows at exactly the same rate. The minus time derivative in the total force cancels the inlet contribution and the target has zero force during this interval.

Worked optical impulse

A 0.60 mJ pulse is fully absorbed after crossing the inlet. Initially and finally the region contains no pulse. Integrating the balance over that full time removes the stored-field term. The target gains $0.60 \times 10^{-3}/c = 2.00 \times 10^{-12}$ N s. For ideal normal reflection the exiting pulse carries the opposite momentum, giving twice that impulse. For a transparent window with parallel faces at normal incidence the outgoing pulse may eventually have the same vacuum momentum as the incoming pulse, while the two faces still experience transient forces.

Convert distributed field momentum into a turning impulse

Now take a long coaxial arrangement: an inner narrow cylindrical shell of radius a carries free charge $+\lambda$ per length; an outer shell of radius b carries $-\lambda$. There is a uniform axial magnetic field B_0 over the annular gap, with end effects negligible. For a definite preparation take an ideal long solenoid producing uniform field throughout $r < b$, and confine the initial electric field to the gap. Between shells $E_r = \lambda/(2\pi\epsilon_0 r)$. The field momentum density points towards negative azimuth because $\hat{\mathbf{r}} \times \hat{\mathbf{z}} = -\hat{\boldsymbol{\phi}}$.

$$g_\phi = -\lambda B_0/(2\pi r),$$

$$L'_{\mathrm{field},z} = \int_a^b r g_\phi \, 2\pi r \, dr = -\tfrac{1}{2}\lambda B_0 (b^2 - a^2).$$

The prime denotes angular momentum per length. While the coil current decays slowly, Faraday’s law gives $2\pi r E_\phi = -\pi r^2 \dot{B}$ for radii up to b. The torque on each shell per length is $r\lambda_{\mathrm{shell}} E_\phi(r)$. Integrating from B_0 to zero yields $+\lambda B_0 a^2/2$ on the inner shell and $-\lambda B_0 b^2/2$ on the outer. Their sum reproduces the initial field angular momentum. Clamps would receive those impulses if the shells were held fixed.

Worked sign and magnitude

With $\lambda = 20$ nC/m, $B_0 = 0.25$ T, $a = 10$ mm and $b = 30$ mm, the initial $L'_z = -2.0 \times 10^{-12}$ kg m/s. That is angular momentum per metre of apparatus. The outer shell receives nine times the inner impulse magnitude and rotates in the opposite sense. Torque from the induced azimuthal electric field supplies this transfer even though both shells started at rest.

Transfer angular momentum from a field into two shells

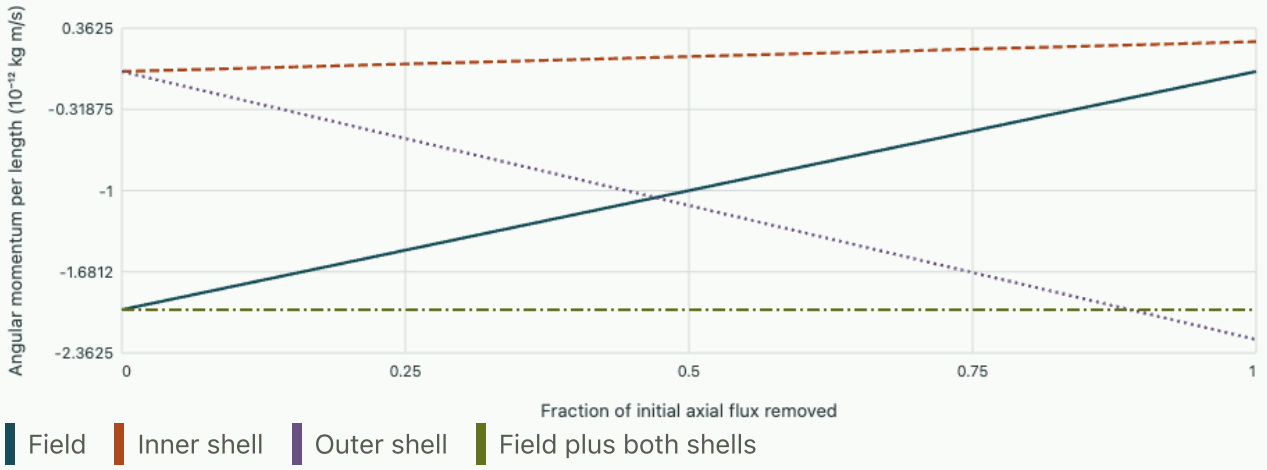
The inner radius is 10 mm and its line charge is +20 nC/m. For a 30 mm outer radius and an initial field of 0.25 T, the final shell impulses are +0.25 and −2.25 in units of 10⁻¹² kg m/s. Trace the four accounts during switch-off.

Open this experiment online

Try the calculation

Outer radius b (mm) **30**

Initial field B (T) **0.25**



After switch-off the inner shell carries 0.25 and the outer shell -2.25 in plotted units. Their sum equals the initial field value -2.

Choose a control surface for momentum

Find a pressure when transmission and reflection coexist

A normal beam of intensity 300 W/m² reaches a free-standing film in vacuum. Reflected power fraction is 0.20 and transmitted fraction is 0.50. Find the time-averaged pressure and identify the momentum carried away.

▼ **Work through the solution**

With incident direction positive, outgoing momentum per area and time is $(T - R)I/c$. The deposited amount is

$$(1 - T + R)I/c = 0.70(300)/c = 0.700\,\mu\text{Pa}.$$

The absorbed part contributes $0.30I/c$ and reflection adds $0.40I/c$. The result assumes forward and backward beams remain normal to the film.

Reverse just the charges

In the coaxial example reverse both shell charges while keeping the axial field and current decay unchanged. Which quantities reverse?

▼ **Work through the solution**

The radial field, stored azimuthal momentum and angular momentum reverse. Faraday's induced E_ϕ follows the same history because the flux decay is unchanged. Each electric force and angular impulse reverses since the charge signs reverse. This independent sign argument checks the cross product in the original momentum density.

Practicum · audit a pulse bouncing inside a partly absorbing cavity

A shutter has just prepared a short $U_0 = 1.20$ mJ pulse travelling to the right in an otherwise empty one-dimensional cavity. Start the accounting after preparation. The left mirror reflects perfectly. At every encounter with the right mirror 60% of incident pulse energy is reflected, 30% passes into vacuum on the right, and 10% is absorbed. Pulse duration is short compared with the flight time across the cavity. Mirrors are mounted on massive force probes; neglect their displacement and delay any re-emission of heat. The challenge is to reconcile two large, oppositely signed readings with a small net impulse.

Audit one encounter and decide which surface terms matter

For energy U just before an encounter with the right mirror, calculate its mechanical impulse. Then find the impulse when the reflected packet hits the left mirror. Which momentum is still present between those events?

▼ Follow the calculation and check the result

The incoming momentum is $+U/c$. The two outgoing packets carry $+0.30U/c$ and $-0.60U/c$. Momentum balance therefore assigns the right mirror $+(1 - 0.30 + 0.60)U/c = +1.30U/c$. This includes reversing the reflected packet and stopping the absorbed packet. The left mirror receives $-2(0.60U)/c = -1.20U/c$ at the next collision. Between those events the cavity still contains $-0.60U/c$ of field momentum. The complete system also contains the $+0.30U/c$ transmitted packet outside the cavity.

Stop the movie at a specified time

Freeze the experiment just after the third right-mirror encounter and before the next left-mirror encounter. Find the remaining energy, radiated output, deposited heat, both probe impulses, and total field momentum.

▼ Follow the calculation and check the result

Three incident energies in units of U_0 form the sequence $1, 0.6, 0.36$. Their sum is 1.96. The output energy is $0.30(1.96)U_0 = 0.588U_0$, heat is $0.196U_0$, and $0.216U_0$ remains. The three fractions sum to unity. Right-probe impulse is

$$+1.30(1.96)U_0/c = +2.548U_0/c.$$

Only two left reflections have occurred, so the left probe reads $-2(0.60 + 0.36)U_0/c = -1.920U_0/c$. Their sum is $+0.628U_0/c$.

The escaped packets carry $+0.588U_0/c$, but the packet still inside carries $-0.216U_0/c$. Including these two contributions gives measured plus field momentum $(0.628 + 0.588 - 0.216)U_0/c = U_0/c$, the initial value.

Counting a third left-hand reflection at this time would violate the preselected event boundary. Since $U_0/c = 4.00 \times 10^{-12}$ N s, the probe readings are approximately $+10.2$ and -7.68 pN s.

Sum the full history in two independent ways

Continue until no macroscopic pulse remains. Find the impulse on each mirror, then derive their combined reading directly from exterior measurements without examining any reflection.

▼ **Follow the calculation and check the result**

The geometric series of right hits sums to $U_0/(1-0.6) = 2.5U_0$. The right impulse is $3.25U_0/c$. Summing every left hit gives

$$-2(0.60)U_0/[(1-0.60)c] = -3.00U_0/c.$$

Thus the pair receives $+0.25U_0/c$. All transmitted packets travel right and together carry $0.30U_0/(1-0.60) = 0.75U_0$. An external vacuum control surface therefore gives $(1-0.75)U_0/c$ immediately. The absorbed $0.25U_0$ accounts for the energy stored as heat.

A detector without enough time resolution can recover the integrated exterior momentum even though it cannot resolve the alternating internal forces. Conversely, two large internal probe signals can coexist with a modest net exterior impulse. In a numerical time march record separate signed transfers and the energy still in flight after each boundary update, then vary the stopping time. A budget that closes only after the very last encounter cannot diagnose an intermediate sign mistake.

Further reading

- Griffiths, §§8.2.2–8.2.4, stress, momentum and angular momentum
- Tong, Electromagnetism and Relativity, energy-momentum and angular momentum

Why a small transparent sphere scatters light

Electrodynamics

A particle can redistribute light without consuming it. The calculation starts with an electrostatic boundary-value problem and ends with power crossing a faraway sphere. Connecting the two limits predicts a wavelength dependence that can be tested with colour.

Find the induced dipole by matching two potentials

Take a sphere of radius a and relative permittivity ϵ_r in vacuum. Its size must be small compared with the wavelength everywhere, including inside the particle. Across its diameter, the incident wave then looks almost like a uniform electric field $E_0\hat{\mathbf{z}}$. Spherical symmetry and the applied field single out the angular factor $\cos\theta$. Choose $\phi_{\mathrm{in}} = -Cr\cos\theta$ and $\phi_{\mathrm{out}} = -E_0 r\cos\theta + D\cos\theta/r^2$. Continuity of tangential electric field is equivalent here to continuity of potential. With zero free surface charge, the normal displacement is also continuous. At $r = a$,

$$C = E_0 - D/a^3,$$

$$\epsilon_r C = E_0 + 2D/a^3.$$

Eliminate C to obtain

$$D = a^3 E_0(\epsilon_r - 1)/(\epsilon_r + 2).$$

The coefficient of the dipole potential is $p/(4\pi\epsilon_0)$. Hence $p = \alpha_0 E_0$ with $\alpha_0 = 4\pi\epsilon_0 a^3(\epsilon_r - 1)/(\epsilon_r + 2)$. The internal field $3E_0/(\epsilon_r + 2)$ explains the denominator. For $\epsilon_r = 1$ the disturbance vanishes. In the perfectly polarizable limit the internal field tends to zero.

Match a small dipole to an outgoing wave

Use complex amplitudes with time dependence $e^{-i\omega t}$. The far electric field is $(k^2/4\pi\epsilon_0 r)[(\hat{\mathbf{n}}\times\mathbf{p})\times\hat{\mathbf{n}}]e^{ikr}$, $k = \omega/c$. Squaring its amplitude and multiplying by $\epsilon_0 c/2$ gives the mean intensity. Divide by the incident intensity $\epsilon_0 c|E_0|^2/2$. The differential scattering cross section is then $k^4|\alpha|^2\sin^2\theta/(16\pi^2\epsilon_0^2)$, with θ measured from the dipole axis. Integrating the angular factor gives $8\pi/3$:

$$\sigma_{\mathrm{sca}} = \frac{k^4|\alpha|^2}{6\pi\epsilon_0^2} \simeq \frac{8\pi}{3}k^4a^6\left|\frac{\epsilon_r - 1}{\epsilon_r + 2}\right|^2.$$

The sixth power originates in the coherent addition of dipoles throughout a volume proportional to a^3. The fourth power in frequency comes from the second time derivative of the dipole in the radiative electric field. A cross section has units of area and measures removed power divided by incident intensity.

Where the scattered energy comes from

The incident field does work at mean rate $\langle\mathbf{E}\cdot\dot{\mathbf{p}}\rangle = (\omega/2)\,\mathrm{Im}(\alpha)|E_0|^2$. Dividing by incident intensity yields

$$\sigma_{\mathrm{ext}} = k\,\mathrm{Im}(\alpha)/\epsilon_0.$$

A strictly real electrostatic α_0 would predict no extracted power while predicting positive scattered power. The static calculation captured the leading dipole amplitude but omitted its small radiation-induced phase delay. For a lossless electric dipole its leading radiative correction satisfies

$$\frac{1}{\alpha} = \frac{1}{\alpha_0} - i\frac{k^3}{6\pi\epsilon_0}.$$

Indeed $\operatorname{Im}\alpha = k^3|\alpha|^2/(6\pi\epsilon_0)$, making extinction equal to scattering. In an absorbing particle their difference is absorbed power per incident intensity. Finite-size dispersion and higher multipoles require additional terms once ka grows.

Size, colour and the limits of inference

A particle doubles in radius

For dilute spheres of the same weakly dispersive material, doubling a gives $2^6 = 64$ times the scattered power per particle at fixed incident intensity. If the total volume of dielectric is held fixed, the number of spheres falls by eight. The whole sample then scatters eight times the power, provided phases between different particles are effectively uncorrelated and multiple scattering is negligible. The two questions describe different samples.

A frequency-independent contrast

For $a = 20$ nm, $\epsilon_r = 2.25$ and vacuum wavelength 600 nm, $ka = 0.209$ and $(\epsilon_r - 1)/(\epsilon_r + 2) = 0.2941$. The leading cross section is 5.58×10^{-19} m², around 4.4×10^{-4} of the geometric area. At 450 nm the model predicts 3.16 times that cross section. A large discrepancy can indicate dispersion, additional multipoles, or a distribution of particle sizes.

Compare the scattering of several particle sizes

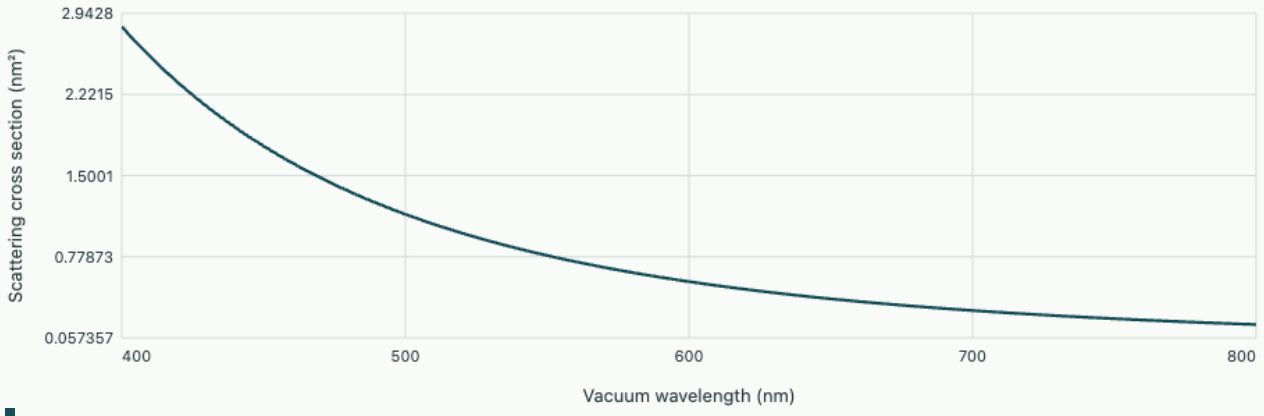
For radius 20 nm, relative permittivity 2.25 and wavelength 600 nm, the predicted cross section is about 0.56 nm². Halve the radius and estimate the drop before reading the plot. The curve uses a frequency-independent dielectric contrast.

Open this experiment online

Try the calculation

Sphere radius (nm) **20**

Relative permittivity **2.25**



Rayleigh sphere

The leading cross section at 600 nm is 0.558 nm². The shortest plotted wavelength gives ka = 0.314; finite-size corrections grow as this ratio grows.

Problem sheet · inverse measurements

Estimate a radius ratio

Two dilute suspensions have the same number density, composition and optical path length. In the single-scattering regime one produces twice the scattered intensity at the same angle. Infer the ratio of particle radii.

▼ Work through the solution

The proportionality to a^6 gives

$$a_2/a_1 = 2^{1/6} = 1.122.$$

A 12.2% change in radius doubles the signal. If counts or material contrasts are unknown, the same intensity change has multiple explanations.

Recover Thomson scattering

A free electron is driven by a weak field. Use $m\ddot{x} = -eE_0 e^{-i\omega t}$ to find its dipole and cross section. Why does the wavelength law change?

▼ Work through the solution

The displacement amplitude is $eE_0/(m\omega^2)$. Multiplying by electronic charge gives

$$\alpha = -e^2/(m\omega^2).$$

The ω^{-4} in $|\alpha|^2$ cancels k^4. The cross section becomes $(8\pi/3)[e^2/(4\pi\epsilon_0 mc^2)]^2$. This classical approximation assumes nonrelativistic electron motion and negligible recoil.

Further reading

- Fließbach, chapter 25 · scattering of light
- Cambridge Electromagnetism, chapter 5 · Rayleigh and Thomson scattering

From wavefronts to rays, focusing and caustics

Electrodynamics

A ray is a prediction extracted from the phase of a wave. Following its amplitude at the same time reveals the point at which a ray calculation loses its validity. A gently graded transparent material gives a solvable example of both effects.

Derive the phase and transport equations together

In a lossless isotropic dielectric with a slowly varying refractive index $n(\mathbf{r})$, consider one scalar polarisation component satisfying $\nabla^2\psi + k_0^2 n^2 \psi = 0$, where $k_0 = \omega/c$. This scalar description works when vector-polarisation coupling is negligible over the scales of interest. Set $\psi = A(\mathbf{r})e^{ik_0 S(\mathbf{r})}$ with a slowly varying real amplitude. Differentiation twice gives

$$0 = \nabla^2 A + ik_0[2\nabla A \cdot \nabla S + A\nabla^2 S] + k_0^2 A(n^2 - |\nabla S|^2).$$

If the material and amplitude vary little over a wavelength, first equate the leading k_0^2 term to zero. At the next order the k_0 term vanishes. The two results are the eikonal equation and transport equation:

$$|\nabla S|^2 = n^2,$$

$$\nabla \cdot (A^2 \nabla S) = 0.$$

The first fixes the optical length accumulated between neighbouring wavefronts. The second transports energy down a thin ray tube. Across a tube of area $\mathcal{A}$ perpendicular to the phase gradient, $A^2 n \mathcal{A}$ is constant to this order. In a uniform medium a spherical tube grows as r^2, giving $A \propto 1/r$, consistent with the outgoing wave.

Follow the normal to a wavefront

Choose arclength s along the ray and $\hat{\mathbf{t}} = \nabla S / n$. In the eikonal equation take a spatial derivative. Since ∇S is a gradient,

$$(\nabla S \cdot \nabla)\nabla S = \nabla(|\nabla S|^2)/2 = n \nabla n.$$

Divide by n. The ray equation is

$$\frac{d}{ds}(n\hat{\mathbf{t}}) = \nabla n.$$

A transverse increase of index bends the ray toward larger optical index. If n depends only on the coordinate normal to layers, the component of $n\hat{\mathbf{t}}$ parallel to them stays constant. Thus $n \sin\theta$ is conserved for an angle measured from the normal, recovering Snell's law. The same equation follows by requiring stationary optical length $\int n\, ds$ when varying the path with fixed endpoints.

Solve a graded-index profile

For paraxial rays travelling close to $+z$, take $n(y) \simeq n_0(1 - g^2 y^2/2)$. Assume $|gy| \ll 1$ and $|dy/dz| \ll 1$. Then $ds \simeq dz$, $\hat{t}_y \simeq y'$, and $d(ny')/dz \simeq n_0 y''$ to leading order. Since $\partial_y n = -n_0 g^2 y$, the trajectory solves $y'' + g^2 y = 0$. Initial height y_0 and slope s_0 give

$$y(z) = y_0 \cos(gz) + \frac{s_0}{g}\sin(gz),$$

$$y'(z) = -g y_0 \sin(gz) + s_0 \cos(gz).$$

For a collimated bundle, all $s_0 = 0$. The transverse magnification $\partial y/\partial y_0 = \cos(gz)$ vanishes simultaneously for all rays at $z_f = \pi/(2g)$. This is a caustic of the ray description. In this planar example the ray density grows as $1/|\cos(gz)|$. Its divergence signals that the neglected $\nabla^2 A$ becomes important. The underlying wave stays finite and develops diffraction near the focus.

A focus and a reconstructed gradient

Refocus an off-axis bundle

Take $g = 20\ \mathrm{m}^{-1}$, initial heights between -1 and $+1$ mm and zero slopes. The predicted first focus is 78.5 mm downstream. The extreme rays acquire slopes of magnitude 0.020 at the focus, well within the paraxial regime. Tilt every initial ray by 0.010 rad. At z_f the common height shifts to $s_0/g = 0.50$ mm. The focal length is insensitive to this common tilt in the parabolic approximation.

Estimate the index from a photograph

Two successive crossings of initially parallel symmetric rays are 120 mm apart. Crossings occur at $gz = \pi/2 + m\pi$, so

$$g = \pi/0.120 = 26.2\ \mathrm{m}^{-1}.$$

At a height of 2 mm, the predicted fractional index decrease from the central index is

$$g^2 y^2/2 = 1.37 \times 10^{-3}.$$

Resolving this small contrast through repeated bending can be easier than resolving it through a single refraction measurement.

Focus and tilt a bundle of paraxial rays

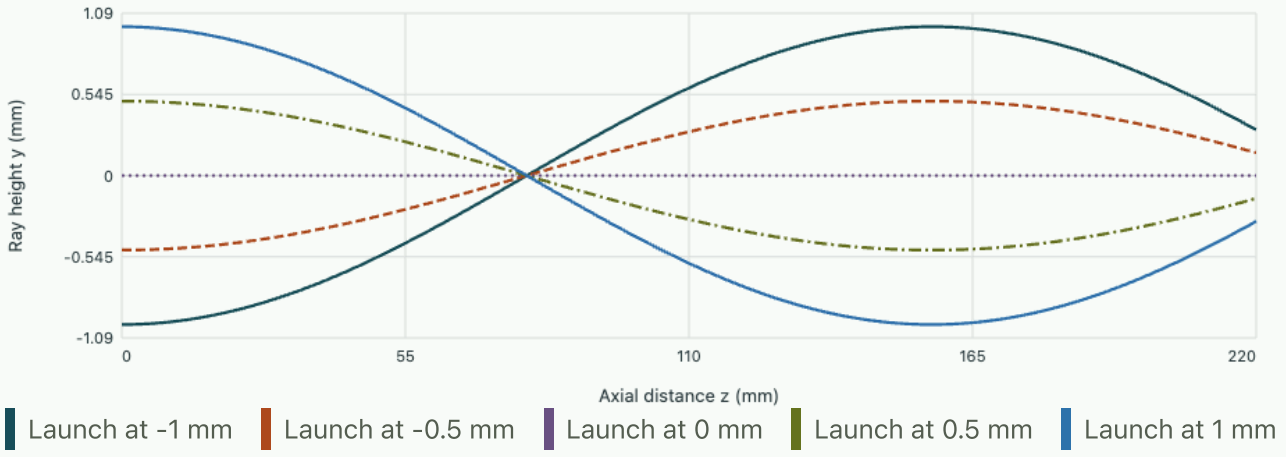
At gradient 20 m⁻¹, five parallel rays meet after 78.5 mm. A common tilt of 10 mrad displaces their meeting point by 0.50 mm. Follow the bundle through a crossing and decide why the geometric density ceases to predict a finite intensity there.

Open this experiment online

Try the calculation

Index gradient g (1/m) **20**

Common input slope (mrad) **0**



The first crossing occurs at z = 78.5 mm and y = 0 mm. Finite-wavelength diffraction limits the intensity at that crossing.

Problem sheet · optical path and diffraction

A travel-time check

A ray passes through two homogeneous layers of thicknesses a, b, indices n_1, n_2, with lateral endpoint separation D. Write travel time as a function of crossing coordinate x. Recover the interface condition.

▼ **Work through the solution**

The time is

$$T(x) = [n_1\sqrt{a^2 + x^2} + n_2\sqrt{b^2 + (D - x)^2}]/c.$$

The derivative vanishes when $n_1 x/\sqrt{a^2 + x^2} = n_2(D - x)/\sqrt{b^2 + (D - x)^2}$. The two fractions are the sines of the angles to the interface normal. Each segment's length has positive second derivative, making this stationary path a local minimum in this geometry.

Estimate the diffraction scale

A circular pupil of numerical aperture 0.030 focuses vacuum light of wavelength 600 nm. Estimate the first dark-ring radius and compare it with a ray focus.

▼ **Work through the solution**

For a uniformly illuminated circular aperture in the paraxial limit the first Airy minimum is $0.61\lambda_0/\mathrm{NA}$, giving 12.2 µm. Different pupil shapes change the numerical factor. The wavelength dependence is absent from the ray crossing, so a ray trace alone cannot distinguish two experiments with identical geometry but different wavelengths.

Further reading

- Fließbach, chapter 38 · geometrical optics
- Fließbach, chapter 36 · interference and diffraction

Applied study

Predict the spacing and visibility of interference fringes from phase, wavelength and aperture size.

What determines the dark fringes behind two slits?

Light reaches a point on a screen from two open slits. Closing one slit can make that point brighter. To predict the result, start with a wave field rather than the power arriving from each slit separately. In a homogeneous lossless medium a scalar component u obeys $\nabla^2 u = v^{-2}\partial_{tt}u$. A monochromatic solution can be written $u = \mathrm{Re}\{A(\mathbf{x})e^{-i\omega t}\}$. The phase of A advances by kr along a path of length r, with $k = 2\pi/\lambda$ and $v = \omega/k$.

The wave equation is linear, so complex amplitudes add. A detector averaging over many cycles records an intensity proportional to the squared magnitude. For two waves of the same frequency and polarisation,

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\Delta\phi.$$

This last term records the relative phase. Random changes of relative phase over a measurement wash it out.

Go from path difference to angular position

Let two identical narrow slits be separated centre to centre by d and illuminated normally and in phase. In the far field the paths at angle θ to the screen normal differ by $d\sin\theta$. Write the two equal amplitudes as $A_0 e^{\pm ikd\sin\theta/2}$. Their sum is $2A_0\cos(kd\sin\theta/2)$, apart from a common phase. Destructive interference therefore occurs at $d\sin\theta = (m + 1/2)\lambda$, with integer m when the angle exists.

A real slit has width w. Every point x across it contributes phase $kx\sin\theta$. Integrating $e^{ikx\sin\theta}$ from $-w/2$ to $w/2$ gives $w\sin\beta/\beta$,

$$\beta = \pi w\sin\theta/\lambda.$$

Uniformly illuminated rectangular slits then give the normalised intensity

$$\frac{I(\theta)}{I(0)} = \left(\frac{\sin\beta}{\beta}\right)^2 \cos^2\left(\frac{\pi d\sin\theta}{\lambda}\right).$$

The first factor sets the envelope. The second makes the fine fringes. The expression assumes equal coherent illumination, a common polarisation and negligible change of amplitude across each slit. For an observation distance z much larger than $(d + w)^2/\lambda$, the quadratic variation of propagation phase across this aperture can be neglected.

Predict a fringe spacing and a missing order

Use red light of wavelength 600 nm, $d = 0.20$ mm, $w = 0.040$ mm and a screen 2.0 m away. Here $(d + w)^2/\lambda = 0.096$ m, well below the screen distance. At small angles the screen position is $y \simeq z\sin\theta$. The first dark fringe is at $z\lambda/(2d) = 3.0$ mm. Adjacent nominal bright orders are 6.0 mm apart.

The first single-slit zero lies at $z\lambda/w = 30$ mm. It coincides with the fifth positive interference order, so that order is missing even with perfect coherence. Changing slit width can remove an order without changing the

underlying spacing of the interference fringes.

What sets the resolving power of a lens?

Apply the aperture integral to a different observation. Two distant point sources viewed through a lens each produce a diffraction pattern in the focal plane. For a rectangular pupil of width W, a single source produces a first zero at angle λ/W in the direction across that width. Translating a second source through an angle $\Delta\theta$ shifts its entire pattern by $f\Delta\theta$ for focal length f.

When a second maximum falls at the first zero of the first source, the angular separation is about λ/W (the Rayleigh criterion for this pupil). With $W = 5.0$ mm, $\lambda = 600$ nm and $f = 100$ mm, this is 1.2×10^{-4} rad or 12 µm between image peaks. For mutually incoherent sources the intensities of the two patterns add. In the slit experiment the coherent amplitudes added. This distinction determines whether overlapping light shows interference fringes. At an interface the relative strength of each optical polarisation can change. Polarisation and reflection at glass.

Exercises on visibility and wavelength

1. Distinguish imbalance from lost coherence

At the observation point one slit alone gives intensity $4I_0$, the other I_0. The phase can be varied. What are the extremes and fringe visibility for perfectly coherent, parallel-polarised illumination? What would completely random relative phase do?

▼ Calculate the extremes

The individual amplitudes have ratio 2:1. In phase,

$$I_{\max} = (2 + 1)^2 I_0 = 9I_0.$$

Out of phase,

$$I_{\min} = (2 - 1)^2 I_0 = I_0.$$

The visibility $(I_{\max} - I_{\min})/(I_{\max} + I_{\min}) = 0.80$. A nonzero minimum can therefore coexist with perfect coherence. With a uniformly random relative phase the time-averaged intensity is $5I_0$ throughout the phase scan and visibility vanishes.

---

2. Change the light source

Replace 600 nm light by 450 nm light in the worked arrangement. Find the first dark position, the first envelope zero and the index of the first missing bright order.

▼ Compare the two length scales

Both screen distances scale with wavelength at small angle, giving $3.0(450/600) = 2.25$ mm and $30(450/600) = 22.5$ mm. The condition for coincidence of a bright order and the first envelope zero gives $m = d/w = 5$, independent of wavelength.

ONESIMULATION LABORATORY

Find the dark fringes of a double slit

Compare recorded OneSimulation runs on this page. To test your own inputs, download the notebook and model files and run the model in an activated OneSimulation environment.

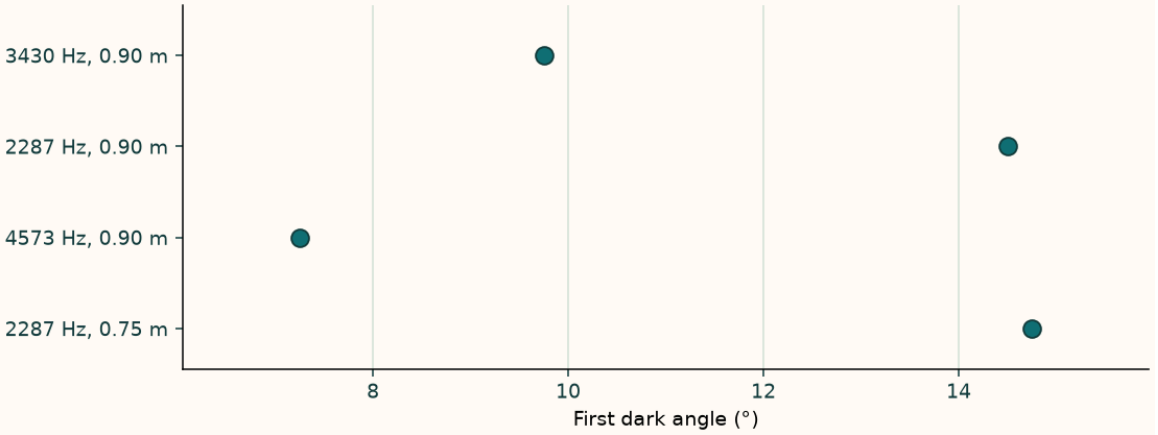
The numerical example drives two acoustic slits in phase. Sound pressure obeys the scalar Helmholtz equation, so the path-difference and aperture calculations transfer directly for the same d/λ and w/λ. This model uses rigid acoustic walls and has no optical polarisation. Its receivers lie at a finite distance of 0.90 m. Two-slit pressure field and angular probes.

Run the notebook with its supplied `ratios = [2, 3, 4]` sweep, $d = 0.30$ m, $w = 0.050$ m and sound speed 343 m/s. Each frequency is a separate solve on the same mesh. Read the first minimum on the positive-angle side of the pressure-amplitude profile. Squaring the pressure magnitude gives a mean-square pressure pattern with the same minima. In a locally travelling plane wave, mean-square pressure divided by acoustic impedance gives the mean intensity.

1. **Baseline.** At 3430 Hz, $d = 3\lambda$. Predict $\arcsin(1/6) = 9.59^\circ$ and compare with the computed minimum.
2. **Lower the frequency.** At $d = 2\lambda$, predict the first minimum at 14.48°. Check that the separation between the two dark fringes nearest the centre increases.
3. **Raise the frequency.** At $d = 4\lambda$, expect 7.18°. Check all three first minima within 0.5° of the far-field estimate. Then compare the outer minima and report where the finite observation distance and the assumed uniform slit illumination begin to matter.

Minima sampled every 0.25° on the receiver arc

Frequency and receiver radius	First dark angle (°)	Second dark angle (°)
3430 Hz, 0.90 m	9.75	30.50
2287 Hz, 0.90 m	14.50	48.75
4573 Hz, 0.90 m	7.25	22.50
2287 Hz, 0.75 m	14.75	49.00



The first dark angle moves inward as frequency rises at the same receiver radius. Moving the receiver closer at fixed frequency shifts it slightly outward.

To test the far-field assumption independently of frequency, move the receiver arc inward from 0.90 to 0.75 m. Evaluate both the computed pressure and the Rayleigh reference on the new arc. The simple far-field angle depends only on wavelength and slit spacing, so any change in the sampled minimum tests the finite-distance approximation.

Download experiment files (ZIP)

Extract the files together and open `double_slit_diffraction.ipynb` in a local OneSimulation environment.

Try it · Finite slits at the measured receiver radius

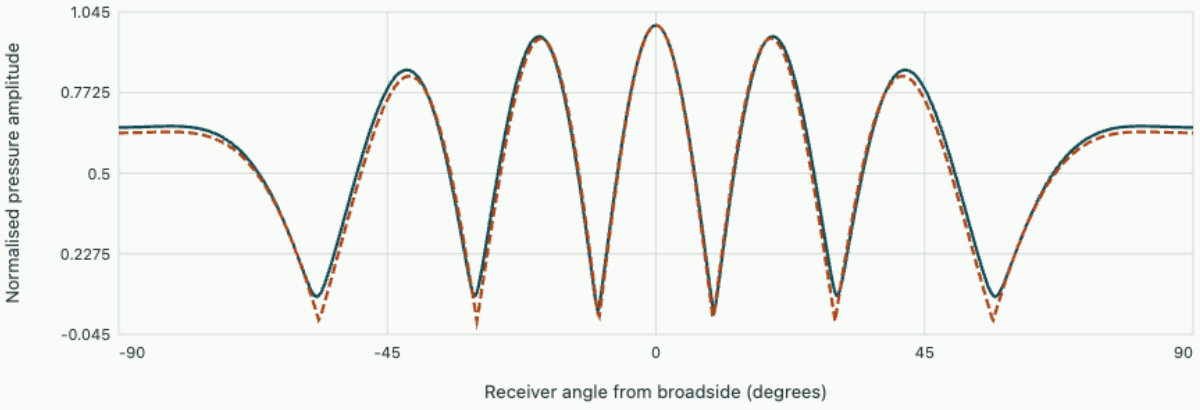
The reference uses the same 0.30 m slit spacing, 0.050 m slit widths and 0.90 m receiver radius as the OneSimulation model. Change separation/wavelength from 3 to 2 or 4, then move the receiver to 0.75 m. This is the notebook's finite-radius, two-dimensional Rayleigh integral with uniform aperture velocity, compared with its Fraunhofer limit. The full guide-and-screen solve additionally resolves the nonuniform slit velocity, so its sampled minima can differ slightly.

Open the corresponding OneSimulation notebook, code and output

Open this live calculation

Try the calculation

Slit separation / wavelength	**3**
Receiver radius (m)	**0.9**
Slit width (m)	**0.05**



Frequency 3430 Hz; spacing 0.30 m; sound speed 343 m/s. First two sampled minima 9.75°, 30.25° (0.25° sampling). This is the notebook's 2D Hankel Rayleigh reference with uniform slit velocity; the full OneSimulation guide/screen solution also resolves edge-dependent aperture velocity.

Problem sheet 4 · Infer slit spacing from a sampled angular minimum

Notation and assumptions

Read each defining equation with its assumptions and units, then follow the link to the argument that gives the equation its meaning.

Fringe visibility

$$\mathcal{V} = (I_{\max} - I_{\min})/(I_{\max} + I_{\min})$$

Use maxima and minima from the same scan. Unequal intensities can reduce contrast in a fully coherent field.

Derivation · Coherence and interferometry

Gaussian Rayleigh range

$$z_R = \pi w_0^2/\lambda$$

Here the waist is the one-over-e field radius and the wavelength is in the medium. At one Rayleigh range its squared beam radius doubles.

Derivation · Gaussian beams and focusing

Group velocity

$$v_g = d\omega/dk$$

This is the local slope of the dispersion curve. A sufficiently narrow-band packet with weak distortion can travel with its envelope at this velocity.

Derivation · Wave packets and dispersion

Numerical aperture

$$\mathrm{NA} = n \sin\theta$$

The largest accepted angle controls which transverse spatial frequencies reach the pupil. Resolution also depends on the optical imaging and coherence assumptions.

Derivation · Images through a finite pupil

Optical path length

$$\mathcal{L} = \int n\, ds$$

Differences of this quantity determine phase delays in geometrical optics. For a pulse in a dispersive medium one must also account for group delay.

Derivation · Optical paths, lenses and images

Wavenumber

$$k = 2\pi/\lambda$$

The wavenumber gives phase change per unit distance in a homogeneous medium. Wavelength must be specified in the medium in which propagation is calculated.

Derivation · Travelling waves and energy on a string

Further reading

- Work through the finite-pupil imaging and coherence material. Compare the predictions of a ray diagram and a diffracting aperture for two closely spaced objects. TU Delft · Interactive Optics
- Start with the normal-mode problems and follow their continuum limit to a travelling wave. Use the examinations for practice with unfamiliar boundary conditions. MIT · Vibrations and Waves

Compare specific topics with the university sources