Waves, radiation and relativity

Follow a disturbance from its source to a measurement. Work through polarisation, dispersion, scattering and changes of observer using measurable quantities.

Contents

1. **From a local field balance to travelling waves and standing modes**
   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.
2. **What an analyser measures about polarisation and coherence**
   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.
3. **Oblique interfaces and films whose reflections cancel**
   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.
4. **Causal material response and the propagation of a finite pulse**
   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.
5. **A guided mode, a cavity line and the energy they carry**
   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.
6. **Constructing the retarded Green function from an outgoing spherical shell**

A short source pulse reaches two receivers at different times and amplitudes. Derive the signal from a localized spacetime impulse, see why the potential is evaluated at a past time, and identify which observables change when the source occupies an extended region.

7. Radiating multipoles and what a detector actually integrates

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.

8. Scattering, extinction and the phase of the forward wave

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.

9. Diffraction from the near field to the focal plane

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.

10. What different observers measure

A moving observer carries clocks, rulers and a calibrated test charge. A statement about the electric field must specify which observer held that test charge at rest. In this chapter we derive how different measurements can agree on the force law and yet assign different electric fields, magnetic fields and charge densities to the same physical apparatus. A current-carrying wire and a light pulse will provide independent checks.

11. Relativistic particle dynamics and the field tensor

A circular track measures momentum per charge. A clock following the particle registers proper time, and a detector records how much kinetic energy it has gained. We will put these observations in one dynamical equation, solve that equation in uniform fields and exhibit a numerical pusher that can get the energy right while its orbit phase slowly slips.

12. Radiated power, mechanical work and energy limits

A charge can radiate while its speed is constant. A source can exchange energy with its nearby field before a pulse reaches a receiver. We will use the flux through a distant sphere, the work at the charge and the energy stored in a local field as three operationally distinct measurements. Agreement requires the relevant integration times and the range of validity of the particle model.

13. How can a field that decays in air transmit power to another prism?

Place a thin air gap between identical prisms. Derive the TE transmission from the two interface conditions and test when an exponential attenuation length predicts the actual throughput.

14. Resolve the self-force before taking a point-charge limit

A finite spherical charge distribution introduces a propagation time. Follow its memory kernel to electromagnetic inertia, radiative recoil and the limits of a truncated equation.

15. From light clocks to collision kinematics

Construct spacetime measurements before transforming fields. Use the same invariant interval to compare clocks, add velocities and close a relativistic collision.

Waves, radiation and relativity · 1. From a local field balance to travelling waves and standing modes

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.

$$v = (\epsilon\mu)^{-1/2},$$

$$Z = \sqrt{\mu/\epsilon},$$

$$\partial_t(E_x + ZH_y) + v\partial_z(E_x + ZH_y) = 0.$$

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.

$$\sin(\omega s/2) = C\sin(kd/2),$$

$$\frac{v_{\mathrm{phase}}}{u} = \frac{\arcsin[C\sin(kd/2)]}{Ckd/2},$$

$$\frac{v_{\mathrm{group}}}{u} = \frac{\cos(kd/2)}{\sqrt{1 - C^2\sin^2(kd/2)}}.$$

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

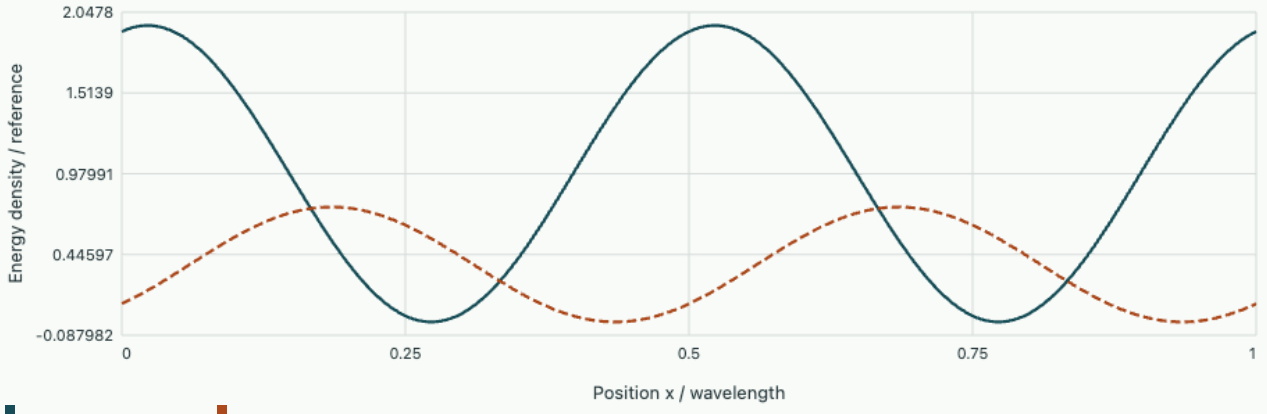
Change the balance of waves travelling in opposite directions. Identify a point where the instantaneous field is large while net power transfer is small.

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

Waves, radiation and relativity · 2. What an analyser measures about polarisation and coherence

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

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

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.

$$2\psi = \operatorname{atan2}(S_2, S_1),$$

$$\sin(2\chi) = S_3/s.$$

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.

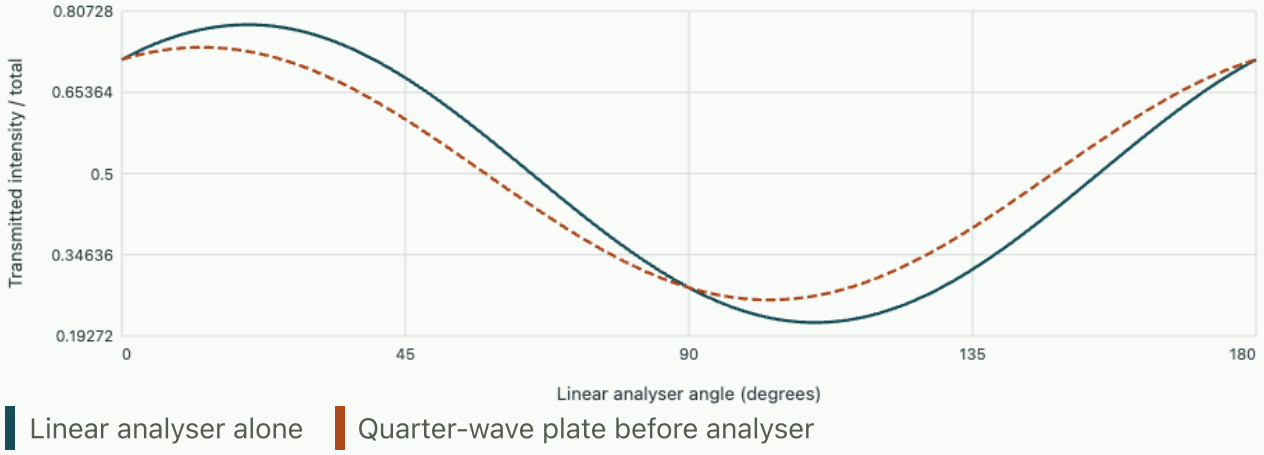
Open this experiment online

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 $S_3^2 \leq 6^2 - 2^2 - 3^2 = 23$, so $|S_3| \leq 4.796$. 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

Waves, radiation and relativity · 3. Oblique interfaces and films whose reflections cancel

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

$$Y_{\text{in}} = \frac{Y_b\cos\delta - iY\sin\delta}{\cos\delta - i(Y_b/Y)\sin\delta}.$$

Reflection from the whole film at its entrance now follows from the same single-interface equation with $Y_2 = Y_{\text{in}}$. At a quarter wave, $\delta = \pi/2$, this simplifies to $Y_{\text{in}} = Y^2/Y_b$. 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 $\delta = 2\pi n d \cos\theta_{\text{film}}/\lambda_0$. 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.

$$\frac{I_{\mathrm{out}}}{I_{\mathrm{in}}} = \frac{(1-R)^2}{|1 - Re^{i\delta}|^2} = \frac{1}{1 + F\sin^2(\delta/2)},$$

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

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

$$T_{\text{peak}} = \frac{(1-R)^2 a}{(1-Ra)^2},$$

$$g_E = R^2 a^2.$$

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

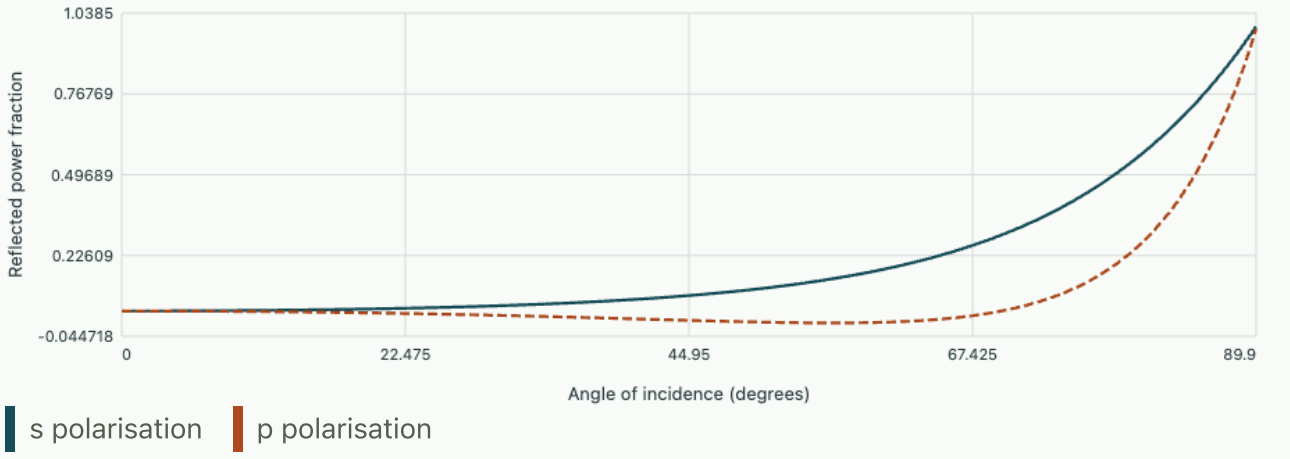
Change incidence angle and the receiving medium. Compare amplitude reflection with the power that actually crosses the interface.

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

$$\lambda_0/1.149 = 522\,\text{nm}.$$

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

Waves, radiation and relativity · 4. Causal material response and the propagation of a finite pulse

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

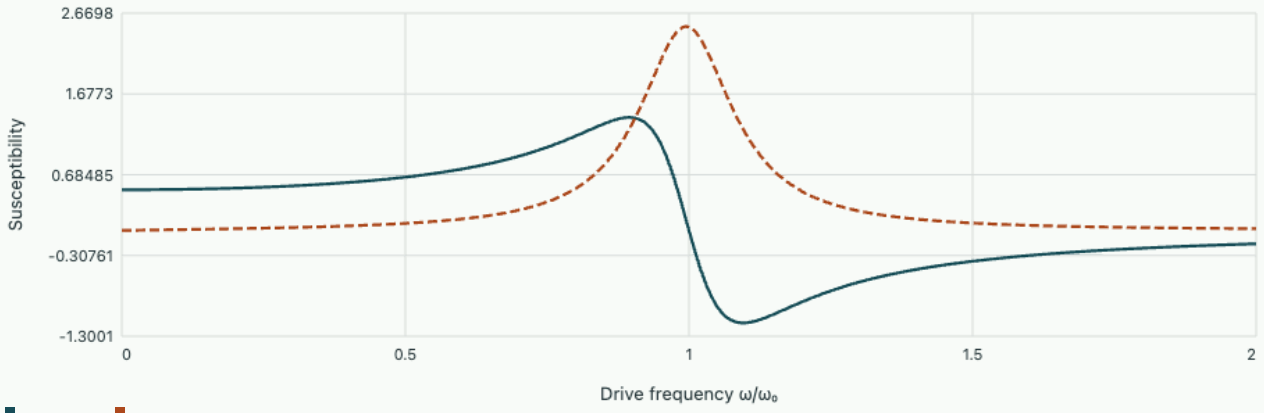
Sweep the drive through a material resonance. Follow the phase response and loss separately before interpreting a pulse arrival time.

Open this experiment online

Try the calculation

Damping γ/ω₀ **0.2**

Oscillator strength ωₚ²/ω₀² **0.5**



Re χ Im χ

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

Waves, radiation and relativity · 5. A guided mode, a cavity line and the energy they carry

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

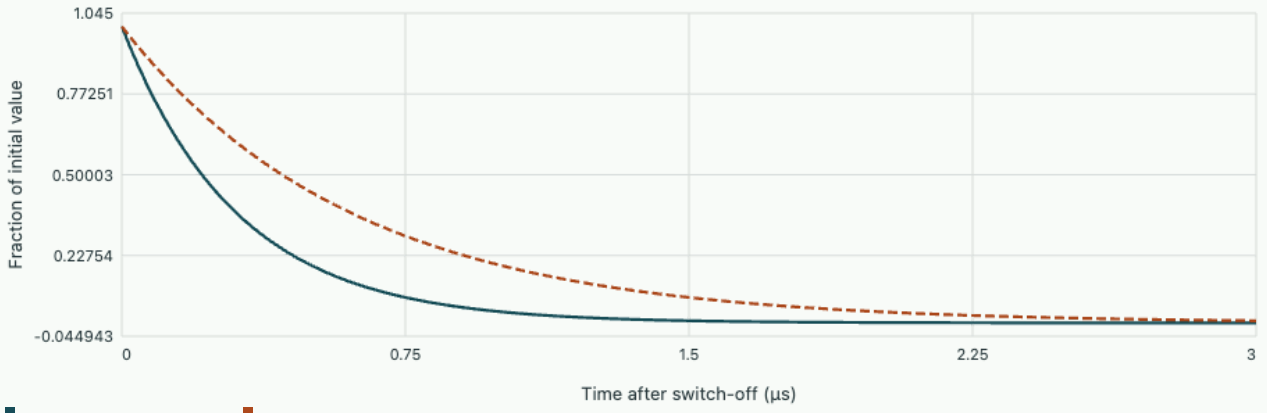
Change loss at a selected cavity mode and compare linewidth with energy decay. State whether each reported rate describes amplitude or energy.

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

Waves, radiation and relativity · 6. Constructing the retarded Green function from an outgoing spherical shell

A short source pulse reaches two receivers at different times and amplitudes. Derive the signal from a localized spacetime impulse, see why the potential is evaluated at a past time, and identify which observables change when the source occupies an extended region.

Turn the curl constraints into wave equations with sources

Use vacuum SI units, a localized charge density ρ and current density $\mathbf{J}$ satisfying $\partial_t \rho + \nabla \cdot \mathbf{J} = 0$. Introducing $\mathbf{B} = \nabla \times \mathbf{A}$ enforces the absence of magnetic charge. Setting $\mathbf{E} = -\nabla\Phi - \partial_t \mathbf{A}$ then enforces Faraday's law. These definitions leave freedom to add $\nabla\Lambda$ to $\mathbf{A}$ while subtracting $\partial_t\Lambda$ from Φ. An individual value of scalar or vector potential therefore depends on convention, whereas electric and magnetic fields constructed from them do not.

Adopt the Lorenz condition $\nabla \cdot \mathbf{A} + c^{-2}\partial_t \Phi = 0$. Insert the potentials into Gauss's and Ampère's laws and use $\nabla \times \nabla \times \mathbf{A} = \nabla\nabla\cdot\mathbf{A} - \nabla^2\mathbf{A}$. The gradient and time-mixed terms cancel in this gauge:

$$(\nabla^2 - c^{-2}\partial_t^2)\Phi = -\rho/\epsilon_0, \qquad \left(\nabla^2 - c^{-2}\partial_t^2\right)\mathbf{A} = -\mu_0 \mathbf{J}.$$

The same scalar propagation operator acts on each component. Taking a divergence and time derivative of these equations shows that the Lorenz-condition combination itself obeys a source-free wave equation when current is conserved. If it vanishes initially with compatible time derivative and boundary conditions it remains zero. Independently prescribing an oscillating point charge and zero current would violate continuity and can undermine both this consistency and any computed radiation field.

Determine the radial form and its source strength

Define a causal distribution $G(\mathbf{R}, \tau)$ solving $(\nabla^2 - c^{-2}\partial_\tau^2)G = -\delta^3(\mathbf{R})\delta(\tau)$ and vanishing for $\tau < 0$. In free space rotational symmetry makes the impulse response depend on $R = |\mathbf{R}|$. For $R > 0$,

$$\nabla^2 G = R^{-1}\partial_R^2 (RG).$$

Thus RG satisfies the ordinary one-dimensional wave equation. The solution for an outward-only signal from the origin has the form $g(\tau - R/c)$. An incoming term $g_{\mathrm{in}}(\tau + R/c)$ would specify radiation arriving from infinity. It can occur with an incident field, but is excluded for the initially quiet source experiment.

To fix g, integrate the equation over a small sphere of radius a and pair its time dependence with a smooth test function. The flux of ∇G through that sphere has leading term $-4\pi g(\tau - a/c)$. Additional terms proportional to $a\,g'$ tend to zero as distributions; the integral of the time derivative likewise vanishes as the sphere shrinks. The prescribed delta requires $-4\pi g(\tau) \to -\delta(\tau)$. Consequently

$$G(\mathbf{R}, \tau) = \frac{\delta(\tau - R/c)}{4\pi R}.$$

This normalization depends on which operator was given unit source. If instead one uses $\partial_t^2 - c^2\nabla^2$, the corresponding unit impulse response carries an additional $1/c^2$. The energy spreading as R^{-2} does not replace the field amplitude R^{-1}; the two concern different observables. In three spatial dimensions the free-space impulse is concentrated on the light cone. This sharp shell property need not hold in a material with memory, a waveguide, or a two-dimensional scalar wave problem.

Perform the time integral before interpreting retardation

A distributed source is a superposition of shifted impulses. Convolve each charge or current component with G over $\mathbf{r}'$ and t', keeping the observation point $\mathbf{r}, t$ fixed. The delta sets $t' = t - R/c$, with $R = |\mathbf{r} - \mathbf{r}'|$. Because t' occurs linearly for a fixed spatial source coordinate, this integration has unit Jacobian. The potentials are

$$\Phi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}', t - R/c)}{R}\, d^3r',$$

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(\mathbf{r}', t - R/c)}{R}\, d^3r'.$$

Every source point is sampled at its own retarded time. Replacing the whole extended source by its history at $t - r/c$ discards phase variation across it. This may be acceptable only when that variation is small or has been explicitly integrated. A charge distribution fixed in space and time recovers the ordinary Coulomb potential after the initial transient passes. The specified radiation condition selects a particular solution; a homogeneous incident pulse may be added if the actual experiment includes one.

For the $e^{-i\omega t}$ convention the frequency-domain kernel is $e^{ikR}/(4\pi R)$ with $k = \omega/c$. Substituting a harmonic source into the time-domain potential gives the positive outward radial phase directly. The incoming Helmholtz kernel $e^{-ikR}/(4\pi R)$ solves the same differential equation away from the origin but has the opposite boundary condition at infinity. One physical check is that outgoing radiation transports energy through a large sphere away from the source.

Background · Why does an ideal line source leave a tail after the front?

The three-dimensional scalar retarded kernel for the wave operator $c^{-2}\partial_t^2 - \nabla^2$ has support on the light cone. Now switch on an impulse simultaneously along an ideal infinite straight line. At perpendicular distance ρ, the nearest segment contributes first, at $t = \rho/c$. More distant segments continue to arrive even though each point source emitted only once. Integrating the point kernel along the line gives, for $t > \rho/c$,

$$G_{\text{line}}(\rho, t) = \int_{-\infty}^{\infty} \frac{\delta(t - \sqrt{\rho^2 + z^2}/c)}{4\pi\sqrt{\rho^2 + z^2}}\, dz = \frac{1}{2\pi\sqrt{t^2 - \rho^2/c^2}}.$$

Before the first arrival the integral vanishes. At later times its two contributing segments lie at $z = \pm\sqrt{c^2t^2 - \rho^2}$. This geometrical tail can be present even with a nondispersive propagation speed. A finite line length cuts it off when the contributions from the ends have arrived.

Jefimenko equations from differentiating the source clock

Let brackets denote source quantities at $t_R = t - R/c$ and $\widehat{\mathbf{R}} = (\mathbf{r} - \mathbf{r}')/R$. At fixed observation time,

$$\nabla R = \widehat{\mathbf{R}},$$

$$\nabla t_R = -\widehat{\mathbf{R}}/c.$$

Therefore differentiating $[\rho]/R$ gives both a derivative of inverse distance and a derivative of charge history. Combining $-\nabla\Phi$ with $-\partial_t\mathbf{A}$ leads to

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\int\left(\frac{[\rho]\widehat{\mathbf{R}}}{R^2} + \frac{[\dot\rho]\widehat{\mathbf{R}}}{cR} - \frac{[\dot{\mathbf{J}}]}{c^2R}\right)d^3r'.$$

Similarly take the curl of $[\mathbf{J}]/R$. The gradient of $1/R$ crossed into J gives $[\mathbf{J}]\times\widehat{\mathbf{R}}/R^2$, while the retarded-time derivative gives $[\dot{\mathbf{J}}]\times\widehat{\mathbf{R}}/(cR)$. The terms individually contain charge and current derivatives, but exact continuity produces important cancellations. For example, a spherically symmetric source with conserved total charge cannot radiate a time-varying electric monopole in vacuum. Outside a symmetric redistribution Gauss's law fixes the radial field by the constant total charge. Predicting radiation from just one term of the retarded electric integral would obscure that cancellation.

Use delay and geometrical spreading as separate diagnostics

Take an ideal compact scalar emitter as a model of a single outgoing far-field polarization trace, with waveform $s(t) = s_0 e^{-(t/T)^2}$. A receiver measuring that field with negligible bandwidth limitation reads $Cs(t - r/c)/r$. At 0.60 m the peak is delayed 2.00 ns and at 1.80 m it is delayed 6.00 ns after emission. The field peak falls by three, so power flux falls by nine. Multiplying each waveform by its distance and shifting its time origin back by r/c should collapse the two records. A mismatch in width after doing so suggests dispersion, source extent, detector bandwidth, multipath or near-field contributions. A single peak-height comparison does not distinguish these effects.

Suppose the emission has compact duration 4 ns and receivers are separated radially by 1.20 m. Their onset times differ by 4.00 ns. If identical receivers have a 1 ns electrical delay, their differential time of flight remains unchanged; if their amplifiers have different group delays, the difference directly biases the inferred propagation speed. A pair of distances can determine one offset and one speed if their absolute radii are known. Three or more distances also permit checking whether a straight-line timing model fits the data.

Translate and rescale two records of a compact retarded pulse

Keep the directional phase when the source is far away

Write $\mathbf{r} = r\mathbf{n}$ with source extent d and $r \gg d$. Expand the distance as $R = r - \mathbf{n}\cdot\mathbf{r}' + [r'^2 - (\mathbf{n}\cdot\mathbf{r}')^2]/(2r) + \cdots$. Relative amplitude variation in $1/R$ is order d/r. The first directional phase $k\mathbf{n}\cdot\mathbf{r}'$ can remain large even when d/r is tiny. For a harmonic current,

$$\widetilde{\mathbf{A}}(\mathbf{r}) \simeq \frac{\mu_0 e^{ikr}}{4\pi r}\int \widetilde{\mathbf{J}}(\mathbf{r}')e^{-ik\mathbf{n}\cdot\mathbf{r}'}\,d^3r'.$$

Neglecting the next, quadratic, phase requires kd^2/r small up to a geometry-dependent coefficient. This is a separate far-zone condition from $r \gg d$. The integral is a spatial Fourier component of the current, evaluated on the sphere of radius k. Measurements from one propagation direction reveal one projected spatial frequency and generally cannot reconstruct an arbitrary extended source.

For two equal emitters separated by d along z, observed at angle θ to z, the two contributions differ by delay $d\cos\theta/c$. In harmonic operation with equal drive phase the amplitude factor is $2\cos(kd\cos\theta/2)$. At 1.00 GHz, $d = 0.150$ m is almost half a free-space wavelength. Along the source axis the two arrivals differ by 0.500 ns, nearly half a period, so the equal contributions cancel. Broadside they arrive together and add. For pulsed excitation with a duration much shorter than 0.500 ns, the end-on receiver sees two time-separated pulses instead of a permanent null across the whole spectrum. The time-domain and frequency-domain statements agree after applying the appropriate bandwidth.

Recover source time from a moving point instead of a fixed grid

The earlier time integral held the source coordinate fixed. A point following a trajectory contributes $\rho(\mathbf{r}', t') = q\,\delta^3[\mathbf{r}' - \mathbf{r}_q(t')]$. If one instead integrates the spatial delta first, the remaining condition is

$$g(t') = t - t' - |\mathbf{r} - \mathbf{r}_q(t')|/c = 0.$$

The source-to-observer distance now depends on the integration time. Its derivative is $-1 + \widehat{\mathbf{R}} \cdot \boldsymbol{\beta}$, giving

$$\int f(t')\delta(g(t'))dt' = \frac{f(t_{\mathrm{ret}})}{1 - \widehat{\mathbf{R}} \cdot \boldsymbol{\beta}_{\mathrm{ret}}}.$$

For subluminal motion this denominator is positive. It records how emission intervals map into observation intervals and must appear when evaluating retarded point-source potentials. Simply replacing position by retarded position in a static Coulomb formula misses this Jacobian. A charge moving towards a distant receiver compresses successive arrival intervals; transverse motion has no first-order time compression along that line of sight. Later units relate this observable to Doppler shifts and relativistic angular radiation patterns.

As a numerical example, two features emitted 10 ns apart by a compact laboratory source whose radial velocity component is $0.20c$ arrive about 8.0 ns apart at a distant fixed receiver, provided the direction and velocity vary negligibly during the interval. The same proper-time interval of a fast moving clock would also require time dilation in converting to the laboratory emission interval. Keeping those two clocks separate avoids applying the Doppler factor twice.

A causal Green function in front of a grounded plane

The free-space Green function encodes an outgoing wave from a point impulse. A boundary changes the response and can be built into the Green function before integrating over sources. As an exactly soluble example choose the scalar wave equation in $z > 0$ with a homogeneous Dirichlet condition on $z = 0$. This is a scalar field problem, and can describe a selected tangential electric component in a geometry whose conducting boundary sets that component to zero. It is not a universal image rule for every vector potential component. The boundary conditions on electromagnetic scalar and vector potentials depend on the chosen gauge and physical surface.

Place a source at $\mathbf{s} = (x_s, y_s, h)$ and its mirror at $\mathbf{s}^* = (x_s, y_s, -h)$. Every point on the plane is equidistant from these points. Hence two opposite-sign delayed spherical responses cancel there at every time. The mirror source lies outside the physical domain and introduces no delta function inside it. Both contributions vanish before their respective light cones. Uniqueness of the retarded boundary-value problem identifies their difference as the required Green function.

$$G_D(\mathbf{r},\mathbf{s},t)=\frac{\delta(t-R/c)}{4\pi R}-\frac{\delta(t-R^*/c)}{4\pi R^*},$$
$$R=|\mathbf{r}-\mathbf{s}|,$$
$$R^*=|\mathbf{r}-\mathbf{s}^*|.$$

A receiver at height 0.40 m is displaced horizontally 1.20 m from an impulse source at height 0.30 m. The direct path is $\sqrt{1.20^2+0.10^2}=1.204$ m and the reflected path is

$$\sqrt{1.20^2+0.70^2}=1.389\,\mathrm{m}.$$

The reflected pulse follows the direct one by 0.617 ns, with opposite sign and amplitude magnitude about 0.867 of the direct amplitude for this scalar geometrical model. A receiver that only rectifies the signal will lose this sign information. For two overlapping pulses their sign then still affects intensity, so rectification before adding the predicted pulses is incorrect.

For a scalar Neumann condition one adds the mirror instead. Although the two path lengths are identical at the boundary, their normal derivatives there are opposite, including the derivative of the retarded-time argument. Thus the sum has zero normal derivative. A rigid acoustic wall and a pressure-release surface furnish experimentally familiar examples of different scalar boundary rules, but the field variable must be named before assigning a reflection sign. Two surfaces can yield the same return time despite opposite returned amplitudes.

If a source is monochromatic, the two impulses become terms e^{ikR}/R and $-e^{ikR^*}/R^*$. Where the amplitudes are nearly equal, a path difference equal to one wavelength gives nearly complete cancellation, while a half wavelength can give reinforcement because the boundary reversal supplies the other half cycle. The image solution therefore connects the time-domain echo test to a spatial fringe prediction. Moving the receiver a few centimetres may change interference dramatically without changing the material reflectance.

Can two receiver traces determine a short source waveform?

Let $g(t)$ denote the waveform emitted by a small scalar source. After correcting a receiver's direct-path amplitude and arrival time, the ideal plane response above has the form $y(t)=g(t)-b\,g(t-\Delta)$. Both $b=R/R^*$ and $\Delta=(R^*-R)/c$ are known from geometry. The transfer function for this normalised signal has modulus squared $1+b^2-2b\cos\omega\Delta$. Frequencies with $\omega\Delta=2\pi n$ have poor signal. Merely subtracting an independently smooth fitted baseline does not recover their phase.

For the preceding geometry $b=0.867$, so response amplitude ranges from 0.133 to 1.867 times the corrected direct-path amplitude. Unregularised division at a minimum magnifies amplitude noise by about 7.5. The spacing between minima is roughly 1.62 GHz. A second receiver at a different height changes b and Δ. Choose a geometry for which its minima avoid the first receiver's vulnerable frequencies across the band of interest. Two receivers at symmetric points with equal path lengths give more counts but no new spectral constraint.

Given complex measurements $Y_j(\omega) = H_j(\omega)G(\omega) + \eta_j$ with independent mean-square noise levels s_j^2, minimise $\sum_j |Y_j - H_j G|^2 / s_j^2$ at each frequency. Differentiating in the real and imaginary parts of G gives

$$\widehat{G} = \frac{\sum_j H_j^* Y_j / s_j^2}{\sum_j |H_j|^2 / s_j^2}.$$

The reciprocal denominator sets the estimated error variance under the model. If both instruments have a common phase calibration error it does not average down like independent noise. If both transfer functions are small, probabilistic regularisation or a finite source-support constraint may stabilise a fit, but the inferred high-frequency waveform then depends on the extra assumptions. Present the bandwidth over which the data genuinely constrain the result.

A clean time-domain cross-check is available when the pulse is shorter than the known path separation. Window the first arrival before any reflected contribution enters and predict both later echoes using the measured first pulse. If the first pulse is not fully separated, the retarded equation can still be solved forward in time by $g(t) = y(t) + b\, g(t - \Delta)$, starting from a known zero prehistory. Its noise after many recurrences accumulates, particularly for b near one. A frequency fit and this delayed recursion use the same physics and have related conditioning. Agreement over the accurately measured band with disagreement confined to a spectral minimum identifies an inverse problem limitation before it is mistaken for a new source feature.

Recover the source phase at the correct emission time

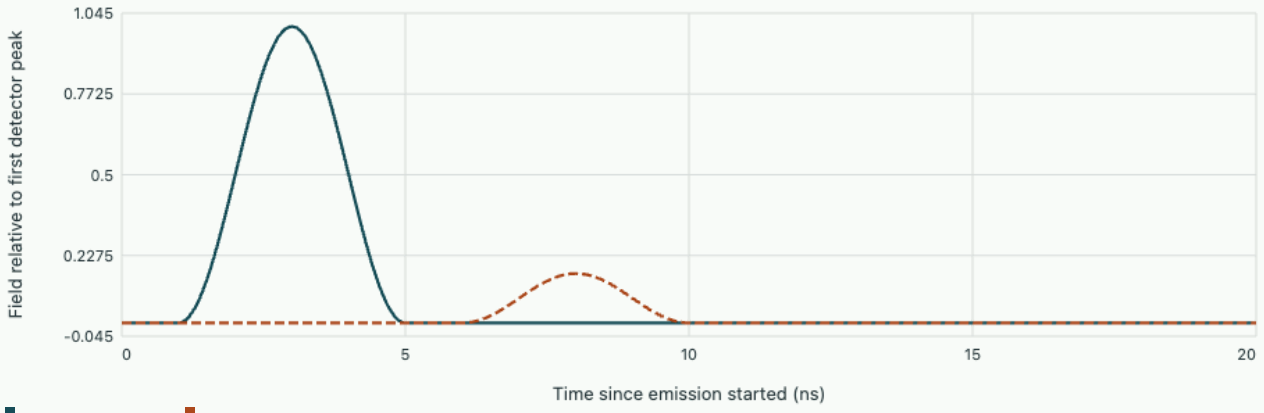
Move the observation point and compare its signal with the source evaluated at the observation time. Quantify the phase error of that shortcut.

Open this experiment online

Try the calculation

Second detector radius (m) **1.8**

Pulse duration (ns) **4**



The second trace starts 5 ns after the first. Its peak is 16.7% of the first peak.

Worked propagation problem · a cell-centred retarded-time approximation

A compact source is represented by cubic cells of edge $d = 2.0$ mm at frequency 5.0 GHz. Each cell is replaced by its centre value in the far-field sum. Estimate the error for a cell with uniform complex current when observed along a grid axis. Distinguish this amplitude error from uncertainty in arrival time.

▼ **Integrate the phase across a cell before summing**

Across one cell the leading far-zone phase is linear, e^{-ikx}, for observation along the local x axis. Averaging exactly across its width multiplies the centre approximation by $\sin(kd/2)/(kd/2)$. Here wavelength is approximately 60.0 mm and $kd/2 = 0.1048$, giving a factor 0.99817 or about 0.18 percent amplitude reduction. The two transverse integrals have no phase variation in this particular direction. An oblique look multiplies three sinc factors evaluated at the separate direction-cosine components. Spatially varying current or strongly curved wavefronts require additional corrections.

The travel time changes by $d/c = 6.67$ ps across the cube along the axis. The centre convention therefore leaves a maximum individual delay discrepancy of 3.34 ps. Its effect on the amplitude of a symmetric uniform cell begins at second order because the first-order phase terms cancel. That cancellation can disappear for a strongly varying current or a clipped boundary cell. Adding the same delay bound over every cell in the source would grossly overestimate a coherent error; conversely, relative error in a near-null total field can be very large even when each absolute contribution is accurate.

Refine spatial resolution while separately refining interpolation of the source history at non-grid retarded times. The former controls geometric phase and boundary representation, the latter controls the recovered temporal spectrum. Compare a signed field at a generic angle, integrated radiated power and absolute field near an interference minimum. No single relative-error statistic is satisfactory for all three. Reporting a tiny percentage error for the total power can hide a phase error that materially changes the inverse source estimate at selected receivers.

How a finite measurement surface can test retardation

A closed surface of time-resolved field probes can test a proposed source history against local energy conservation. Integrate the measured outward Poynting vector over the surface at each receiver time. Between two concentric surfaces in free space with no charges, the difference between outward powers equals the rate at which field energy in the intervening shell changes, with the sign determined by whether the shell is filling or emptying. During a leading pulse edge the outward flux at the smaller sphere can exceed that at the larger one without any violation of conservation.

With a compact central source and far-zone probes, rescale radiation amplitudes by radius and shift their signals by the radial travel-time difference before imposing pointwise agreement. For a radial separation of thirty centimetres the shift is approximately one nanosecond. A modest timing mismatch can produce an apparent amplitude discrepancy on a steep edge even when both probes are correctly calibrated. Integrating through the complete pulse removes that sensitivity if the integration windows include its entire support.

Residual disagreement can identify omitted directional dependence, environmental reflections or nonradiative near-field terms. Repeat at increasing radii and compare their scaling. A coherent field proportional to inverse squared radius becomes relatively weaker compared with the outgoing radiation, while a reflected contribution may vary nonmonotonically as phase changes. A delayed free-space Green function makes distinct predictions for these different tests.

Tutorial · infer an unknown delay and test the amplitude law

A detector moved successively to 0.45, 0.90 and 1.35 m records field-peak times of 2.20, 3.70 and 5.20 ns and field amplitudes 6.0, 3.0 and 2.2 mV/m. Estimate speed and common timing offset. Which amplitude is least consistent with a single outgoing far field, and what additional observation could distinguish an additive background from a receiver gain change?

▼ Fit the delay and audit the independent product

Each 0.45 m step adds 1.50 ns, giving $v = 3.00 \times 10^{8}$ m/s. Extrapolation to r=0 gives 0.70 ns offset. The products rE are 2.70, 2.70 and 2.97 mV, with the last 10% higher than the common value. A stationary additive background field could interfere coherently with the local signal; a gain error would scale the waveform. A source-off record tests an additive background under the same trigger. Moving the detector back to the first radius or swapping two calibrated receivers tests whether the excess

follows the position or the instrument. Arrival-time agreement alone cannot verify absolute amplitude calibration.

Tutorial · decide when one common retarded time suffices

A 60 mm source stands 3.0 m from a receiver near broadside. It radiates near 20 GHz in vacuum. Bound the relative amplitude variation from replacing $1/R$ by $1/r$. For a source lying in a transverse disk of radius 30 mm, estimate the largest omitted quadratic phase when evaluating a far-field integral. Would neglecting the entire directional phase be justified for observation 20° away from broadside?

▼ **Compare three different expansion parameters**

A conservative bound using the full size is order $0.060/3 = 0.020$; the actual broadside amplitude variation across a transverse disk is much smaller. At the rim the omitted optical path is approximately $0.030^2/(2 \cdot 3) = 0.00015$ m. Wavelength is about 15.0 mm, so the omitted phase is 0.063 rad. Off broadside along a diameter the retained linear path difference over 60 mm is $0.060 \sin 20^\circ = 0.0205$ m, or 8.60 rad. Constant amplitude and linearized phase can therefore be excellent approximations while treating all parts of the source as simultaneous at the receiver is seriously wrong.

David Tong, University of Cambridge · electromagnetic radiation, Green functions and retarded potentials

Waves, radiation and relativity · 7. Radiating multipoles and what a detector actually integrates

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 $p_0 = 1.0\times10^{-12}$ 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, $m_0 = 3.14\times10^{-6}$ A m² and the equivalent effective electric-moment magnitude for equal power is $m_0/c = 1.05\times10^{-14}$ 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 $Q_{ij} = 3D_{ij} - \delta_{ij} \operatorname{tr} D$. 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

$$f_{\text{pole}} = \frac{\int_0^{\theta_a} \sin^3\theta \, d\theta}{\int_0^{\pi} \sin^3\theta \, d\theta} = \frac{2 - 3\cos\theta_a + \cos^3\theta_a}{4}.$$

For small θ_a expand the integrand rather than subtracting nearly equal cosines in floating point: $f_{\text{pole}} \simeq 3\theta_a^4/16$. 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

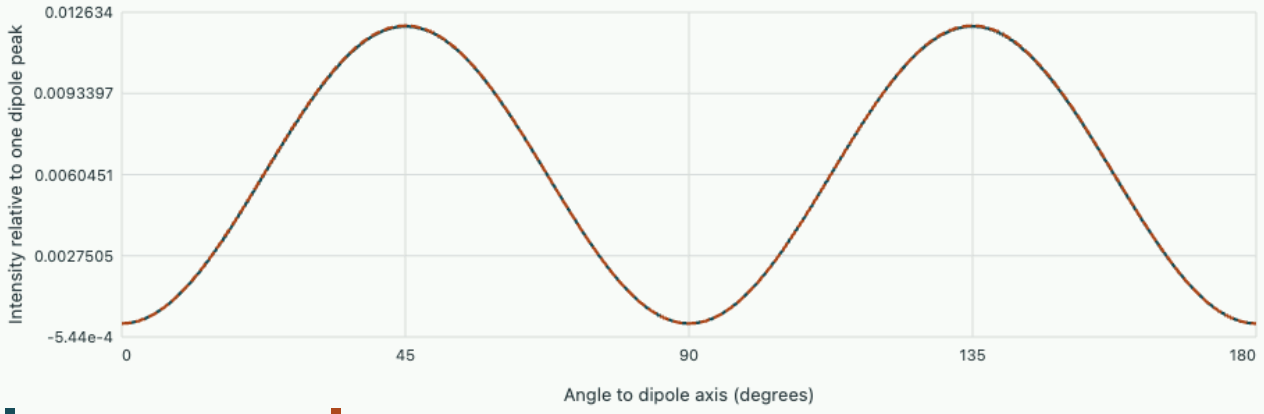
Change the separation or relative phase of the source elements and locate directions in which the leading radiation vanishes.

Open this experiment online

Try the calculation

Separation d/λ **0.035**

Opposing amplitude deficit **0**



Full phase difference Long-wavelength expansion

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

Waves, radiation and relativity · 8. Scattering, extinction and the phase of the forward wave

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

$$\mathbf{f}(\mathbf{n}) = \frac{k^2\alpha}{4\pi}\left[\mathbf{e}_0 - \mathbf{n}(\mathbf{n}\cdot\mathbf{e}_0)\right],$$

$$\sigma_{\mathrm{sca}} = \frac{k^4}{6\pi}|\alpha|^2.$$

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

$$\operatorname{Im} a_\ell - |a_\ell|^2 = \frac{1 - |S_\ell|^2}{4},$$

$$\sigma_{\mathrm{abs}} = \frac{\pi}{k^2}\sum_\ell (2\ell+1)(1-\eta_\ell^2).$$

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

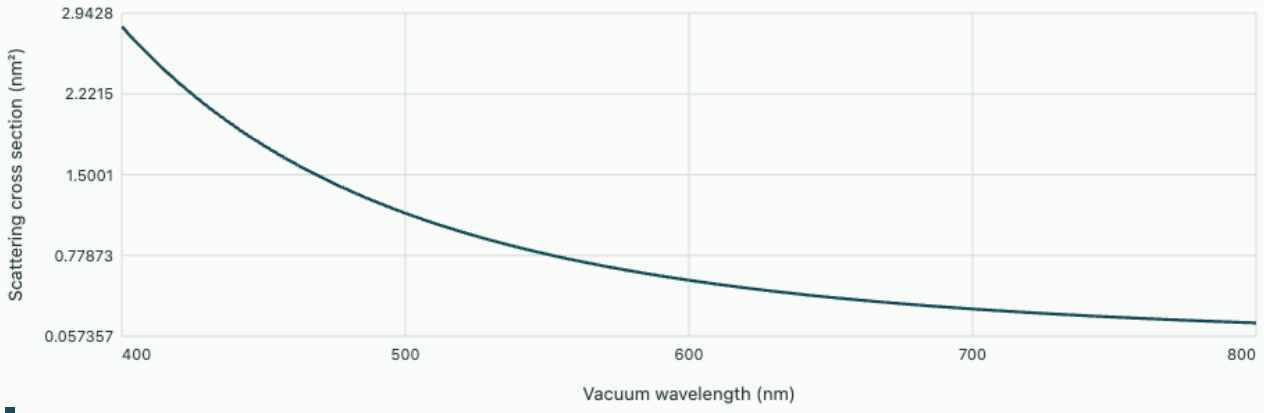
Change the ratio of forcing to resonance frequency. Decide which part of the response determines a scattering measurement and compare it with the power balance in the text.

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

Waves, radiation and relativity · 9. Diffraction from the near field to the focal plane

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

$$M(s) = \frac{2}{\pi}\left[\arccos s - s\sqrt{1 - s^2}\right], \quad 0 \le s \le 1,$$

$$M(s) = 0 \quad (s > 1).$$

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.

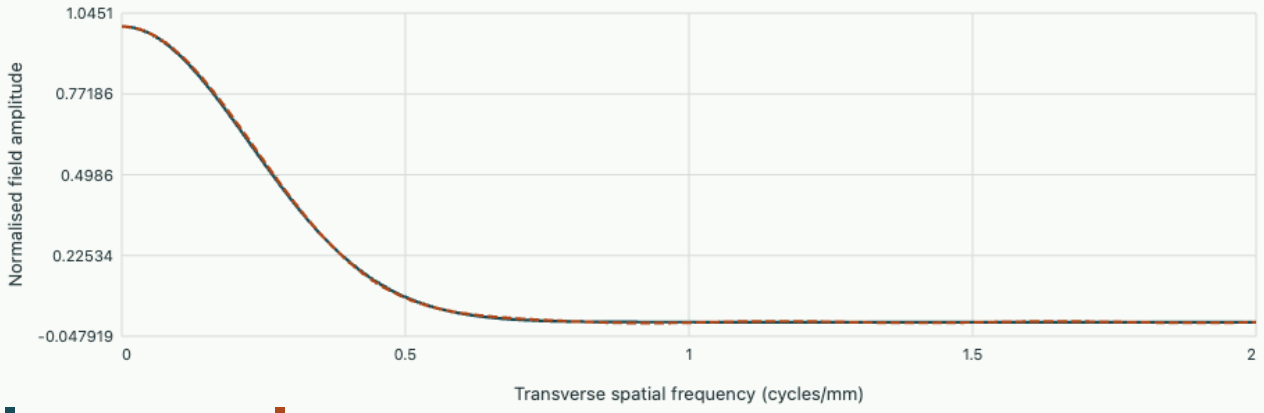
Open this experiment online

Try the calculation

Gaussian aperture radius w (mm) **1**

Window half-width / w **2**

Samples across window **30**



Infinite Gaussian Sampled finite window

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 $I/I_{\text{peak}} \simeq s^2/x_0^2$. 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

Waves, radiation and relativity · 10. What different observers measure

A moving observer carries clocks, rulers and a calibrated test charge. A statement about the electric field must specify which observer held that test charge at rest. In this chapter we derive how different measurements can agree on the force law and yet assign different electric fields, magnetic fields and charge densities to the same physical apparatus. A current-carrying wire and a light pulse will provide independent checks.

Events, synchronisation and directed boosts

Use two inertial coordinate systems with aligned axes. The primed origin moves at speed v in the positive laboratory x direction. Choose coincident origins when both clocks read zero and synchronise distant clocks in each frame by light signals. Homogeneity makes the transformation linear. The worldline of the primed origin requires

$$x' = A(x - vt).$$

A light ray moving along either axial direction must still obey $x' = \pm ct'$. Substituting its two possible worldlines into the linear map fixes the relative coefficients of x and t in t'. Reversing the transformation by changing the sign of v fixes the remaining scale.

$$x' = \gamma(x - vt),$$

$$t' = \gamma(t - vx/c^2),$$

$$y' = y,$$

$$z' = z,$$

$$\gamma = (1 - v^2/c^2)^{-1/2}.$$

The squared interval $c^2\Delta t^2 - |\Delta \mathbf{r}|^2$ is unchanged, a direct algebra check that detects a surprising number of sign choices. Two sparks separated by 3.0 m along a laboratory bench and simultaneous on laboratory clocks have $\Delta t' = -7.51$ ns for $v = 0.60c$, if the second event is at larger x. Their ordering in this frame is consistent with their spacelike separation. Events joined by a light signal are null-separated and retain their order under this boost.

Length measurements need two simultaneous endpoint readings in the chosen frame. A bar at rest in the lab and parallel to the axis has length L/γ in the primed frame. Substituting the simultaneous laboratory endpoint events into the boost instead gives γL, but their primed times differ. This distinction reappears when we count charge per unit length in a moving wire. The result of that count depends on the spacetime slice used to perform it.

A current is a rate of charge crossing a surface

Begin with a single population of particles, each carrying charge q, whose proper number density is n_0. Their counting volume is Lorentz-contracted only along their motion. The lab assigns density $n = \gamma_u n_0$

and current density $qn\mathbf{u}$. Both are components of $j^\mu = qn_0 u^\mu$, where $u^\mu = \gamma_u(c, \mathbf{u})$ is the four-velocity. Independent species can be superposed. A total four-current can consequently be spacelike even though every material particle moves more slowly than light.

$$\rho' = \gamma(\rho - vJ_x/c^2),$$

$$J'_x = \gamma(J_x - v\rho),$$

$$J'_y = J_y,$$

$$J'_z = J_z.$$

The usual continuity law becomes $\partial_\mu j^\mu = 0$ with $\partial_0 = c^{-1}\partial_t$. Transformation of the derivative and of the current cancels exactly in this contraction. An isolated finite charge distribution has an observer-independent total charge if each observer integrates through a complete spacelike slice. One way to see this is to connect the two slices by a large timelike enclosing wall. Four-dimensional flux conservation sets the difference between the two charge integrals equal to the current crossing the wall. For a wall outside all the charge trajectories that last contribution vanishes.

For a charge-neutral straight wire carrying conventional current $I = 2.0$ A along $+x$, a frame moving parallel to the wire at $0.20c$ assigns line density $\lambda' = -\gamma v I/c^2 = -1.36$ nC/m. Transverse cross-section is unchanged. There is no requirement that electrons themselves travel at $0.20c$. Both the mobile and immobile species contribute to the transformed density. The ideal wire is infinitely long, so one cannot integrate its transformed charge density out to infinity and apply the finite isolated system argument.

Obtaining field components from measured forces

Place a test charge at rest at an event in the primed frame. In the lab it has velocity $v\mathbf{e}_x$ and force $q(\mathbf{E} + \mathbf{v}\times\mathbf{B})$. To transform the force, transform the corresponding momentum change. For a small transverse kick, $dp'_y = dp_y$ while $dt' = dt/\gamma$ on the test charge's path. Thus $F'_y = \gamma F_y$ and similarly for z. For the parallel component use $dp'_x = \gamma(dp_x - v\,d\mathcal{E}/c^2)$ and $d\mathcal{E}/dt = vF_x$. The result is

$$F'_x = F_x.$$

Dividing by q gives the electric field. Requiring the same force law for charges moving transversely, or transforming the magnetic flux measured by a moving loop, gives the magnetic field.

$$\mathbf{E}'_\parallel = \mathbf{E}_\parallel,$$

$$\mathbf{B}'_\parallel = \mathbf{B}_\parallel,$$

$$\mathbf{E}'_\perp = \gamma(\mathbf{E} + \mathbf{v}\times\mathbf{B})_\perp,$$

$$\mathbf{B}'_\perp = \gamma(\mathbf{B} - \mathbf{v}\times\mathbf{E}/c^2)_\perp.$$

Imagine a uniform $+y$ electric field of 90.0 kV/m and a $+z$ magnetic field with $cB = 150$ kV/m. For a boost along $+x$, the y component of $\mathbf{v}\times\mathbf{B}$ is negative. At $\beta = 0.60$ the electric field vanishes. The

remaining magnetic field has $cB' = 1.25(150 - 54) = 120$ kV/m or 0.400 mT. A particle comoving with that frame feels no force at that event. A different particle moving across the magnetic field can still curve.

Vary the observer speed and compare the electric and magnetic field invariants

Compare two force measurements near an ideal wire

Return to the 2.0 A wire and put the observation point 1.0 cm along $+y$. Ampère's law gives $B_z = \mu_0 I/(2\pi r) = 40.0$ microtesla. The lab radial electric field is zero in this idealised symmetric model. A positive test charge moving along $+x$ at $0.20c$ experiences an inward magnetic force per unit charge of 2.40 kV/m. In its momentary rest frame the wire's negative line charge produces an inward radial electric field. The field from Gauss's law is $|\lambda'|/(2\pi\epsilon_0 r) = 2.45$ kV/m. The difference in the reported transverse forces is precisely the factor $\gamma = 1.0206$ derived from the momentum kick.

There is no contradiction between a field with zero laboratory divergence outside the wire and a different electric field in the primed frame. The charge density at the wire itself also transforms. If the laboratory wire carries a real driving surface charge and a return conductor, add their fields and charges before transforming. A claimed measurement of this ideal result without specifying the return path and electrostatic shielding needs a separate model for those additional fields.

When can a field be made purely electric or magnetic?

Direct substitution in the transverse transformation leaves the scalars $B^2 - E^2/c^2$ and $\mathbf{E}\cdot\mathbf{B}$ unchanged. The first is a difference, so retain its sign. If the second is nonzero, no observer sees either field vanish. If the fields are orthogonal and $E < cB$, choose $\mathbf{v} = \mathbf{E}\times\mathbf{B}/B^2$. The vector identity for $(\mathbf{E}\times\mathbf{B})\times\mathbf{B}$ cancels $\mathbf{E}$, and the required speed is below c. With $E > cB$, choosing $c^2\mathbf{E}\times\mathbf{B}/E^2$ instead cancels $\mathbf{B}$.

When both invariants vanish one must still distinguish a field identically zero from a nonzero null field. A vacuum travelling plane wave satisfies $E = cB$ and orthogonality everywhere. The advertised boost would then require speed c. An observer can lower its measured frequency and amplitude continuously by following the wave at increasing speed but can never transform away its nonzero field. A superposition of waves need not be null. A standing wave has positions and times where only one of the fields is present.

Phase matching across inertial frames

Every observer agrees on whether a particular spacetime event is a crest. The phase $\mathbf{k}\cdot\mathbf{r} - \omega t$ therefore labels the same event in each frame. Substituting the inverse coordinate transformation gives the transformation of frequency and wavevector. For vacuum propagation at laboratory angle θ to the boost, divide the transformed longitudinal wavenumber by the transformed magnitude.

$$\omega' = \gamma\omega(1 - \beta\cos\theta),$$

$$\cos\theta' = \frac{\cos\theta - \beta}{1 - \beta\cos\theta}.$$

An observer chasing a collinear 600 nm light wave at $0.60c$ sees half its frequency and a wavelength of 1200 nm. A beam entering at 90° to the boost appears at $\cos\theta' = -0.60$ and frequency 1.25ω. Asking for a transverse Doppler shift without saying in whose frame the angle is a right angle leaves the problem under-specified. A detector counts phase fronts crossing its own trajectory and its own clock sets the elapsed time.

A boosted light beam also transports momentum

Frequency alone does not specify the energy measured by a moving detector. Its collecting surface and clock must be assigned to an observer. For a monochromatic collimated beam in vacuum, one route is to transform both fields and calculate their new Poynting vector. Another route is to transform the stress-energy tensor, whose energy density, momentum density and momentum flux describe the same field together. This second route shows explicitly which measured quantity is being transformed and extends to superposed beams with zero mean momentum in one frame.

For a right-moving plane wave along x, the energy density is $u = S_x/c$, momentum density is u/c and normal momentum flux is u. In coordinates with $x^0 = ct$, the relevant block has $T^{00} = T^{01} = T^{10} = T^{11} = u$. Boost both indices using the same Lorentz matrix as for coordinates. The transformed energy density follows directly:

$$T'^{00} = \gamma^2(T^{00} - 2\beta T^{01} + \beta^2 T^{11}) = \gamma^2(1-\beta)^2 u.$$

The frequency Doppler factor enters twice. One factor corresponds to lower energy per photon and one to the changed number per unit volume. For a finite pulse, its length at a common primed time grows for an observer following it, so the total energy changes by only one Doppler factor. Volumes measured between equal-time spatial endpoints must be computed on the appropriate observer's simultaneity slice, just as for a moving rod.

Now superpose two equal-intensity mutually incoherent plane waves in opposite axial directions. In the laboratory $T^{01} = 0$, while $T^{11} = T^{00} = u$ for the combined radiation. An observer moving at $0.60c$ obtains $u' = \gamma^2(1+\beta^2)u = 2.125u$, with net flux $S'_x = -2\gamma^2\beta\,cu = -1.875\,cu$. The counterpropagating beam that the observer meets head-on is blueshifted, while the other is redshifted. Energy density remains greater than the magnitude of momentum density times c, as required for a superposition of distinct propagation directions. This check fails if one transforms a zero Poynting vector as though it were the spatial component of a four-vector with no stress information.

The direction-independent radiation field in an ideal cavity has $T^{ii} = u/3$ for each spatial axis in its rest frame rather than $T^{11} = u$. The one-third comes from averaging n_i^2 over a sphere. The moving observer then sees $u' = \gamma^2 u(1 + \beta^2/3)$ and a net longitudinal flux. At $\beta = 0.60$ the energy-density

factor is 1.75. An isotropic thermal field becomes directionally distorted; assigning one Doppler factor to the whole sky would give neither the correct integrated energy density nor its directional spectrum.

A moving parallel-plate capacitor tests surface sources

Choose large parallel conducting plates at $z = \pm d/2$ carrying opposite uniform surface densities $\pm\sigma$. Ignore edge effects in the central region. In the lab the field between them has magnitude $E_0 = \sigma/\epsilon_0$ and there is no magnetic field. A frame boosted along $+x$ sees both plates move at $-v$. Their areas contract along x, so their surface charge magnitudes become $\gamma\sigma$; their equal and opposite surface currents point in opposite senses. Use Gauss's law and the magnetic jump condition on each sheet to reconstruct the transformed fields. The current sheets reinforce magnetically between the plates and cancel outside, just as electric sheets of opposite charge do electrostatically.

For laboratory electric field along $+z$, the moving observer has $E'_z = \gamma E_0$ and $B'_y = +\gamma v E_0/c^2$. Cross the primed plate velocity $-v\mathbf{e}_x$ with this magnetic field to find a force along $-z$ on a positive surface charge, partly cancelling the enhanced electric attraction towards the opposite plate. When finding the force on a sheet, use the fields of the other sheet or the arithmetic mean of the two limiting fields. Substituting the total interior field without removing the self contribution would double the result.

At $E_0 = 2.00$ MV/m and $\beta = 0.60$, the primed interior field is 2.50 MV/m and its magnetic field 5.00 mT. Per positive surface charge the transverse attraction from the opposite sheet has magnitude $\gamma E_0(1-\beta^2)/2 = E_0/(2\gamma)$. This is 0.800 MV/m, compared with 1.00 MV/m in the laboratory. The transformation factor is inverse γ because the charge in this example is at rest in the original laboratory; the earlier force derivation placed the test charge at rest in the primed frame. Interchanging those initial conditions without changing the time conversion reverses the factor.

Multiplying the primed force per charge by the larger surface density $\gamma\sigma$ restores the same force per transverse plate area in this tangential boost. The cancellation supplies a local cross-check among length contraction, field transformation and mechanical force. A complete calculation of the total energy or momentum of an accelerated finite capacitor must also account for stresses in its supports and for the fields near its edges. Taking the uniform central field times a transformed volume is adequate for a local comparison, but does not by itself describe that whole closed mechanical system.

Test a field transformation with an invariant

Change observer speed, compare the separately transformed fields and inspect the invariant combination. Agree on the event before comparing measurements.

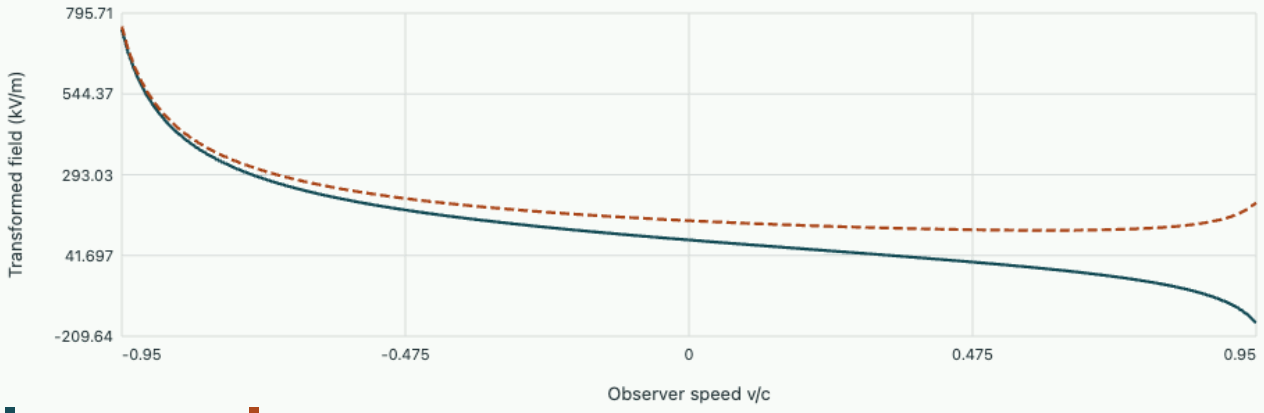
Open this experiment online

Try the calculation

Electric field E (kV/m) **90**

Magnetic field cB (kV/m) **150**

Observer speed β **0.6**



At the selected speed E′ = 0 kV/m and cB′ = 120 kV/m. E′² − c²B′² = -14400 (kV/m)².

Worked observer problem · a null of one component does not imply a vanishing field

At one event a laboratory finds $\mathbf{E} = (30, 40, 0)$ kV/m and $c\mathbf{B} = (0, 0, 50)$ kV/m. Another laboratory reports that it can remove the magnetic field by boosting along the direction of $\mathbf{E} \times \mathbf{B}$. Test the claim and predict the fields for a boost of speed $0.80c$ along this direction. Does the local measurement establish that this is an infinite plane wave?

▼ **Evaluate the invariants and transform the transverse field**

The electric and scaled magnetic fields both have magnitude 50 kV/m and are perpendicular. Both quadratic Lorentz invariants vanish, while the fields themselves are nonzero. With $\mathbf{v}$ along $\mathbf{E} \times \mathbf{B}$, both field vectors are perpendicular to the boost. The cross terms reduce their magnitudes by a common factor $\gamma(1 - \beta) = 1/3$ for this speed, so $E' = cB' = 16.7$ kV/m. Their directions in this transverse configuration are unchanged. No subluminal boost removes either nonzero field, as follows also by applying the inverse transformation to an alleged zero field.

A pulse at one point can have these local field values, as can the field near a suitable superposition at a particular event. A local tensor and its invariants contain no information about all the field derivatives elsewhere. To establish a plane-wave model measure spatial phase across transverse probes, verify the propagation relation at separated stations and bound amplitude variation over the required observation region. Infinity is an idealisation established only to a scale specified by the experiment.

As an error check, compare the measured difference $E^2 - c^2B^2$ with the propagated uncertainties of both meters. Small nonzero estimated differences near a null may be statistically insignificant, while a systematic alignment error in $\mathbf{E} \cdot \mathbf{B}$ can obscure orthogonality. Classifying a field as electric-dominated or magnetic-dominated from a difference of nearly equal large squares requires an uncertainty budget. The exact mathematical classification and the experimental confidence in assigning it answer different questions.

Recover a source direction from two clocks and a known speed

Two perfectly synchronised receivers travel with equal speed in opposite directions along the laboratory axis. They sample the same monochromatic plane wave. Their measured frequencies are proportional to $1 - \beta\cos\theta$ and $1 + \beta\cos\theta$, with a common Lorentz factor. Taking their difference divided by their sum determines $\beta\cos\theta$ without prior knowledge of emitted frequency. Their sum then fixes that frequency once the speed is calibrated. The construction combines two observer assignments rather than treating a transverse Doppler correction as an unknown empirical shift.

For speeds of half the speed of light and a measured frequency ratio of three to two, with the receiver travelling toward negative x recording the higher value, the inferred cosine is 0.40. If their average measured frequency is 600 THz, the emitted laboratory frequency is about 520 THz. A second direction on the opposite side of the boost axis has the same cosine, so the two receivers alone cannot determine the full incoming unit vector.

A third receiver with a nonparallel velocity supplies another directional projection. Detector clocks must be calibrated in their respective rest frames and each must see the same stable source frequency. If the source itself chirps between noncoincident measurements, that chirp can imitate an angular shift. Using overlapping signals and comparing phase histories addresses this limitation more directly than quoting only two peak frequencies.

Tutorial · Infer charge and current from two observers

A uniform channel carries an unknown lab line charge λ and current I along its axis. Observers moving at $+0.30c$ and $-0.30c$ measure line densities 1.0 and 3.0 nC/m respectively. Recover the laboratory quantities. Can a third observer travel alongside this current so that the measured current vanishes?

▼ Recover the two independent quantities and classify the current

Adding the readings gives $2\gamma\lambda = 4.0$ nC/m, so $\lambda = 1.908$ nC/m. Subtraction gives $2\gamma\beta I/c = 2.0$ nC/m and hence $I = 0.953$ A. Reinsert both signs in the original equations as a check. The speed needed to cancel the current would be $I/\lambda = 5.00 \times 10^8$ m/s. It exceeds the speed of light. The current four-vector is spacelike since $I^2 > (c\lambda)^2$.

A frame with zero line charge can instead move at $v = c^2\lambda/I = 0.600c$. A spacelike total current is possible for superposed charge carriers of opposite sign. The calculation would rule out a model

consisting solely of one sign of carriers all propagating with a common subluminal drift velocity. It does not rule out the measured pair of densities.

Tutorial · What does a moving pulse detector count?

A 10.0 ns rectangular-envelope light pulse moves in $+x$ in vacuum. A detector moves along the same direction at $0.60c$. Compute the interval between its encounters with the leading and trailing edges using laboratory kinematics, then its own clock. Predict its measured peak intensity and fluence relative to a stationary detector for a broad collimated beam.

▼ **Track the two crossing events and transform the flux**

The pulse is $c(10.0\,\mathrm{ns})$ long in the lab. Its back approaches the moving detector at coordinate closing speed $c - v = 0.40c$, so the two events occur 25.0 ns apart in laboratory time. Both events lie on the detector worldline, so its proper interval is 20.0 ns. This agrees with the inverse Doppler factor of two. A pulse of enough cycles to define a carrier now has half the frequency and twice its envelope duration.

For a transverse plane wave the field amplitude transforms as $\gamma(1 - \beta) = 0.50$ times the laboratory value. The primed Poynting flux is therefore one quarter as large. Integrating over twice the proper duration gives one half the original energy per unit transverse area. The cross section perpendicular to the boost is unchanged. Photons are registered with half their laboratory energy and the same total number if the beam is entirely collected, an independent check on the fluence result.

Cambridge Electromagnetism · electromagnetism and relativity

Waves, radiation and relativity · 11. Relativistic particle dynamics and the field tensor

A circular track measures momentum per charge. A clock following the particle registers proper time, and a detector records how much kinetic energy it has gained. We will put these observations in one dynamical equation, solve that equation in uniform fields and exhibit a numerical pusher that can get the energy right while its orbit phase slowly slips.

What the track, clock and calorimeter measure

Consider a particle of invariant mass m, signed charge q and ordinary velocity $\mathbf{v}$. Inertial observers agree on the proper time between neighbouring events on its timelike worldline,

$$d\tau = dt\sqrt{1 - v^2/c^2}.$$

Adopt coordinates $x^\mu = (ct, x, y, z)$ and metric $g = \mathrm{diag}(1, -1, -1, -1)$. Dividing the displacement by proper time produces a vector of fixed Minkowski norm. Multiplying it by invariant mass gives four-momentum.

$$u^\mu = \gamma(c, \mathbf{v}),$$
$$p^\mu = m u^\mu = (\mathcal{E}/c, \mathbf{p}),$$
$$\mathcal{E}^2 - c^2|\mathbf{p}|^2 = m^2 c^4.$$

Energy includes the rest term. Momentum is $\gamma m \mathbf{v}$. Differentiating the last equation on a trajectory gives

$$d\mathcal{E}/dt = \mathbf{v} \cdot d\mathbf{p}/dt.$$

This identity replaces the nonrelativistic step of dotting $m\mathbf{a}$ with velocity. A magnetic field can rotate momentum with fixed magnitude and energy, whereas an electric field can change energy by doing work.

Suppose an electron begins at rest and gains 10.0 keV of kinetic energy across an electrostatic potential difference in vacuum. Its rest energy is 511 keV, giving $\gamma = 1.01957$ and $v/c = \sqrt{1 - 1/\gamma^2} = 0.1950$. The nonrelativistic estimate is 0.1978. In a modest energy measurement the discrepancy could lie below the experimental spread. For a precise orbit calculation the 1.96 percent change in relativistic cyclotron period is measurable independently of that discrepancy in speed.

A linear map that produces the Lorentz force

Seek a local equation linear in field and four-velocity and covariant under boosts. For the energy component to give $q\mathbf{E} \cdot \mathbf{v}$, and for the spatial components to give the electric force and the oriented cross product, construct the antisymmetric field tensor. With the stated metric and $\epsilon^{123} = 1$, set

$$F^{0i} = -E_i/c,$$

$$F^{i0} = E_i/c,$$

$$F^{ij} = -\epsilon^{ijk} B_k,$$

$$\frac{dp^\mu}{d\tau} = qF^{\mu\nu} u_\nu.$$

Lowering the index on u reverses the signs of its spatial components. The spatial product becomes $q\gamma(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, and division by γ converts the proper-time derivative to $d\mathbf{p}/dt$. The time component similarly yields

$$d\mathcal{E}/dt = q\mathbf{E} \cdot \mathbf{v}.$$

Contract the force with u_μ. The symmetric product $u_\mu u_\nu$ annihilates the antisymmetric tensor, which proves that the evolution preserves the mass shell.

To recover Maxwell's kinematic constraints from potentials, put $A^\mu = (\phi/c, \mathbf{A})$ and $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. Then $F^{0i} = (\partial_t A_i + \partial_i \phi)/c = -E_i/c$, while the spatial-spatial components are the magnetic curl. Adding a four-gradient to A^μ leaves the tensor unchanged because ordinary derivatives commute. The particle action $S = -mc^2 \int d\tau - q \int A_\mu dx^\mu$ gives the same force after integrating variations by parts with fixed endpoints. Gauge changes add only an endpoint term to this action. A trajectory measured with local forces cannot distinguish such potential representatives.

Compare the potentials and field-based description in the field-tensor chapter

Constant force, variable coordinate acceleration

Put a particle initially at rest in a uniform static electric field along x. Let $a_0 = qE/m$ be signed. Integration of $dp_x/dt = qE$ gives $p_x = ma_0 t$, so the speed tends to c even while momentum increases linearly. Proper-time evolution is simplest in terms of rapidity η, where $u^0 = c\cosh\eta$, $u^x = c\sinh\eta$. The two tensor equations imply $d\eta/d\tau = a_0/c$.

$$t = \frac{c}{a_0}\sinh(a_0\tau/c),$$

$$x = \frac{c^2}{a_0}\{\cosh(a_0\tau/c) - 1\},$$

$$\frac{dv_x}{dt} = \frac{a_0}{\gamma^3}.$$

The initial time and position are zero. The work $qEx = mc^2(\cosh\eta - 1)$ equals kinetic energy, which checks both the integration constant and the dimensional factor in the displacement. For an electron use the direction of the electric force as positive to simplify signs. At 20.0 kV/m it gains 1.0 keV in 5.0 cm. The energy gain follows from the static potential even though the speed does not grow exactly as

the square root of distance. The field region must last long enough for the particle to cross it and fringe fields can alter the entry trajectory.

Compare momentum, coordinate time and proper time in the charged-worldline experiment

Helical motion, radius and period

A uniform magnetic field along z keeps p_z, energy and $|\mathbf{p}_\perp|$ fixed. Write $p_x + ip_y$ as a complex number. The Lorentz force reduces to a first-order rotation equation and a single time integration gives the corresponding position. With positive B, positive charges starting along $+x$ initially turn towards $-y$.

$$\frac{d}{dt}(p_x + ip_y) = -i\frac{qB}{\gamma m}(p_x + ip_y),$$

$$R = \frac{p_\perp}{|q|B},$$

$$T = \frac{2\pi\gamma m}{|q|B}.$$

The helical pitch is $v_z T$. Radius gives transverse momentum and a calibrated period gives total energy through γ. A track seen only in projection along the field gives no direct evidence that longitudinal momentum vanishes. A second view or a timing measurement can resolve this ambiguity.

At 10.0 keV the electron has $pc = \sqrt{K(K + 2mc^2)} = 101.6$ keV. With entry perpendicular to a 2.0 mT field its radius is 0.169 m and its period 18.2 ns. Doubling the field halves both radius and period at fixed injection energy. Doubling momentum at fixed field doubles radius but changes period by the ratio of Lorentz factors. These are different tests of a track calibration. A beam with a spread of energies will dephase relative to a narrow timing gate even when every single particle remains in a perfectly uniform field.

A drift frame gives an exact crossed-field solution

For constant crossed fields with $E < cB$, transform to the frame moving at $\mathbf{v}_d = \mathbf{E} \times \mathbf{B}/B^2$. There $E' = 0$ and $B' = \sqrt{B^2 - E^2/c^2}$; calculate the transformed initial momentum and integrate a helix. Transform two successive equal-phase events back to the lab. Their perpendicular displacement in the electric-free frame is zero. If the motion has no component parallel to $\mathbf{B}$, the entire displacement there is zero, and the average lab cycle velocity is exactly $\mathbf{v}_d$. The detailed lab trace need not be circular and the instantaneous lab energy may oscillate.

Take $E = 30.0$ kV/m along y and $B = 0.50$ mT along z. The drift is along $+x$ at 6.00×10^7 m/s. A particle injected at that velocity crosses the uniform fields on a straight line because the forces cancel. An initially stationary positive particle instead begins by accelerating along $+y$. These observations can use the same field with different injected particles. If E exceeds cB, this electric-free frame is unavailable. Transform to a magnetic-free frame and integrate electric acceleration instead.

Continuous rotation and a discrete experiment

A naive forward step $\mathbf{p}_{n+1} = \mathbf{p}_n + q\Delta t\, \mathbf{v}_n \times \mathbf{B}$ adds a vector perpendicular to the previous momentum. Its squared norm increases by the squared magnitude of that kick. The apparent heating will persist in perfectly static uniform magnetic fields. For a purely magnetic update, evaluate the direction using the mean of old and new momenta. At fixed γ, the resulting implicit two-dimensional linear map can be solved algebraically and is an exact rotation by a modified angle.

$$\delta = 2\arctan\left(\frac{\omega_c \Delta t}{2}\right),$$

$$\omega_c = \frac{qB}{\gamma m}.$$

The sign of rotation follows the earlier complex equation. Dotting the update with the mean momentum gives zero change in squared norm, which verifies the energy property even for relatively large time steps. A common electromagnetic pusher surrounds the magnetic rotation by two electric half-kicks. In spatially varying fields, specify at what time and position each field is sampled, and check spatial interpolation and time convergence. For ultrarelativistic nearly cancelling electric and magnetic fields, preserving rotation in a pure field alone does not ensure cancellation is represented accurately.

Solve the constant tensor without transforming every time step

When uniform electric and magnetic fields are constant in spacetime, the Lorentz equation is a linear constant-coefficient equation for four-velocity as a function of proper time. Nonlinearity in ordinary time hides that simple structure, because the rate at which proper time accumulates changes along the orbit. Define the mixed tensor $M^{\mu}{}_{\nu} = (q/m)F^{\mu\alpha}g_{\alpha\nu}$. In a uniform field it does not vary along the trajectory. A formal solution is then

$$u(\tau) = e^{M\tau}u(0),$$

$$x(\tau) - x(0) = \int_0^{\tau} e^{Ms}u(0)\, ds.$$

The coordinate formula is written as an integral since M can be singular, as it is in a purely magnetic field. Blindly replacing this integral by $M^{-1}(e^{M\tau} - I)$ fails in exactly such simple cases. A power series for the integral or a block matrix exponential treats zero eigenvalues without an artificial inverse. Antisymmetry of the tensor implies $M^{\mathsf{T}}g + gM = 0$. Differentiating $e^{M^{\mathsf{T}}\tau}ge^{M\tau}$ shows that it stays equal to g, making preservation of $u^{\mu}u_{\mu} = c^2$ a structural property of the exact mapping.

For parallel constant fields along z, the matrix decomposes into a boost in the ct, z subspace and a rotation in x, y. Write $\alpha = qE/(mc)$, $\Omega = qB/m$, and denote the initial four-velocity components by u_0^{μ}. The proper-time solution is

$$u^0 = u_0^0 \cosh \alpha\tau + u_0^z \sinh \alpha\tau,$$
$$u^z = u_0^z \cosh \alpha\tau + u_0^0 \sinh \alpha\tau,$$
$$u^x + i u^y = (u_0^x + i u_0^y) e^{-i\Omega\tau}.$$

The longitudinal and transverse operations commute because they act in different subspaces. In lab time the transverse rotation slows as the electric field increases energy, although the proper-time rotation frequency stays constant. A particle therefore traces a helix whose transverse radius $p_\perp/(|q|B)$ is fixed while the pitch evolves. A single measured turn time without a tagged position in the accelerating region is insufficient for an energy reconstruction.

As a numerical test use an electron, quoting magnitudes, with $B = 1.0$ mT, $E = 10.0$ kV/m and initial purely transverse speed $0.10c$. One magnetic turn takes 35.72 ns in proper time and accumulates electric rapidity magnitude $|\alpha|\tau \simeq 0.2096$. The final Lorentz factor is $\gamma_i \cosh(0.2096) = 1.0272$, giving final kinetic energy about 13.9 keV, compared with 2.57 keV initially. The lab-time turn duration is $\gamma_i \sinh(|\alpha|\tau)/|\alpha| \simeq 36.2$ ns. Reversing the electric field changes the direction of the axial displacement but leaves final energy and turn duration unchanged for this initially transverse particle. Reversing the initial longitudinal velocity would generally break that energy symmetry.

Infer uncertainty and test units in a momentum reconstruction

A track fitter commonly reports field, radius, pitch, time and a fitted covariance matrix. The dynamical relation turns these observed quantities into momentum. Let L_z denote a signed pitch per turn. With known signed direction of circulation and calibrated absolute field magnitude,

$$|\mathbf{p}| = |q| B \sqrt{R^2 + (L_z/2\pi)^2}.$$

For small errors, logarithmic differentiation gives weights for fractional radius and pitch errors:

$$d\log p = d\log B + w_R\, d\log R + w_L\, d\log |L_z|,$$
$$w_R = \frac{R^2}{R^2 + (L_z/2\pi)^2},$$
$$w_L = 1 - w_R.$$

For $R = 0.20$ m and $L_z = 0.30$ m the radius accounts for 94.6 percent of the squared transverse-plus-longitudinal geometric factor. Independent relative uncertainties of 0.20 percent in field, 0.50 percent in radius and 2.0 percent in pitch yield about 0.525 percent uncertainty in total momentum. A common detector scale controlling both radius and pitch would enter with correlated signs and should not be counted as two independent errors. When the pitch tends to zero, use the original derivatives with respect to L_z; a reported fractional error on a quantity consistent with zero is ill defined.

Differentiating $K = \sqrt{m^2c^4 + p^2c^2} - mc^2$ gives

$$d\log K = (\gamma + 1) d\log p / \gamma.$$

The coefficient approaches two at low energy and one at very high energy. At the approximately 3.70 keV energy of the tutorial the above momentum error would imply about a 1.05 percent energy error. A kinetic-energy estimate from a small difference of nearly equal numbers can also lose numerical precision at much lower energies. Rationalise it as $K = p^2c^2/(\sqrt{m^2c^4 + p^2c^2} + mc^2)$ when working near rest.

Several independent checks use this reconstruction without assuming the track fitter is correct. The turn time inferred from the fitted γ can be compared with a separate timing channel. Reverse the magnetic field to reverse the direction of curvature while keeping the speed and energy approximately fixed. Translate the beam within a mapped uniform field and look for changing fitted radius that correlates with position, an indication of geometric or field-map bias. Finally plot the residual of the magnetic force law separately for transverse magnitude and angle. A time integrator may conserve momentum magnitude for symmetry reasons even when its field interpolation biases the phase or shifts the guiding centre systematically.

Advance a charge using relativistic momentum

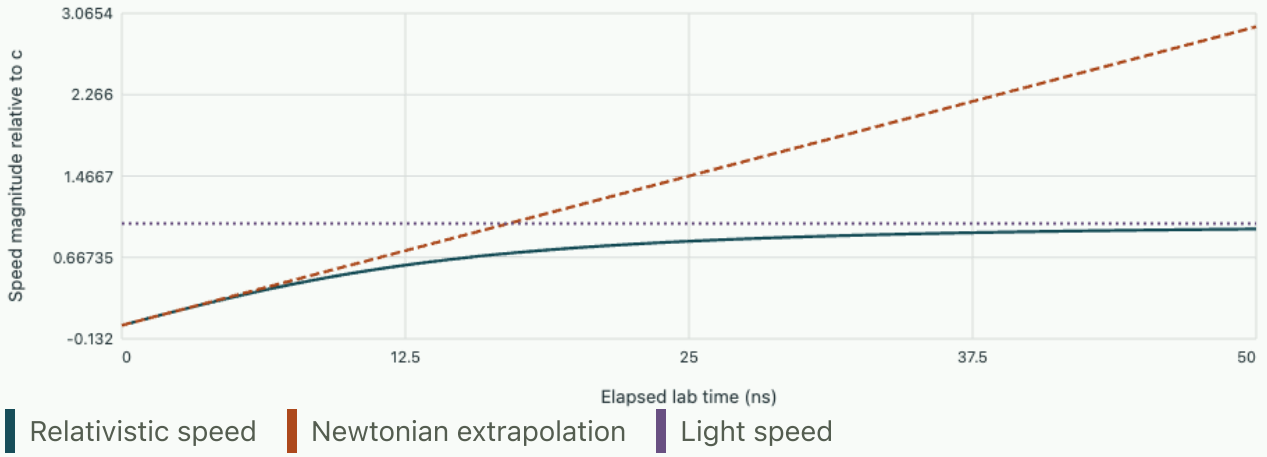
Increase the impulse while following speed, energy and proper time. Identify the range where a Newtonian momentum update becomes inaccurate.

Open this experiment online

Try the calculation

Electric field magnitude (kV/m) **100**

Lab time for a reading (ns) **17**



At 17 ns laboratory time the electron speed is 0.706 c. Its clock has advanced 15 ns and it has gained 211 keV.

Worked timing problem · infer an accelerating voltage without a radius

A tagged electron bunch completes ten nearly circular turns in a calibrated 5.0 mT magnetic field. Two time stamps indicate an elapsed 73.0 ns. The field is perpendicular to the orbit. Assuming complete turns and negligible electrical work during the timed interval, estimate the injection energy. What timing precision would distinguish a one-percent change in that energy?

▼ **Use the relativistic period sensitivity rather than the energy fraction directly**

The nonrelativistic turn period is

$$2\pi m/(eB) = 7.145\,\mathrm{ns}.$$

The measured period is 7.30 ns, giving $\gamma \simeq 1.0217$ and $K = (\gamma - 1)mc^2 \simeq 11.1$ keV. The relative energy sensitivity is much larger than the relative period change when γ is close to one. A one-percent change in K corresponds to $\Delta\gamma \simeq 2.17 \times 10^{-4}$, or relative period change 2.12×10^{-4}. Over the ten turns this gives approximately 15.5 ps.

To resolve that change spectroscopically, uncertainties in the elapsed time and field scale must together be significantly smaller than the equivalent change. A constant pickup cable delay cancels if the same electronics register both ends of the interval, while a drift of its timebase scale does not. Timing over more turns improves the total accumulated signal but also increases exposure to field inhomogeneity, energy loss and dephasing among particles with different injection energies. A narrower timing histogram does not automatically eliminate a systematic field calibration.

Finally verify the integer turn count independently. Missing a pickup crossing would cause a much larger change in inferred energy than the desired resolution. The measured wave packet may broaden so strongly that later crossings overlap. Recording the full pickup waveform and tracking ordered crossings, along with an independent rough radius estimate, can constrain the turn assignment before fitting the small relativistic shift. A perfect circular image with no time labels cannot supply that missing count.

Why momentum and kinetic energy do not gain fixed fractions together

In an electrostatic accelerating gap, the kinetic energy increment for a particle passing between two equipotential surfaces is its charge times the potential drop, regardless of injection energy. This statement follows by integrating the Lorentz power law over the actual path. A magnetic guiding field can alter the path and transit time while doing no direct work. A time-dependent electric field, a missed equipotential endpoint or appreciable radiation changes the simple inference and must be addressed separately.

For an electron entering with 10 keV and acquiring an additional 10 keV, kinetic energy doubles. Its initial momentum times c is about 101.6 keV and its final value about 144.4 keV. The radius in the same subsequent perpendicular magnetic field increases by a factor approximately 1.42. Doubling the radius would instead require doubling momentum, which demands a substantially greater energy increment at this low initial energy. Comparing radii before and after a known voltage supplies an independent check on the sign and magnitude of a particle's charge-to-mass model.

A measured energy gain below the nominal voltage does not by itself diagnose radiation. The potential actually traversed, finite angular acceptance, entrance or exit fringe fields, collision losses and biases in reconstructing longitudinal momentum can all matter. Repeat with reversed field polarity or several injection energies and compare the kinetic-energy difference rather than the two radii directly.

Tutorial · Reconstruct a track with two measurements

A particle known to be an electron or positron crosses a uniform $+z$ field of 1.0 mT. The projected radius is 0.200 m and the pitch magnitude is 0.300 m. Recover the transverse and longitudinal momentum, energy and period. What else must the detector record to determine the sign of charge?

▼ **Reconstruct both momentum components**

The transverse momentum is $|q|BR = 3.20 \times 10^{-23}$ kg m/s. During a cycle the parallel displacement is $p_z T/(\gamma m) = 2\pi p_z/(|q|B)$, so $|p_z| = 7.65 \times 10^{-24}$ kg m/s. The quadrature total is 3.29×10^{-23} kg m/s, yielding $p/(mc) = 0.1206$, $\gamma = 1.00724$, and approximately 3.70 keV kinetic energy. The period is about 36.0 ns.

Neither radius nor pitch magnitude carries the signed circulation. Record the direction of traversal relative to the calibrated direction of magnetic field. Reversing both charge and the direction of traversal yields the same unlabelled transverse ring. If timing or two ordered detectors supply a direction, the sign follows from the first Lorentz kick.

Tutorial · Can a constant-energy circle be trusted?

A pusher uses 20 uniform steps per exact cyclotron period in a uniform magnetic field. It reproduces initial kinetic energy to machine precision for 50 turns. Estimate the phase drift from the midpoint rotation formula. Design a refinement diagnostic that distinguishes a time-discretisation bias from a field calibration error.

▼ **Compare the exact and discrete angles**

The exact unsigned angle per step is

$$h = 2\pi/20 = 0.31416.$$

The discrete one is

$$2\arctan(h/2) = 0.31161.$$

The accumulated deficit over 1,000 steps is about 2.55 rad, roughly 0.41 turn. The method can keep every track point on a closed momentum circle yet miss the pickup electrode by a large fraction of a revolution. Coordinate radius may depend on how positions were staggered, so perform the angular comparison on momentum first.

Double the number of steps at fixed physical end time and take the difference in unwrapped phases. The leading angular error $h^3/12$ per step becomes a quadratic global error, so halving the step should reduce it by about four. A calibration error in $qB/(\gamma m)$ persists under time refinement. Track both the expected frequency and the convergence of the time series; comparing the coordinates at a single late time can confuse complete turns with near agreement.

Cambridge Electromagnetism · four-vectors, fields and charged matter

Waves, radiation and relativity · 12. Radiated power, mechanical work and energy limits

A charge can radiate while its speed is constant. A source can exchange energy with its nearby field before a pulse reaches a receiver. We will use the flux through a distant sphere, the work at the charge and the energy stored in a local field as three operationally distinct measurements. Agreement requires the relevant integration times and the range of validity of the particle model.

Integrate the far field of one slowly moving charge

For a source moving slowly compared with c, the radiation field at a large distance is transverse to the viewing direction $\mathbf{n}$. To leading order in v/c, its electric amplitude is $q\mathbf{n}\times(\mathbf{n}\times\mathbf{a})/(4\pi\epsilon_0 c^2 r)$ evaluated at retarded time. The magnetic radiation field is $\mathbf{n}\times\mathbf{E}/c$. Form the radial Poynting flux and integrate over the sphere. Choose the polar axis along the instantaneous acceleration.

$$\frac{dP}{d\Omega}=\frac{q^2a^2}{16\pi^2\epsilon_0 c^3}\sin^2\theta,$$

$$\int \sin^2\theta\, d\Omega=\frac{8\pi}{3},$$

$$P=\frac{q^2a^2}{6\pi\epsilon_0 c^3}.$$

The angular integral contains a factor of $\sin\theta$ from solid angle in addition to the square from transverse projection. The result is instantaneous emission power, not the average over a sinusoid. For acceleration $a_0\cos\omega t$, the average is one half the power computed with peak acceleration a_0. No acceleration component along the velocity is required. A charge in circular motion radiates even when the externally applied magnetic field does no work.

A trapped electron executing small prescribed oscillations at 100 MHz with amplitude 1.0 micrometre has peak speed 628 m/s and peak acceleration 3.95×10^{11} m/s squared. Its mean radiated power is 4.45×10^{-31} W. The energy scale of this illustrative classical orbit is only 1.80×10^{-25} J, much smaller than one oscillator quantum at 100 MHz. The field calculation is meaningful as a classical scaling or for an appropriately prepared mean current. Interpreting it as a literal well-defined single-electron mechanical orbit would require addressing the state preparation. Raising the amplitude to 1.0 mm raises energy and mean emitted power by a factor of a million, and keeps v/c below 0.003.

Integrate the measured dipole pattern and compare with the total radiated power

Transform instantaneous power using the rest frame

In the instantaneously comoving frame the nonrelativistic expression applies at that event, with the proper acceleration. The dipolar pattern there carries zero net three-momentum when integrated over all directions, by front–back symmetry. The radiation four-momentum emitted during $d\tau$ is therefore $(P_0 d\tau/c,\mathbf{0})$. A boost multiplies its energy by γ, but also multiplies the emission-event time interval by

γ. Thus the total laboratory power allocated per emitter coordinate time equals P_0. The directional detector signal need not share this invariance.

At the instant of rest the proper acceleration is the Euclidean norm of the spatial part of $du^\mu/d\tau$. Its square in any frame follows from the invariant $-a^\mu a_\mu$. Differentiating $u^\mu = \gamma(c, \mathbf{v})$, with $d\gamma/dt = \gamma^3 \mathbf{v}\cdot\mathbf{a}/c^2$, and separating acceleration parallel and perpendicular to velocity gives

$$P = \frac{q^2}{6\pi\epsilon_0 c^3}\left(\gamma^6 a_\parallel^2 + \gamma^4 a_\perp^2\right) = \frac{q^2\gamma^6}{6\pi\epsilon_0 c^3}\left[a^2 - |\boldsymbol{\beta}\times\mathbf{a}|^2\right].$$

For a fixed applied parallel electric force,

$$a_\parallel = qE/(\gamma^3 m).$$

The γ^6 cancels, so the ideal point-charge radiated power in a constant longitudinal field is independent of instantaneous speed. This conclusion keeps the field fixed. Holding coordinate acceleration fixed while increasing speed would require an increasing input force, producing a different power law. Always identify what is actually kept fixed before applying the exponents.

At fixed perpendicular magnetic field, $a_\perp = |q|vB/(\gamma m)$, so power varies as $\gamma^2 v^2 B^2$. A gyrating electron gradually loses energy if there is no replenishing electric field. Over a short tracked interval one can treat its radius and frequency as constant only after comparing radiated energy with kinetic energy. The large-mass dependence makes the corresponding correction much smaller for heavy particles at comparable speeds and charges.

Arrival time and a detector's angular window

Let radiation from emission position $\mathbf{r}_s(t_e)$ travel to a fixed distant receiver. Its arrival time is

$$t_o = t_e + |\mathbf{R} - \mathbf{r}_s(t_e)|/c.$$

Differentiating this relation gives

$$dt_o/dt_e = 1 - \mathbf{n}\cdot\boldsymbol{\beta}.$$

A forward-moving charge can deliver the energy emitted over a long source time in a much shorter arrival interval. Angular radiation patterns specified per emission time and a directional Poynting flux specified per arrival time then differ by this Jacobian.

Explore how observation direction and motion change the received field

At $\beta = 0.90$ directed toward a distant receiver, radiation from two source events separated by 5.0 ns arrives about 0.50 ns apart if the direction remains nearly unchanged. The same events viewed transversely arrive 5.0 ns apart. A slow photodiode could register comparable integrated pulse energies after geometric corrections while reporting very different peak shapes from the true fields. If the

acceleration also changes direction during the interval, account for the angular antenna pattern before attributing the whole difference to arrival-time compression.

A finite-time mechanical energy balance

Poynting's theorem balances electromagnetic energy, its outward flux and work on charges. In a small region surrounding an idealised particle one can separate a far-field flux associated with radiation from energy temporarily stored with the near field. In the nonrelativistic, weak-reaction expansion the leading self-force is $m\tau_0\dot{\mathbf{a}}$, where $\tau_0 = q^2/(6\pi\epsilon_0 mc^3)$. Dotting with velocity and applying the product rule gives

$$\mathbf{F}_{\mathrm{self}}\cdot\mathbf{v} = \frac{d}{dt}(m\tau_0\mathbf{a}\cdot\mathbf{v}) - m\tau_0 a^2,$$

$$P_{\mathrm{rad}} = m\tau_0 a^2.$$

Adding the external mechanical work shows that $W_{\mathrm{ext}} = \Delta K + E_{\mathrm{rad}} - [m\tau_0\mathbf{a}\cdot\mathbf{v}]_{t_1}^{t_2}$. This last boundary term accounts for reversible energy exchange with the bound field in this separation. It vanishes if acceleration is zero at both endpoints, or over complete periods when the mechanical state repeats. During a limited interval positive radiated energy and instantaneous mechanical damping need not coincide.

A prescribed interval of constant acceleration illustrates the distinction. Within that interval $\dot{\mathbf{a}} = 0$ and the local leading self-force vanishes, yet the far-zone power is positive. The endpoint term changes because $\mathbf{a}\cdot\mathbf{v}$ grows with time. A physical pulse has switching intervals and any comparison of initial and final energies must include them. Treating an endlessly accelerated point charge as a mechanical system with ordinary finite reservoir energies introduces boundary questions absent in a compact laboratory pulse.

Compare emitted energy, mechanical work and the endpoint contribution in a pulse

When a weak-reaction or classical model stops answering the question

The electron time τ_0 is approximately 6.27×10^{-24} s. For a driving angular frequency ω the ratio $\tau_0\omega$ estimates the size of the leading reaction force relative to inertial force in a harmonic nonrelativistic problem. At 100 MHz it is 3.94×10^{-15}, tiny per cycle. Many cycles can sum to an appreciable energy loss when all other damping mechanisms are very weak. For a nonrelativistic harmonic oscillator the cycle-averaged Larmor loss equals $\tau_0\omega^2$ times its stored mechanical energy, giving an energy lifetime $1/(\tau_0\omega^2)$.

A perturbative approximation replaces acceleration inside the small self-force by the acceleration from the external force to leading order. For a prescribed, smooth, nonrelativistic $F_{\mathrm{ext}}(t)$, this yields

$$ma \simeq F_{\mathrm{ext}} + \tau_0\dot{F}_{\mathrm{ext}}.$$

Errors scale with further time derivatives. Treating the unexpanded third-order point-particle equation as an arbitrary initial-value solver admits a homogeneous exponentially growing acceleration. That mode varies on a timescale the weak-reaction expansion did not resolve in the first place.

For a collective source also check dimensions against wavelength. A cloud much smaller than the wavelength can add fields coherently when all charges have the same driven phase. Then total power can scale as the squared number of radiators, while random phases usually leave only a contribution proportional to their number after averaging. At larger dimensions include the spatial Fourier transform of the source. Finally, photon recoil and quantisation can matter even when a classical average flux describes an ensemble. Compare $\hbar\omega$ with the mechanical energy change that the measurement actually resolves.

Recover a scattering cross section from an energy balance

A free nonrelativistic electron in a weak nearly monochromatic incident wave is accelerated by the electric field. Let its peak incident field be E_0, assume a quiver speed small compared with c, and ignore the leading magnetic force in calculating the quiver. Then $a_0 = |q|E_0/m$. The time-averaged Larmor power divided by incident time-averaged intensity $\epsilon_0 c E_0^2/2$ gives a quantity with the dimensions of area. The amplitude cancels:

$$\sigma_T = \frac{q^4}{6\pi\epsilon_0^2 m^2 c^4} = \frac{8\pi}{3} r_e^2,$$

$$r_e = \frac{q^2}{4\pi\epsilon_0 m c^2}.$$

For an electron $r_e = 2.82$ femtometres and $\sigma_T = 6.65 \times 10^{-29}$ square metres. The radius here is a combination of charge, mass and speed of light, not an independently measured mechanical size of the electron. For an incident intensity of 100 watts per square metre the weak-drive prediction is 6.65×10^{-27} W per free electron. Calculating an expected photon count requires wavelength, collection solid angle and detector efficiency as well as that power.

There is a subtle question about work done by the incident wave. At lowest order the acceleration is exactly proportional to the applied field and the velocity lags it by a quarter cycle, giving zero average $q\mathbf{E}\cdot\mathbf{v}$. Yet the radiation formula predicts positive time-averaged outgoing energy. The resolution is an order count. Radiation power is already of the order of the small reaction correction, and an input-work calculation to that same order must include the correction to velocity that is in phase with the drive. The perturbatively reduced self-force introduces that component. Equivalently, far from the particle the incident–scattered interference contributes extinction equal to scattering in the absence of absorption.

For a charge bound by a spring, the quiver changes near and below its resonance. Well below resonance the displacement is nearly independent of driving frequency, so radiation power falls with the fourth power of frequency at fixed field. Far above resonance, but well within the classical low-recoil regime, the inertial free-electron response is recovered. This explains why applying the Thomson cross section indiscriminately to all the electrons in a transparent dielectric can give a misleading

attenuation. Bound response, correlations, interfaces and coherent forward propagation can dominate a bulk optical measurement.

Count coherent radiators through a spatial form factor

Consider N nominally identical dipoles driven at positions $\mathbf{r}_j$ by a plane wave with wavevector $\mathbf{k}_i$. Their induced moments acquire incident phases $e^{i\mathbf{k}_i\cdot\mathbf{r}_j}$. Observation in direction $\mathbf{n}$ supplies the outgoing source factor $e^{-ik\mathbf{n}\cdot\mathbf{r}_j}$, so define the transfer wavevector $\mathbf{Q}=\mathbf{k}_i-k\mathbf{n}$. With equal amplitudes and a fixed observed polarisation, the field relative to one dipole is $\sum_j e^{i\mathbf{Q}\cdot\mathbf{r}_j}$. Squaring gives N self terms and ordered cross terms. If positions are independently drawn from a normalised spatial distribution $w(\mathbf{r})$,

$$\left\langle \left| \sum_j e^{i\mathbf{Q}\cdot\mathbf{r}_j} \right|^2 \right\rangle = N + N(N-1)|\widehat{w}(\mathbf{Q})|^2.$$

At exact forward scattering $\mathbf{Q}=0$ and all phases add regardless of source extent. Away from the forward direction a broad cloud can lose mutual coherence even though its drive remains perfectly monochromatic. The total power integrated over a sphere need not follow the on-axis scaling, because the enhanced forward cone narrows as the cloud grows. At sufficiently large density the incident field on one dipole includes scattered fields from others, invalidating the independent-dipole derivation. The formula alone cannot predict arbitrarily large emitted power from an optically thick weakly fed sample.

For a Gaussian cloud with one-coordinate standard deviation s,

$$|\widehat{w}|^2 = e^{-s^2Q^2},$$
$$Q = 2k\sin(\theta/2).$$

Take wavelength 600 nm, $s=5.0$ micrometres and $N=100$. At one degree, $sQ=0.914$ and the averaged intensity factor is about 4.39×10^3 rather than either 100 or 10,000. At five degrees $sQ=4.57$; the coherent excess becomes negligible and the factor approaches 100. An aperture spanning both these angles requires an angular integral, not the central value multiplied by its nominal solid angle.

One way to measure this distinction uses two nominally identical clouds with different sizes but equal independently counted particles. Record angular intensity through a series of apertures, keeping polarisation and illumination profile fixed. A growing forward peak with a narrowing angular width points toward a coherent form factor. If broad-angle intensity changes too, check whether the local density modifies each dipole's response or the sample heats. Compare a phase-sensitive measurement of forward extinction with a calibrated total scattered power rather than treating the forward peak as a direct measure of lost beam energy.

An energy audit of the complete apparatus must include the drive. If radiation from N synchronised elements is substantially larger than the sum for isolated elements, maintaining specified oscillations can demand additional work from the source through mutual fields. If the source instead supplies fixed

power, the element amplitudes generally change. The external energy supply, field distribution and measurement solid angle are all part of the physical question. This accounting links a many-source optics measurement back to the finite-time mechanical work balance for an individual charge.

Relate an accelerating source to the energy it radiates

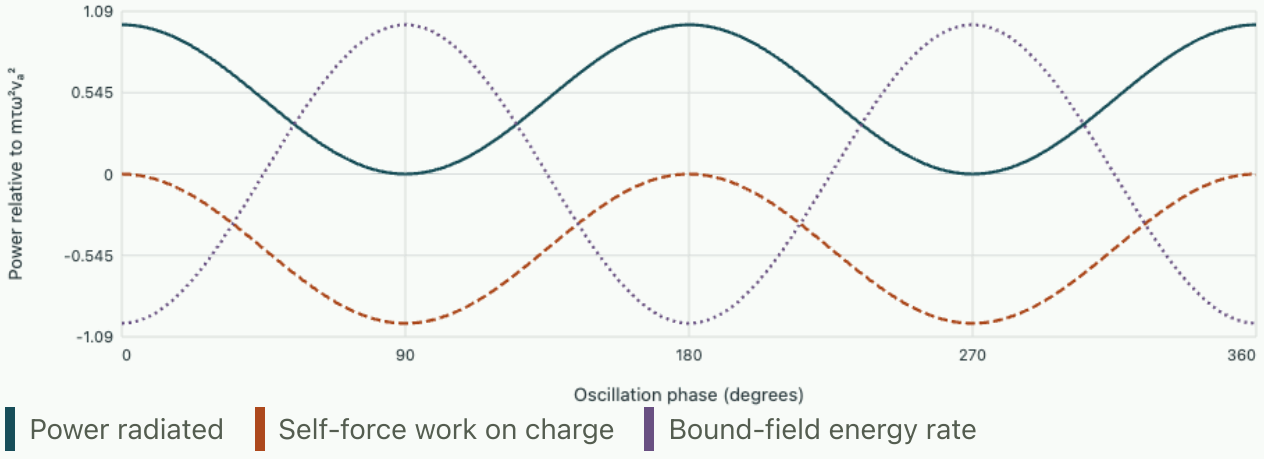
Change the time scale of a smooth motion. Compare instantaneous power with the integrated balance over the complete event.

Open this experiment online

Try the calculation

Mean velocity / oscillating speed **0**

Emission phase (degrees) **0**



At 0°, radiation, self-force work and bound-field rate are 1, 0, -1 in the plotted units. Their sum is zero.

Worked detector problem · energy in a cone around a dipole axis

A nonrelativistic charge acceleration pulse has known total radiated energy 1.00 pJ. An ideal detector collects the entire cone of half-angle 15° about the direction of acceleration, at a distance large compared with the source and the radiation wavelength. Estimate the recorded energy. Compare it with a detector of the same small solid angle centred perpendicular to the acceleration.

▼ **Integrate the projected electric field before applying the photon count**

For acceleration with fixed direction during the pulse, the radiated energy per unit solid angle is proportional to $\sin^2\theta$. The axial fraction is

$$(3/4)\int_0^{\alpha}\sin^3\theta\,d\theta=(3/4)(2/3-\cos\alpha+\cos^3\alpha/3).$$

At fifteen degrees this gives about 8.61×10^{-4} and a collected energy of 0.861 fJ. The small-angle approximation $3\alpha^4/16$ gives 0.881 fJ, consistent at this angle. A detector that assigns its on-axis zero intensity to the whole cone would incorrectly predict no signal.

At right angles, intensity varies slowly over a narrow aperture. Its leading fraction is $3\Omega/(8\pi)$, where $\Omega=2\pi(1-\cos15^\circ)$. This gives about 0.0256 or 25.6 fJ, with a small finite-angle correction. For an exact transverse cone average n_z^2 in coordinates about the detector axis. The ring average is $\sin^2\delta/2$ at polar offset δ, so the exact fraction becomes $(3/8)[(1-\cos\alpha)+(1-\cos^3\alpha)/3]$, or approximately 0.0251. Equal solid angle alone does not imply comparable collected energy.

These estimates assume unit efficiency across the entire pulse spectrum and a stable acceleration direction. A real detector with frequency-dependent response measures an integral over frequency and collection angle. Calibrate both before scaling its reading to total pulse energy. If the acceleration changes direction during the event, integrate the angular pattern appropriate to each emission time instead of assigning all the energy to a single measured peak direction.

A radiofrequency oscillator's finite lifetime creates a spectral width

An energy loss mechanism can be inferred from a time trace and independently tested in the frequency domain. Suppose an isolated weakly damped oscillator loses mechanical energy exponentially with time constant τ_E. Its position amplitude then falls as $e^{-t/(2\tau_E)}$. Near a well-resolved oscillation frequency its Fourier amplitude has a simple pole shifted below the real axis by $1/(2\tau_E)$. The squared Fourier amplitude is Lorentzian, with full width in ordinary frequency $1/(2\pi\tau_E)$.

If the measured energy lifetime in the earlier trapped-electron example were limited by its predicted radiation alone, the resonance width would be roughly four tenths of a microhertz. A measurement lasting one second cannot directly resolve such a line through the uncomplicated Fourier width of that record. Truncating a long-lived signal after that interval multiplies it by the acquisition window and convolves its spectrum with a much broader instrumental response. A broad fitted peak from the short record is consequently insufficient evidence of a short physical lifetime.

Compare records of several durations and perform a phase-coherent fit where appropriate. Technical drive-frequency jitter broadens an ensemble spectrum while a single undriven decay can remain long. Additional true energy dissipation shortens the ringdown itself. Pairing a spectral-width measurement with a time-domain energy measurement thus distinguishes loss of phase coherence from loss of oscillator energy, two effects that simple line shapes alone often confound.

Tutorial · Does the field energy explain a measured ringdown?

A nonrelativistic electron is driven at 100 MHz in a harmonic potential. After the drive ceases, the measured energy decays with time constant 0.10 s. Calculate the radiative energy lifetime, the energy fraction radiated in 0.10 s under radiation alone and whether a measured exponential proves radiative damping.

▼ **Construct and compare the decay rates**

Using $\omega = 2\pi 10^8$ rad/s and the electron τ_0, the radiation-only energy decay rate is 2.47×10^{-6} per second and its time constant is 4.04×10^5 s, about 4.68 days. The fraction radiated in 0.10 s is $1 - \exp(-2.47 \times 10^{-7}) \simeq 2.47 \times 10^{-7}$. The observed decay rate is about four million times larger. Under the assumptions of this single-particle free-space model it calls for further coupling or dissipation.

The amplitude decay time of an oscillator is twice its energy decay time. Confirm whether the measuring circuit reports amplitude, squared amplitude or extracted energy before fitting. A conducting environment can modify the density of radiation modes, and a measurement circuit can add a load. Compare separate runs with different coupling to that circuit and compare a calculated angular power with the collected signal. A single exponential identifies a rate, not a physical mechanism.

Tutorial · Extract a finite-pulse balance from an acceleration trace

During a short nonrelativistic interval the acceleration is $a(t) = a_0 t/T$, the velocity initially vanishes and motion is one-dimensional. Calculate the radiated energy and integrated self-force work over that interval. Then explain what additional information an experiment needs to compare final mechanical energy with radiation after the entire pulse.

▼ **Retain the endpoint before interpreting the sign**

Integrating the acceleration gives

$$v(t) = a_0 t^2/(2T).$$

The radiated energy is $m\tau_0 a_0^2 T/3$. At the final endpoint $m\tau_0 a v = m\tau_0 a_0^2 T/2$. Thus integrated self-force work is the boundary term minus radiated energy, $+m\tau_0 a_0^2 T/6$. A direct integration of $(m\tau_0 a_0/T)v(t)$ gives the same positive answer, despite an outward flux throughout the ramp.

At T the acceleration is still nonzero. A smooth continuation back to zero changes both the radiated energy and boundary contribution. To compare initial and final states without a bound-field endpoint term, measure through that continuation and wait for the pulse to reach the collecting surface. Specify angular coverage and collection efficiency if only a fraction of the radiation reaches the apparatus. An

on-axis detector along the acceleration could see no leading dipole signal even though the sphere integral is positive.

Cambridge Electromagnetism · accelerating charges and radiation

Waves, radiation and relativity · 13. How can a field that decays in air transmit power to another prism?

Place a thin air gap between identical prisms. Derive the TE transmission from the two interface conditions and test when an exponential attenuation length predicts the actual throughput.

Propagate both boundary data through the gap

A monochromatic TE wave in a prism of index n has tangential wavenumber $k_x = nk_0 \sin\theta$, normal wavenumber $k = nk_0 \cos\theta$, and $k_0 = 2\pi/\lambda$. Above the critical angle the air equation becomes

$$E'' = \kappa^2 E,$$

where $\kappa = (k_x^2 - k_0^2)^{1/2}$. Matching E and its derivative across each nonmagnetic interface retains both independent exponential solutions. Discarding the growing solution would preclude matching the outgoing wave in the second prism.

From the exit to the entrance the transfer matrix has diagonal entries $\cosh\kappa d$ and off-diagonal entries $-\sinh(\kappa d)/\kappa$ and $-\kappa\sinh(\kappa d)$. Insert (t, ikt) at the exit and $(1+r, ik(1-r))$ at the entrance. Eliminating r and taking the squared modulus gives the power transmission:

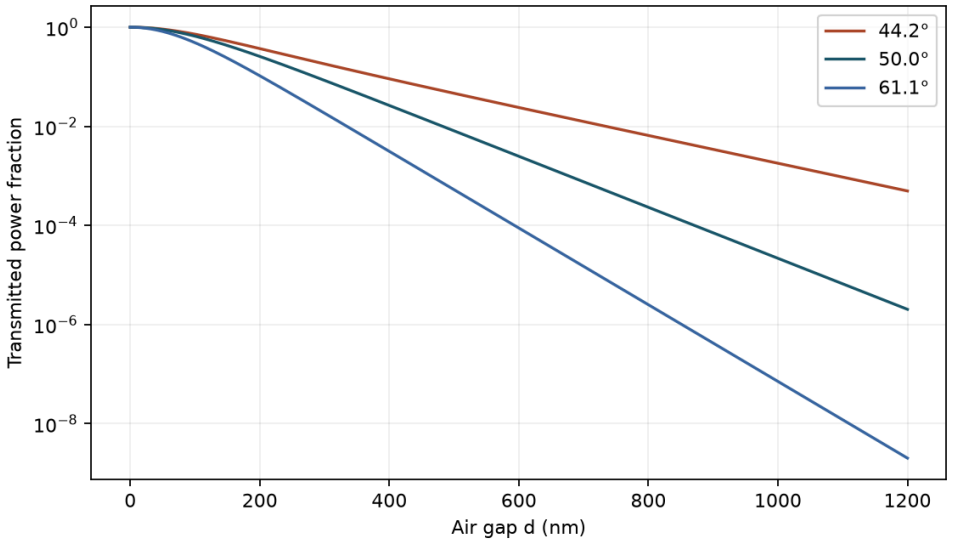
$$T = \left[1 + \frac{(k^2+\kappa^2)^2}{4k^2\kappa^2}\sinh^2(\kappa d)\right]^{-1}.$$

Both exterior media are identical, so transmitted power fraction equals squared transmitted amplitude. The finite flux through the gap comes from interference between its two evanescent components. The single exponential at an isolated interface carries no average normal power.

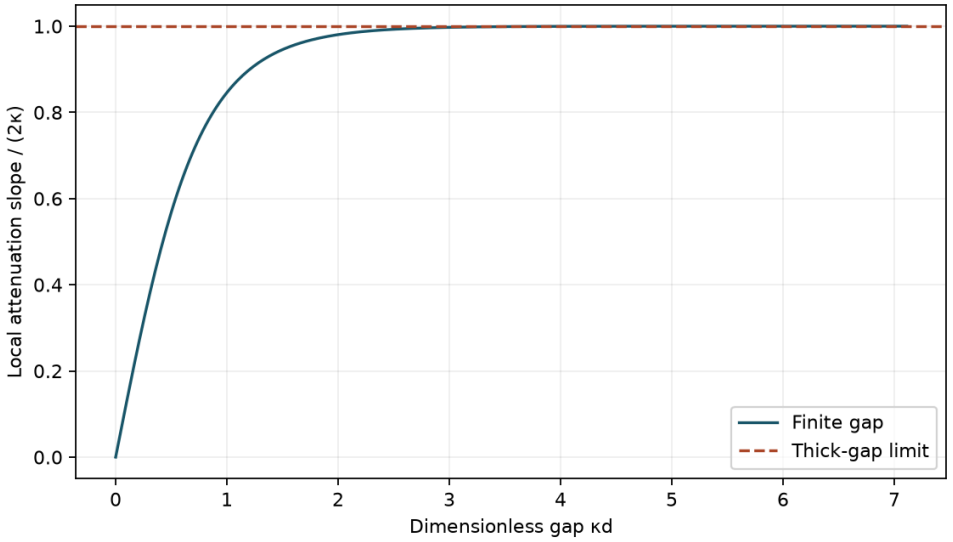
The penetration length sets a slope

At a large gap the logarithmic slope $-d\ln T/dd$ tends to 2κ. At contact $T = 1$ with zero slope, because the leading correction in gap thickness is quadratic. Measuring a straight-line slope over a very small gap would therefore return a biased estimate of the penetration depth. Use the first graph to choose the window for a fit and the second graph to assess the bias. Take a vacuum wavelength of 600 nm, index 1.50 and incidence angle 50°. The critical angle is 41.81°. The amplitude penetration length $1/\kappa$ is 168.72 nm and the coefficient of $\sinh^2(\kappa d)$ is 1.312. For gaps of 50, 100 and 200 nm the transmitted fractions are 0.894, 0.659 and 0.257. Doubling the gap therefore does not simply square the measured transmission. Only the sufficiently thick-gap logarithmic slope approaches a constant.

One can invert a single transmission reading when the angle and indices are independently known. Write $A = (k^2+\kappa^2)^2/(4k^2\kappa^2)$. Then $d = \kappa^{-1}\,\mathrm{asinh}\,\sqrt{(T^{-1}-1)/A}$. The reading 0.659 returns a gap of about 100 nm. Near $T = 1$ a small absolute power error makes a large relative error in the inferred gap. This follows directly from the quadratic expansion $1 - T \simeq A(\kappa d)^2$.



A higher transverse wavenumber causes a steeper thick-gap decay. At contact every angle has unit transmission.



Normalising by twice the decay constant turns the thick-gap limit into one. The early points reveal why a thin-gap slope fit fails.

Exercise · halve a thick-gap signal

How much extra separation is required to halve the transmitted power deep in the exponential regime?

▼ **Work through the solution**

The ratio tends to $\exp(-2\kappa\Delta d)$. Setting it to one half gives

$$\Delta d = \ln 2/(2\kappa).$$

This uses a power attenuation length. Confusing amplitude and intensity introduces a factor of two.

Exercise · choose a boundary at the far side

What changes if the second prism is removed while the incident prism and angle remain the same?

▼ **Work through the solution**

The gap becomes a half-space and boundedness removes the growing component. The transmitted evanescent amplitude can still be finite at the interface, while the average normal power flux becomes zero. The limit of the two-prism transmission agrees as d tends to infinity.

Try it · Evanescent gap

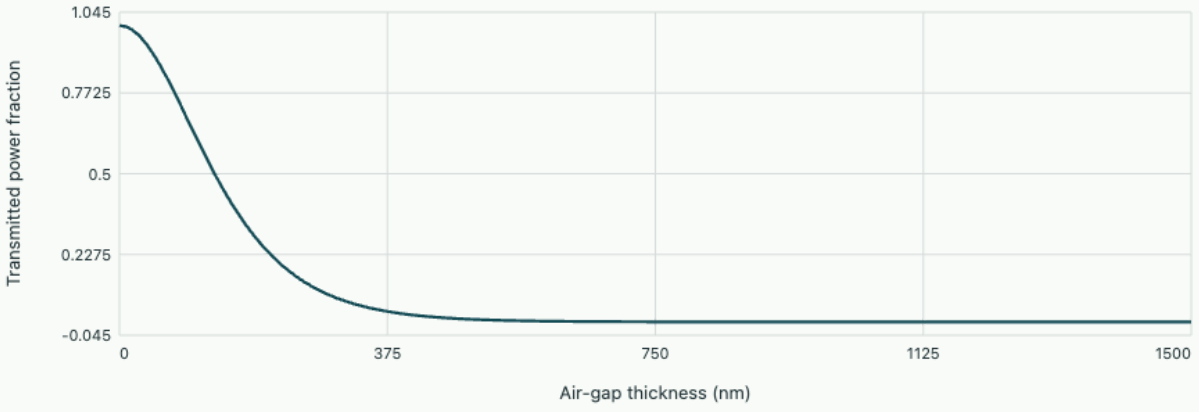
Change the inputs and compare the recalculated curves with the preceding derivation and baseline figure. The axes give the units; the numerical readout states the model assumptions and any limiting cases. Reset the inputs to return to the reference setting.

Open this live calculation

Try the calculation

Incidence angle (degrees) **50**

Wavelength (nm) **600**



Frustrated total internal reflection

Symmetric TE gap with surrounding index 1.5. Amplitude decay length 168.72 nm; critical angle 41.810°. At zero gap transmission is one.

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

Waves, radiation and relativity · 14. Resolve the self-force before taking a point-charge limit

A finite spherical charge distribution introduces a propagation time. Follow its memory kernel to electromagnetic inertia, radiative recoil and the limits of a truncated equation.

Give the charge a specified size and mechanical support

Take a thin spherical shell of radius a carrying fixed uniform total charge q. Prescribe rigid translational motion at speed much smaller than c, and linearise the self-field in that velocity. Mechanical stresses maintain the spherical distribution. Rotation, relativistic contraction and deformable internal charge modes lie outside this model. A finite radius is essential because two different surface elements communicate after a finite delay.

In the linearised retarded-field integral, the instantaneous longitudinal Coulomb pair forces cancel in the net force when using Coulomb gauge. The transverse inertial kernel contains the angular projector I–nn. Spherical averaging gives 2I/3. The resulting vector force is expressible as a pair-distance integral of the retarded acceleration.

$$\mathbf{F}_{\mathrm{self}}(t) = -\frac{2}{3c^2}\iint \frac{dq\,dq'}{4\pi\epsilon_0 R}\,\mathbf{a}(t-R/c).$$

This is the linear, rigid, spherically symmetric kernel. An independently incident electromagnetic field belongs to the applied force. A deformable shell would require an additional internal response model. The delayed time appears inside the acceleration. Pulling it out before averaging would retain electromagnetic inertia but remove the radiative correction. Here R is the distance between two surface points, evaluated in the unperturbed shell geometry.

Evaluate the pair distribution geometrically

Fix one point on the shell. A second point at polar angle θ from it is separated by R=2a sin(θ/2). The fraction of surface area in the annulus is sinθ dθ/2. Since R dR=a² sinθ dθ, the normalised distribution of pair distances is R dR/(2a²) on 0≤R≤2a. This cancels the kernel's inverse distance. Integrating the acceleration over the remaining retarded-time interval gives the exact linear delay formula for this model.

$$\mathbf{F}_{\mathrm{self}}(t) = \frac{q^2}{12\pi\epsilon_0 a^2 c}\,[\mathbf{v}(t-2a/c)-\mathbf{v}(t)].$$

Recall · A delay equation needs a history

A second-order particle equation is normally supplied with position and velocity at one instant. The shell equation samples velocity over the preceding interval of length 2a/c. A compatible history and the incoming electromagnetic field are additional information. Replacing that history by extra arbitrary derivatives changes the initial-value problem.

Separate inertia from the first dissipative correction

If the acceleration varies on a time scale T much longer than 2a/c, Taylor-expand the delayed velocity. The first term is proportional to acceleration and the second to its derivative.

$$\mathbf{F}_{\mathrm{self}} = -m_{\mathrm{em}}\mathbf{a} + \frac{q^2}{6\pi\epsilon_0 c^3}\dot{\mathbf{a}} + O(q^2 a\ddot{\mathbf{a}}/(\epsilon_0 c^4)),$$

$$m_{\mathrm{em}} = \frac{q^2}{6\pi\epsilon_0 a c^2}.$$

The electrostatic shell energy is U=q²/(8πε₀a), so this electromagnetic inertial coefficient is 4U/(3c²). The coefficient refers to the electromagnetic part of a constrained shell. Covariant energy-momentum accounting also includes the stresses that maintain its shape. Identifying this isolated coefficient with the complete rest mass would omit those stresses.

Move the inertial term to the left of m_bare a=F_ext+F_self and define the measured inertial mass m=m_bare+m_em. The remaining leading local term is the Abraham–Lorentz force mτ dot a, with τ=q²/(6πε₀mc³). In the limit a→0 at fixed q the electromagnetic inertia diverges. Holding the measured mass fixed then requires a singular adjustment of the mechanical parameter. The finite-shell model therefore does not justify an unrestricted point limit.

Check the mechanical power and the stored near field

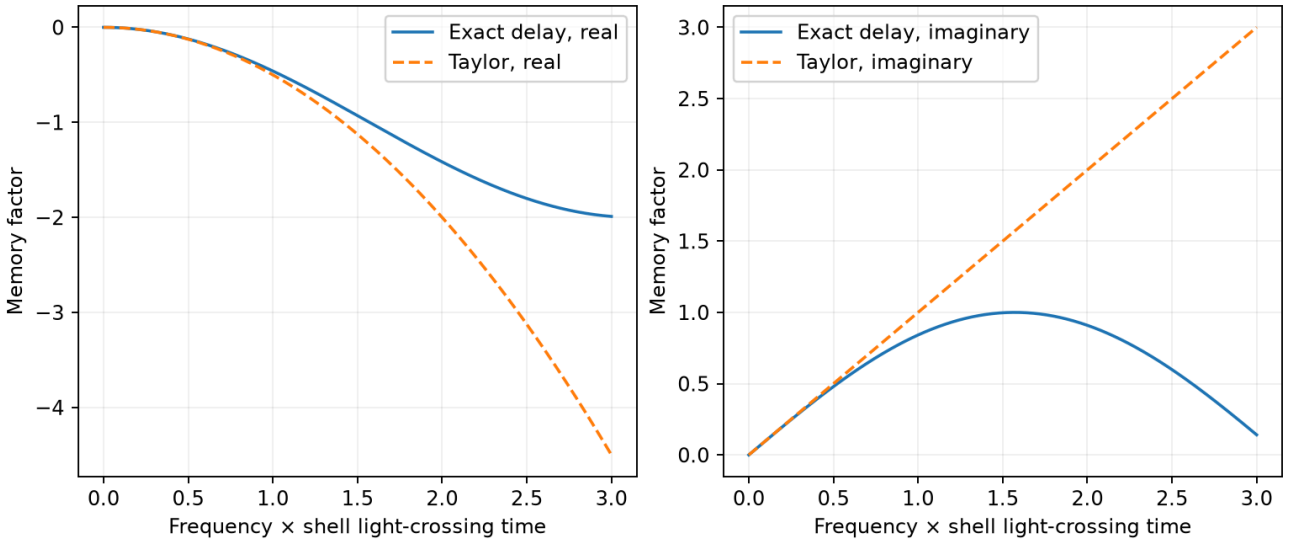
Dot the recoil term with velocity and use the product rule. The first term is a total derivative. The second is the Larmor power with a minus sign.

$$m\tau\dot{\mathbf{a}}\cdot\mathbf{v} = \frac{d}{dt}(m\tau\mathbf{a}\cdot\mathbf{v}) - m\tau a^2.$$

Defining the near-field energy contribution E_S=−mτa·v gives d(K+E_S)/dt=F_ext·v−P_rad. Over a pulse whose endpoint acceleration vanishes, the integrated near-field change is zero. At an intermediate instant it can be nonzero, so mechanical power loss and radiated power need not agree instant by instant.

Remember · A derivative expansion has a frequency range

For harmonic velocity $v \propto e^{-i\omega t}$, the exact delay factor is $e^{i\omega(2a/c)} - 1$. Its Taylor expansion is controlled by $|\omega| 2a/c$. Solving the truncated differential equation can create modes outside that range. Check the frequency of a proposed solution against the approximation that produced the equation.



The low-frequency approximation agrees near the origin. At order-one light-crossing phase the exact delay and its truncated differential approximation differ substantially.

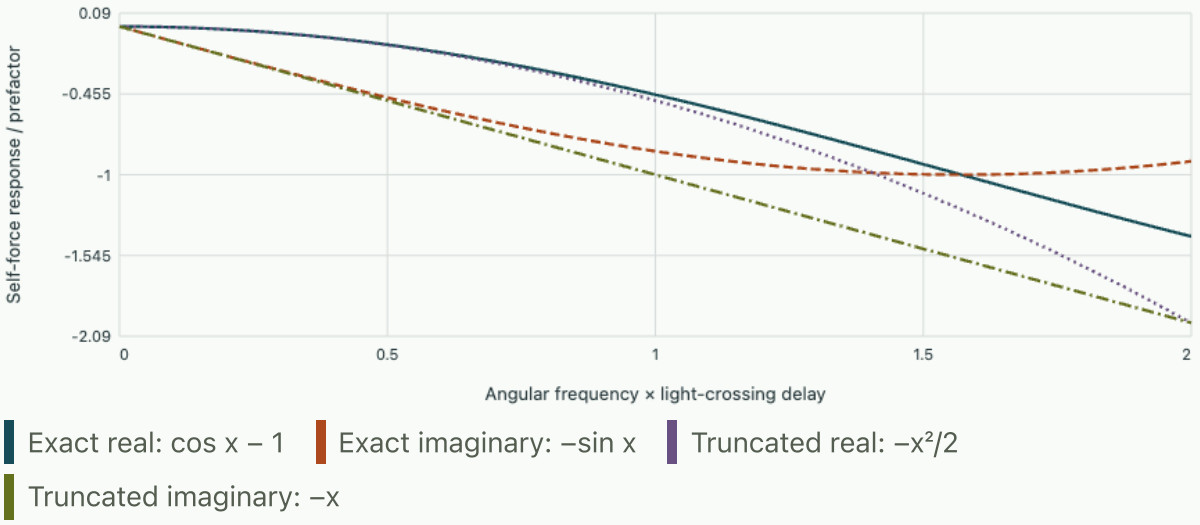
Try it · Delay force

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

Maximum light-crossing phase 2



For velocity proportional to exp(iωt), the delayed shell factor is exp(−iωτ)−1 with τ=2a/c. The differential approximation requires |ωτ| ≪ 1.

Reduce the order on the physical slow branch

The truncated force-free equation a=τ dot a admits an exponentially growing acceleration. Its time scale is τ, while the local expansion was intended to describe slowly varying applied forcing. On the perturbative branch replace a on the small correction's right-hand side by F_ext/m. To first order in τ this gives m a=F_ext+τ dF_ext/dt, with a total derivative along the leading trajectory.

For a harmonic restoring force F_ext=−mω₀²x, order reduction gives ẍ+τω₀²ẋ+ω₀²x=0. The amplitude decay rate is τω₀²/2. The average Larmor power divided by oscillator energy gives the same energy decay rate τω₀². This comparison tests both the sign and the factor two between amplitude and energy.

Exercises on limits and energy

1. For an electron, $\tau \approx 6.27 \times 10^{-24}$ s. Estimate the amplitude-decay time for $\omega_0 = 10^{15}$ s^{-1} within the classical oscillator model.

▼ Work through the solution

The amplitude rate is 3.135×10^{6} s^{-1} and the time is 3.19×10^{-7} s. The small parameter $\tau\omega_0$ is 6.27×10^{-9}. A specific atomic transition additionally requires quantum state and coupling information.

---

2. Evaluate the ratio of the exact shell memory factor to its first-order inertial approximation when $\omega 2a/c = 0.1$. Which next term captures loss?

▼ Work through the solution

The exact factor is $\cos 0.1 - 1 + i \sin 0.1$. Relative to $i0.1$ it is approximately $0.998334 + i0.049958$. The real negative term $-(\omega 2a/c)^2/2$ in the force factor gives the first dissipative correction. Keeping only the linear imaginary term gives reactive inertia.

---

3. A force has compact support and the acceleration vanishes at both endpoints. Show that net radiated energy equals external work minus the change in kinetic energy.

▼ Work through the solution

Integrate the balance for $K + E_S$. The endpoint values of E_S vanish, leaving $W_{ext} - \Delta K = \int P_{rad}\, dt$. Truncating the interval before acceleration vanishes would require its near-field endpoint contribution.

For the finite-history model and its asymptotic limits, see the original analysis by Herbert Spohn. Radiative damping, a case study.

Waves, radiation and relativity · 15. From light clocks to collision kinematics

Construct spacetime measurements before transforming fields. Use the same invariant interval to compare clocks, add velocities and close a relativistic collision.

Make the two postulates operational

Assume inertial experiments obey the same physical laws and that vacuum light signals have the same speed c in each inertial frame. A light clock consists of two parallel mirrors separated transversely by distance d in its rest frame. One round trip takes Δτ=2d/c. An observer seeing the clock move at speed v measures a longer diagonal light path. For half of the round trip, the right triangle gives $(c\Delta t/2)^2=d^2+(v\Delta t/2)^2$. Solving yields Δt=γΔτ.

Here proper time means the reading of a clock carried along the two events. For arbitrary smooth timelike motion, split the trajectory into short inertial pieces and integrate.

$$d\tau = dt\sqrt{1 - v^2/c^2},$$

$$\Delta\tau = \int_{t_1}^{t_2} \sqrt{1 - v(t)^2/c^2}\, dt.$$

> **Recall · A reading at one place differs from comparing distant clocks**
>
> Proper time uses successive events at the same moving clock. Measuring the length of a moving rod instead requires its two endpoints at the same time in the observer's frame. Changing that simultaneity condition changes the pair of events used in the measurement.

Construct the Lorentz map and its diagram

Homogeneity makes the map between aligned inertial coordinates linear. The moving origin has x=vt, fixing x′=A(x−vt). Requiring each ray x=±ct also to satisfy x′=±ct′ fixes t′=A(t−vx/c²). Symmetry under reversing v determines A=γ. The y and z coordinates stay unchanged for this boost.

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta \\ -\gamma\beta & \gamma \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix}, \qquad (c\Delta t)^2 - \Delta x^2 = (c\Delta t')^2 - \Delta x'^2.$$

On a diagram with x horizontal and ct vertical, the moving time axis is x=βct. The moving simultaneity axis is ct=βx. A light ray always has slope ±1. For events separated by Δx=3 m and Δt=0, a frame with β=0.6 assigns Δt′=−7.505 ns. Their interval is spacelike, so their order can change without reversing the order of a causal signal. For timelike events, a subluminal boost preserves the temporal order.

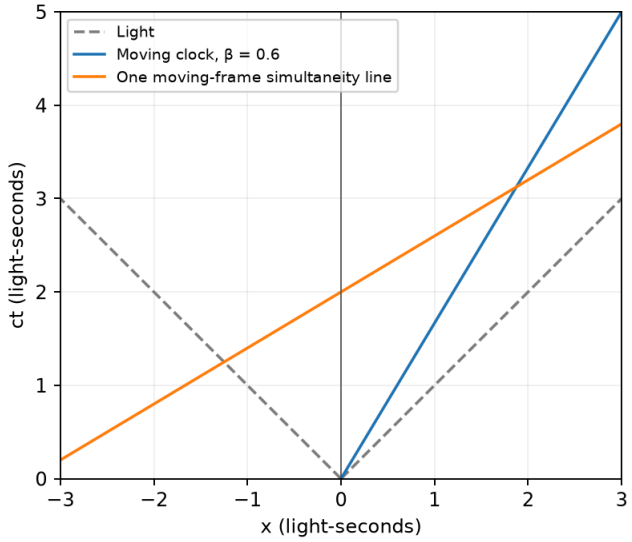
A rod of rest length L_0 has endpoint separation $\Delta x'=L_0$. Its laboratory length uses Δt=0, giving Δx′=γΔx and $L=L_0/\gamma$. Requiring Δt′=0 instead would describe another pair of events. Differentiating the

transformation gives the velocity composition rule.

$$u'_x = \frac{u_x - v}{1 - u_x v/c^2},$$

$$u'_y = \frac{u_y}{\gamma(1 - u_x v/c^2)}.$$

Two speeds of 0.8c in the same direction compose to 1.6c/1.64=0.97561c. Light still gives c. Introducing rapidity η through β=tanhη turns successive collinear boosts into addition of η, which is convenient when several observer changes are chained.



A moving time axis lies inside the light cone. Its simultaneity lines tilt relative to the laboratory horizontal.

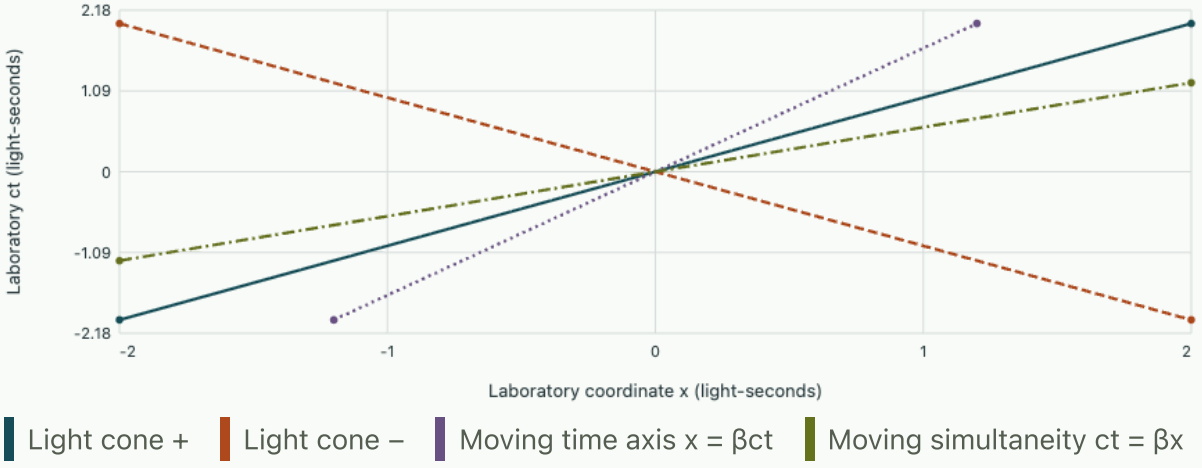
Try it · Spacetime

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

Observer speed v/c **0.6**



Lorentz factor γ = 1.25. Axes tilt with opposite roles; light directions remain invariant.

Compare clocks that meet again

In a chosen laboratory frame, let one clock remain at the origin while another travels out and back at speed 0.6c for a total laboratory duration of 10 years, with a short idealised turnaround. The stay-at-home clock records 10 years, and the travelling one records 8 years. Both readings can be compared at the same final event. The travelling clock uses two inertial frames separated by the turnaround. Its change of simultaneity prevents combining the outward and inward time-dilation arguments as if there were one travelling inertial frame.

Acceleration can be spread over a finite time and inserted into the proper-time integral. The measured age difference follows the whole path. Merely knowing a peak acceleration or a final relative speed does not determine it.

Derive the quantities that survive a collision

For a free particle use the invariant action S=−mc²∫dτ, giving L=−mc²√(1−v²/c²). Differentiate with respect to v to obtain p=γmv. The Legendre transform p·v−L gives E=γmc². Eliminate v to obtain the invariant mass relation.

$$E^2 - c^2p^2 = m^2c^4,$$

$$K = (\gamma - 1)mc^2.$$

For an isolated collision add the incoming four-momenta and equate them to the outgoing sum. The invariant of the total four-momentum defines the centre-of-momentum energy, including energy that becomes internal excitation. The velocity of that frame along the total momentum is V=c²P/E_total.

As an example, two identical particles of mass m approach with speeds ±0.6c and stick. Each has γ=1.25. The total momentum vanishes, and the final invariant mass is 2.5m. The difference from 2m is stored internal energy 0.5mc². In another inertial frame the composite moves, but its invariant mass is the same.

Find a production threshold with one stationary target

Let an incident particle with mass m and total energy E hit an identical stationary target. Squaring the total four-momentum gives s=2m²c⁴+2mc²E. Consider endothermic production with minimum final total rest mass Mf≥2m. At threshold it has no relative kinetic energy in its centre-of-momentum frame, so s=Mf²c⁴.

$$E_{\text{threshold}} = \frac{(M_f^2 - 2m^2)c^2}{2m}.$$

If Mf=3m, the required incident total energy is 3.5mc² and its kinetic energy is 2.5mc². The newly created rest energy is only mc², but momentum conservation forces part of the available laboratory energy into overall recoil. A collider with oppositely directed equal momenta uses the incoming kinetic energy differently.

Remember · Use total energy when constructing a four-vector

The time component is E/c with E=K+mc². Substituting K/c would fail the invariant relation even for a particle at rest. Kinetic energy is convenient for a calorimeter, while total energy is required in the Lorentz transformation and threshold invariant.

Exercises with events and four-momenta

1. A moving clock at 0.8c records 3 μs between two flashes. Find their laboratory time and displacement.

▼ **Work through the solution**

$\gamma = 5/3$ gives $\Delta t = 5$ μs. The displacement is 0.8c times 5 μs≈1199.17 m. Direct substitution gives $c^2\Delta t^2 - \Delta x^2 = c^2(3\ \mu s)^2$.

---

2. Two photons of energies 2 MeV and 1 MeV travel in opposite directions. Determine the invariant mass of their combined system and its centre-of-momentum velocity.

▼ **Work through the solution**

Total energy is 3 MeV and cP=1 MeV. Hence $Mc^2 = \sqrt{8}$ MeV≈2.828 MeV and V=c/3 toward the more energetic photon. Each individual photon has zero rest mass.

---

3. At β=0.6 a rod has laboratory length 0.80 m. Determine its rest length, then find the laboratory time difference for endpoint events simultaneous in the rod frame.

▼ **Work through the solution**

$L_0 = \gamma L = 1.00$ m. For $\Delta t' = 0$ and $\Delta x' = L_0$, the inverse transformation gives $\Delta t = \gamma v L_0 / c^2 = 2.502$ ns. Those events cannot be used directly for the laboratory length measurement.