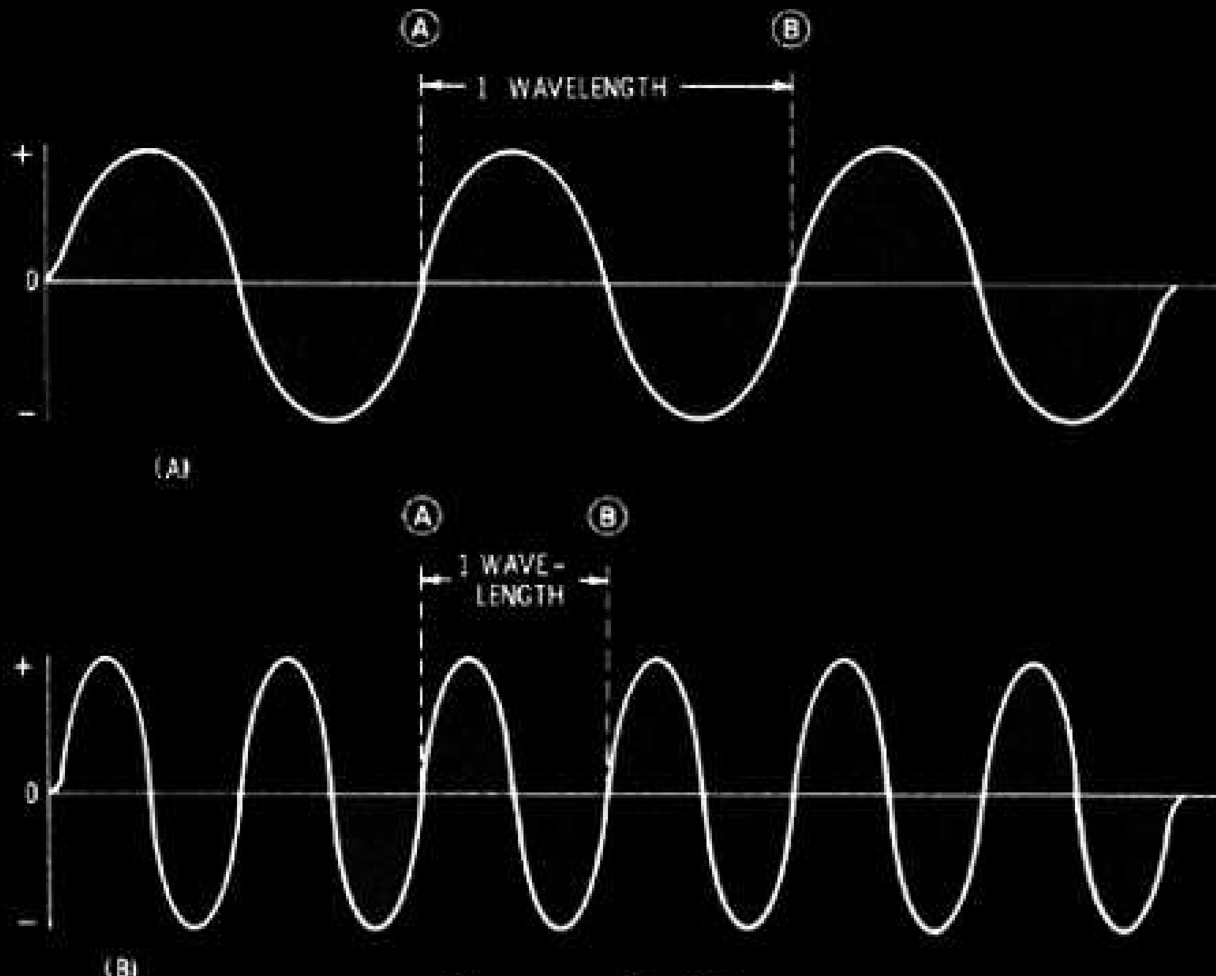
Wave

Physics and Engineering

(Concepts and Applications)



Larue Sutter

Mark Teel

First Edition, 2012

ISBN 978-81-323-2001-2

Published by:
Library Press
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1 - Wave Equation

Chapter 2 - Phase Velocity and Group Velocity

Chapter 3 - Plane Waves and Standing Waves

Chapter 4 - Dispersion (optics)

Chapter 5 - Seismic Wave

Chapter 6 - Electromagnetic Radiation

Chapter 7 - Wave Function

Chapter 8 - Airy Wave Theory

Chapter 9 - Rogue Wave

Chapter 10 - Cnoidal Wave

Chapter 11 - Stokes Drift

Chapter 12 - Breaking Wave

Chapter 13 - Radiation Stress

Chapter 14 - Internal Wave & Kelvin Wave

Chapter 15 - Wave Radar

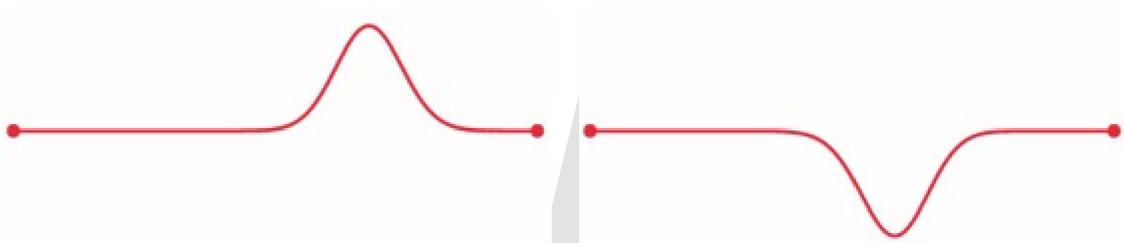
Chapter 16 - Wave Shoaling & Wind Wave Model

Chapter 17 - Mild–Slope Equation

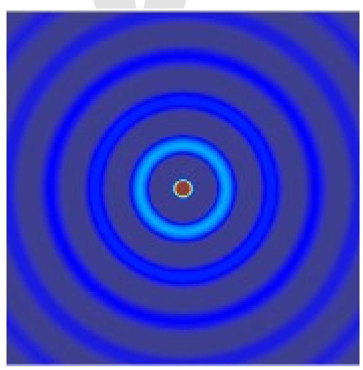
Chapter 1

Wave Equation

The **wave equation** is an important second-order linear partial differential equation of waves, such as sound waves, light waves and water waves. It arises in fields such as acoustics, electromagnetics, and fluid dynamics. Historically, the problem of a vibrating string such as that of a musical instrument was studied by Jean le Rond d'Alembert, Leonhard Euler, Daniel Bernoulli, and Joseph-Louis Lagrange.



A pulse traveling through a string with fixed endpoints as modeled by the wave equation



Spherical waves coming from a point source

Introduction

The wave equation is the prototypical example of a hyperbolic partial differential equation. In its simplest form, the wave equation refers to a scalar function $u=(x_1, x_2,\ldots,x_n,t)$ that satisfies:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \nabla^2 u$$

where $\nabla^2$is the (spatial) Laplacian and where *c* is a fixed constant equal to the propagation speed of the wave. This is known as the non-dispersive wave equation. For a sound wave in air at 20°C this constant is about 343 m/s. For the vibration of a string the speed can vary widely, depending upon the linear density of the string and the tension on it. For a spiral spring (a slinky) it can be as slow as a meter per second. More realistic differential equations for waves allow for the speed of wave propagation to vary with the frequency of the wave, a phenomenon known as dispersion. In such a case, *c* must be replaced by the phase velocity:

$$v_\mathrm{p} = \frac{\omega}{k}.$$

Another common correction in realistic systems is that the speed can also depend on the amplitude of the wave, leading to a nonlinear wave equation:

$$\frac{\partial^2 u}{\partial t^2} = c(u)^2 \nabla^2 u$$

Also note that a wave may be superimposed onto another movement (for instance sound propagation in a moving medium like a gas flow). In that case the scalar *u* will contain a Mach factor (which is positive for the wave moving along the flow and negative for the reflected wave).

The elastic wave equation in three dimensions describes the propagation of waves in an isotropic homogeneous elastic medium. Most solid materials are elastic, so this equation describes such phenomena as seismic waves in the Earth and ultrasonic waves used to detect flaws in materials. While linear, this equation has a more complex form than the equations given above, as it must account for both longitudinal and transverse motion:

$$\rho \ddot{\mathbf{u}} = \mathbf{f} + (\lambda + 2\mu)\nabla(\nabla \cdot \mathbf{u}) - \mu\nabla \times (\nabla \times \mathbf{u})$$

where:

- $\lambda$and $\mu$are the so-called Lamé parameters describing the elastic properties of the medium,
- $\rho$is the density,
- $\mathbf{f}$is the source function (driving force),
- and $\mathbf{u}$is the displacement vector.

Note that in this equation, both force and displacement are vector quantities. Thus, this equation is sometimes known as the vector wave equation.

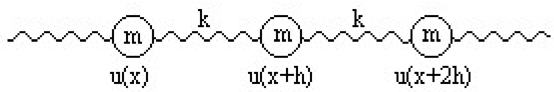
Variations of the wave equation are also found in quantum mechanics, plasma physics and general relativity.

Scalar wave equation in one space dimension

Derivation of the wave equation

From Hooke's law

The wave equation in the one dimensional case can be derived from Hooke's law in the following way: Imagine an array of little weights of mass *m* interconnected with massless springs of length *h* . The springs have a stiffness of *k*:



Here *u(x)* measures the distance from the equilibrium of the mass situated at *x*. The forces exerted on the mass *m* at the location *x* + *h* are:

$$F_{Newton} = m \cdot a(t) = m \cdot \frac{\partial^2}{\partial t^2} u(x+h,t)$$
$$F_{Hooke} = F_{x+2h} + F_x = k\left[u(x+2h,t) - u(x+h,t)\right] + k[u(x,t) - u(x+h,t)]$$

The equation of motion for the weight at the location *x*+*h* is given by equating these two forces:

$$m\frac{\partial^2 u(x+h,t)}{\partial t^2} = k[u(x+2h,t) - u(x+h,t) - u(x+h,t) + u(x,t)]$$

where the time-dependence of *u*(*x*) has been made explicit.

If the array of weights consists of *N* weights spaced evenly over the length *L* = *N h* of total mass *M* = *N m*, and the total stiffness of the array *K* = *k*/*N* we can write the above equation as:

$$\frac{\partial^2 u(x+h,t)}{\partial t^2} = \frac{KL^2}{M}\frac{[u(x+2h,t) - 2u(x+h,t) + u(x,t)]}{h^2}$$

Taking the limit $N \rightarrow \infty, h \rightarrow 0$(and assuming smoothness) one gets:

$$\frac{\partial^2 u(x,t)}{\partial t^2} = \frac{KL^2}{M}\frac{\partial^2 u(x,t)}{\partial x^2}$$

$(KL^2)/M$ is the square of the propagation speed in this particular case.

From the generic scalar transport equation

Starting with the generic scalar transport equation without diffusion,

$$\frac{\partial \phi}{\partial t} + \frac{\partial (u\phi)}{\partial x} = S_\phi,$$

we differentiate with respect to *t* to get

$$\frac{\partial^2 \phi}{\partial t^2} + \frac{\partial^2 (u\phi)}{\partial x \partial t} = \frac{\partial S_\phi}{\partial t}.$$

Assuming that S_φ and *u* are constant, we may write

$$\frac{\partial^2 \phi}{\partial t^2} + u\frac{\partial}{\partial x}\frac{\partial \phi}{\partial t} = 0.$$

Substituting for the time derivative of φ we get

$$\frac{\partial^2 \phi}{\partial t^2} + u\frac{\partial}{\partial x}\left[S_\phi - \frac{\partial (u\phi)}{\partial x}\right] = \frac{\partial^2 \phi}{\partial t^2} - u^2\frac{\partial^2 \phi}{\partial x^2} = 0,$$

where *u* is the speed of propagation of the scalar φ which, in general, is a function of time and position.

General solution

The one dimensional wave equation is unusual for a partial differential equation in that a very simple general solution may be found. Defining new variables:

$$\xi = x - ct \quad ; \quad \eta = x + ct$$

changes the wave equation into

$$\frac{\partial^2 u}{\partial \xi \partial \eta} = 0$$

which leads to the general solution

$$u(\xi,\eta) = F(\xi) + G(\eta) \quad \Rightarrow \quad u(x,t) = F(x - ct) + G(x + ct)$$

In other words, solutions of the 1D wave equation are sums of a right traveling function *F* and a left traveling function *G*. "Traveling" means that the shape of these individual arbitrary functions with respect to *x* stays constant, however the functions are translated left and right with time at the speed *c*. This was derived by Jean le Rond d'Alembert.

Another way to arrive at this result is to note that the wave equation may be "factored":

$$\left[\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right]\left[\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right] u = 0$$

$$\Downarrow$$

$$\frac{\partial u}{\partial t} - c\frac{\partial u}{\partial x} = 0 \qquad \text{and} \qquad \frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0$$

These last two equations are advection equations, one left traveling and one right, both with constant speed *c*.

For an initial value problem, the arbitrary functions *F* and *G* can be determined to satisfy initial conditions:

$$u(x,0) = f(x)$$
$$u_t(x,0) = g(x)$$

The result is d'Alembert's formula:

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

In the classical sense if $f(x) \in C^k$ and $g(x) \in C^{k-1}$ then $u(t,x) \in C^k$. However, the waveforms *F* and *G* may also be generalized functions, such as the delta-function. In that case, the solution may be interpreted as an impulse that travels to the right or the left.

The basic wave equation is a linear differential equation and so it will adhere to the superposition principle. This means that the net displacement caused by two or more waves is the sum of the displacements which would have been caused by each wave individually. In addition, the behavior of a wave can be analyzed by breaking up the wave into components, e.g. the Fourier transform breaks up a wave into sinusoidal components.

Scalar wave equation in three space dimensions

The solution of the initial-value problem for the wave equation in three space dimensions can be obtained from the solution for a spherical wave. This result can then be used to obtain the solution in two space dimensions.

Spherical waves

The wave equation is unchanged under rotations of the spatial coordinates, and therefore one may expect to find solutions that depend only on the radial distance from a given point. Such solutions must satisfy

$$u_{tt} - c^2 \left(u_{rr} + \frac{2}{r} u_r \right) = 0.$$

This equation may be rewritten as

$$(ru)_{tt} - c^2 (ru)_{rr} = 0;$$

the quantity *ru* satisfies the one-dimensional wave equation. Therefore there are solutions in the form

$$u(t,r) = \frac{1}{r} F(r - ct) + \frac{1}{r} G(r + ct),$$

where *F* and *G* are arbitrary functions. Each term may be interpreted as a spherical wave that expands or contracts with velocity *c*. Such waves are generated by a point source, and they make possible sharp signals whose form is altered only by a decrease in amplitude as *r* increases. Such waves exist only in cases of space with odd dimensions.

Solution of a general initial-value problem

The wave equation is linear in *u* and it is left unaltered by translations in space and time. Therefore we can generate a great variety of solutions by translating and summing spherical waves. Let $\varphi(\xi,\eta,\zeta)$ be an arbitrary function of three independent variables, and let the spherical wave form *F* be a delta-function: that is, let *F* be a weak limit of continuous functions whose integral is unity, but whose support (the region where the function is non-zero) shrinks to the origin. Let a family of spherical waves have center at (ξ,η,ζ), and let *r* be the radial distance from that point. Thus

$$r^2 = (x - \xi)^2 + (y - \eta)^2 + (z - \zeta)^2.$$

If *u* is a superposition of such waves with weighting function φ, then

$$u(t,x,y,z) = \frac{1}{4\pi c} \iiint \varphi(\xi,\eta,\zeta) \frac{\delta(r - ct)}{r} d\xi \, d\eta \, d\zeta;$$

the denominator $4\pi c$ is a convenience.

From the definition of the delta-function, *u* may also be written as

$$u(t,x,y,z) = \frac{t}{4\pi} \iint_S \varphi(x + ct\alpha, y + ct\beta, z + ct\gamma) d\omega,$$

where α, β, and γ are coordinates on the unit sphere *S*, and ω is the area element on *S*. This result has the interpretation that *u*(*t*,*x*) is *t* times the mean value of φ on a sphere of radius *ct* centered at *x*:

$$u(t,x,y,z) = tM_{ct}[\phi].$$

It follows that

$$u(0,x,y,z) = 0, \quad u_t(0,x,y,z) = \phi(x,y,z).$$

The mean value is an even function of *t*, and hence if

$$v(t,x,y,z) = \frac{\partial}{\partial t}\left(tM_{ct}[\psi]\right),$$

then

$$v(0,x,y,z) = \psi(x,y,z), \quad v_t(0,x,y,z) = 0.$$

These formulas provide the solution for the initial-value problem for the wave equation. They show that the solution at a given point *P*, given (*t*,*x*,*y*,*z*) depends only on the data on the sphere of radius *ct* that is intersected by the **light cone** drawn backwards from *P*. It does *not* depend upon data on the interior of this sphere. Thus the interior of the sphere is a lacuna for the solution. This phenomenon is called **Huygens' principle**. It is true for odd numbers of space dimension, where for one dimension the integration is performed over the boundary of an interval w.r.t. the Dirac measure. It is not satisfied in even space dimensions. The phenomenon of lacunas has been extensively investigated in Atiyah, Bott and Gårding (1970, 1973).

Scalar wave equation in two space dimensions

In two space dimensions, the wave equation is

$$u_{tt} = c^2\left(u_{xx} + u_{yy}\right).$$

We can use the three-dimensional theory to solve this problem if we regard *u* as a function in three dimensions that is independent of the third dimension. If

$$u(0,x,y) = 0, \quad u_t(0,x,y) = \phi(x,y),$$

then the three-dimensional solution formula becomes

$$u(t,x,y) = tM_{ct}[\phi] = \frac{t}{4\pi}\iint_S \phi(x + ct\alpha,\, y + ct\beta)d\omega,$$

where α and β are the first two coordinates on the unit sphere, and dω is the area element on the sphere. This integral may be rewritten as an integral over the disc *D* with center (*x*,*y*) and radius *ct*:

$$u(t,x,y) = \frac{1}{2\pi c}\iint_D \frac{\phi(x+\xi, y+\eta)}{\sqrt{(ct)^2 - \xi^2 - \eta^2}} d\xi\, d\eta.$$

It is apparent that the solution at (*t*,*x*,*y*) depends not only on the data on the light cone where

$$(x-\xi)^2 + (y-\eta)^2 = c^2t^2,$$

but also on data that are interior to that cone.

Problems with boundaries

One space dimension

A flexible string that is stretched between two points $x = 0$ and $x = L$ satisfies the wave equation for $t > 0$ and $0 < x < L$. On the boundary points, *u* may satisfy a variety of boundary conditions. A general form that is appropriate for applications is

$$-u_x(t,0) + au(t,0) = 0,$$
$$u_x(t,L) + bu(t,L) = 0,$$

where *a* and *b* are non-negative. The case where u is required to vanish at an endpoint is the limit of this condition when the respective *a* or *b* approaches infinity. The method of separation of variables consists in looking for solutions of this problem in the special form

$$u(t,x) = T(t)v(x).$$

A consequence is that

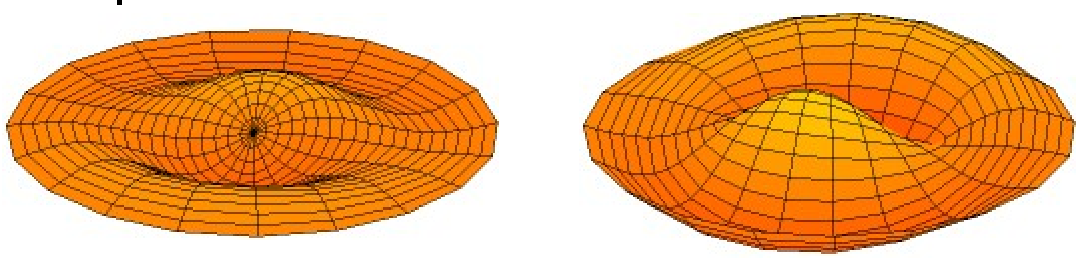
$$\frac{T''}{c^2T} = \frac{v''}{v} = -\lambda.$$

The eigenvalue λ must be determined so that there is a non-trivial solution of the boundary-value problem

$$v'' + \lambda v = 0,$$
$$-v'(0) + av(0) = 0, \quad v'(L) + bv(L) = 0.$$

This is a special case of the general problem of Sturm–Liouville theory. If *a* and *b* are positive, the eigenvalues are all positive, and the solutions are trigonometric functions. A solution that satisfies square-integrable initial conditions for *u* and u_t can be obtained from expansion of these functions in the appropriate trigonometric series.

Several space dimensions



A solution of the wave equation in two dimensions with a zero-displacement boundary condition along the entire outer edge.

The one-dimensional initial-boundary value theory may be extended to an arbitrary number of space dimensions. Consider a domain *D* in *m*-dimensional *x* space, with boundary *B*. Then the wave equation is to be satisfied if *x* is in *D* and $t > 0$. On the boundary of *D*, the solution *u* shall satisfy

$$\frac{\partial u}{\partial n} + au = 0,$$

where *n* is the unit outward normal to *B*, and *a* is a non-negative function defined on *B*. The case where *u* vanishes on *B* is a limiting case for *a* approaching infinity. The initial conditions are

$$u(0, x) = f(x), \quad u_t = g(x),$$

where *f* and *g* are defined in *D*. This problem may be solved by expanding *f* and *g* in the eigenfunctions of the Laplacian in *D*, which satisfy the boundary conditions. Thus the eigenfunction *v* satisfies

$$\nabla \cdot \nabla v + \lambda v = 0,$$

in *D*, and

$$\frac{\partial v}{\partial n} + av = 0,$$

on *B*.

In the case of two space dimensions, the eigenfunctions may be interpreted as the modes of vibration of a drumhead stretched over the boundary *B*. If *B* is a circle, then these eigenfunctions have an angular component that is a trigonometric function of the polar angle θ, multiplied by a Bessel function (of integer order) of the radial component. Further details are in Helmholtz equation.

If the boundary is a sphere in three space dimensions, the angular components of the eigenfunctions are spherical harmonics, and the radial components are Bessel functions of half-integer order.

Inhomogenous wave equation in one dimension

The inhomogenous wave equation in one dimension is the following:

$$c^2 u_{xx}(x,t) - u_{tt}(x,t) = s(x,t)$$

with initial conditions given by

$$u(x,0) = f(x)$$
$$u_t(x,0) = g(x)$$

The function $s(x,t)$ is often called the source function because in practice it describes the effects of the sources of waves on the medium carrying them. Physical examples of source functions include the force driving a wave on a string, or the charge or current density in the Lorenz gauge of electromagnetism.

One method to solve the initial value problem (with the initial values as posed above) is to take advantage of the property of the wave equation that its solutions obey causality. That is, for any point (x_i,t_i), the value of $u(x_i,t_i)$depends only on the values of $f(x_i+ct_i)$and $f(x_i-ct_i)$and the values of the function $g(x)$between $(x_i-ct_i)$and (x_i+ct_i). This can be seen in d'Alembert's formula, stated above, where these quantities are the only ones that show up in it. Physically, if the maximum propagation speed is c, then no part of the wave that can't propagate to a given point by a given time can affect the amplitude at the same point and time.

In terms of finding a solution, this causality property means that for any given point on the line being considered, the only area that needs to be considered is the area encompassing all the points that could causally affect the point being considered. Denote the area that casually affects point (x_i,t_i)as R_C. Suppose we integrate the nonhomogenous wave equation over this region.

$$\iint_{R_C} \left(c^2 u_{xx}(x,t) - u_{tt}(x,t)\right) dx dt = \iint_{R_C} s(x,t) dx dt.$$

To simplify this greatly, we can use Green's theorem to simplify the left side to get the following:

$$\int_{L_0+L_1+L_2} \left(-c^2 u_x(x,t) dt - u_t(x,t) dx\right) = \iint_{R_C} s(x,t) dx dt.$$

The left side is now the sum of three line integrals along the bounds of the causality region. These turn out to be fairly easy to compute

$$\int_{x_i-ct_i}^{x_i+ct_i} -u_t(x,0) dx = -\int_{x_i-ct_i}^{x_i+ct_i} g(x) dx.$$

In the above, the term to be integrated with respect to time disappears because the time interval involved is zero, thus $dt = 0$.

For the other two sides of the region, it is worth noting that $x \pm ct$ is a constant, namingly $x_i \pm ct_i$, where the sign is chosen appropriately. Using this, we can get the relation $dx \pm c dt = 0$, again choosing the right sign:

$$\begin{aligned} &\int_{L_1} \left(-c^2 u_x(x,t) dt - u_t(x,t) dx\right) \\ &= \int_{L_1} \left(c u_x(x,t) dx + c u_t(x,t) dt\right) \\ &= c \int_{L_1} du(x,t) = c u(x_i, t_i) - c f(x_i + c t_i). \end{aligned}$$

And similarly for the final boundary segment:

$$\begin{aligned} &\int_{L_2} \left(-c^2 u_x(x,t) dt - u_t(x,t) dx\right) \\ &= -\int_{L_2} \left(c u_x(x,t) dx + c u_t(x,t) dt\right) \\ &= -c \int_{L_2} du(x,t) = -\left(c f(x_i - c t_i) - c u(x_i, t_i)\right) \\ &= c u(x_i, t_i) - c f(x_i - c t_i). \end{aligned}$$

Adding the three results together and putting them back in the original integral:

$$-\int_{x_i-ct_i}^{x_i+ct_i} g(x)dx+cu(x_i,t_i)-cf(x_i+ct_i)+cu(x_i,t_i)-cf(x_i-ct_i)=\iint_{R_C} s(x,t)dxdt$$

$$2cu(x_i,t_i)-\int_{x_i-ct_i}^{x_i+ct_i} g(x)dx-cf(x_i+ct_i)-cf(x_i-ct_i)=\iint_{R_C} s(x,t)dxdt$$

$$2cu(x_i,t_i)=\int_{x_i-ct_i}^{x_i+ct_i} g(x)dx+cf(x_i+ct_i)+cf(x_i-ct_i)+\iint_{R_C} s(x,t)dxdt$$

$$u(x_i,t_i)=\frac{f(x_i+ct_i)+f(x_i-ct_i)}{2}+\frac{1}{2c}\int_{x_i-ct_i}^{x_i+ct_i} g(x)dx+\frac{1}{2c}\int_0^{t_i}\int_{x_i-c(t_i-t)}^{x_i+c(t_i-t)} s(x,t)dxdt.$$

In the last equation of the sequence, the bounds of the integral over the source function have been made explicit. Looking at this solution, which is valid for all choices (x_i,t_i) compatible with the wave equation, it is clear that the first two terms are simply d'Alembert's formula, as stated above as the solution of the homogenous wave equation in one dimension. The difference is in the third term, the integral over the source.

Other coordinate systems

In three dimensions, the wave equation, when written in elliptic cylindrical coordinates, may be solved by separation of variables, leading to the Mathieu differential equation.

Electromagnetic Wave Equation

The **electromagnetic wave equation** is a second-order partial differential equation that describes the propagation of electromagnetic waves through a medium or in a vacuum. The homogeneous form of the equation, written in terms of either the electric field **E** or the magnetic field **B**, takes the form:

$$\left(\nabla^2-\mu\epsilon\frac{\partial^2}{\partial t^2}\right)\mathbf{E} = 0$$
$$\left(\nabla^2-\mu\epsilon\frac{\partial^2}{\partial t^2}\right)\mathbf{B} = 0$$

where $c=\frac{1}{\sqrt{\mu\epsilon}}$ is the speed of light in the medium, and ∇^2 is the Laplace operator. In a vacuum, $c = c_0$ = 299,792,458 meters per second, which is the speed of light in free space. The electromagnetic wave equation derives from Maxwell's equations. It should also be noted that in most older literature, **B** is called the *magnetic flux density* or *magnetic induction*.

The origin of the electromagnetic wave equation

Conservation of charge

Conservation of charge requires that the time rate of change of the total charge enclosed within a volume *V* must equal the net current flowing into the surface *S* enclosing the volume:

$$\oint_S \mathbf{J} \cdot d\mathbf{A} = -\frac{d}{dt} \int_V \rho \cdot dV$$

where **J** is the current density (in amperes per square meter) flowing through the surface and ρ is the charge density (in coulombs per cubic meter) at each point in the volume.

From the divergence theorem, this relationship can be converted from integral form to differential form:

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t}$$

Ampère's circuital law prior to Maxwell's correction

In its original form, Ampère's circuital law relates the magnetic field **B** to the current density **J**:

$$\oint_C \mathbf{B} \cdot d\mathbf{l} = \iint_S \mu \mathbf{J} \cdot d\mathbf{A}$$

where *S* is an open surface terminated in the curve *C*. This integral form can be converted to differential form, using Stokes' theorem:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$$

Inconsistency between Ampère's circuital law and the law of conservation of charge

Taking the divergence of both sides of Ampère's circuital law gives:

$$\nabla \cdot (\nabla \times \mathbf{B}) = \nabla \cdot \mu_0 \mathbf{J}$$

The divergence of the curl of any vector field, including the magnetic field **B**, is always equal to zero:

$$\nabla \cdot (\nabla \times \mathbf{B}) = 0$$

Combining these two equations implies that

$$\nabla \cdot \mu_0 \mathbf{J} = 0$$

Because $\mu_0$is nonzero constant, it follows that

$$\nabla \cdot \mathbf{J} = 0$$

However, the law of conservation of charge tells that

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t}$$

Hence, as in the case of Kirchhoff's circuit laws, Ampère's circuital law would appear only to hold in situations involving constant charge density. This would rule out the situation that occurs in the plates of a charging or a discharging capacitor.

Maxwell's correction to Ampère's circuital law

Maxwell conceived of displacement current in connection with linear polarization of a dielectric medium. The concept has since been extended to apply to the vacuum. The justification of this virtual extension of displacement current is as follows:

Gauss's law in integral form states:

$$\oint_S \mathbf{E} \cdot d\mathbf{A} = \frac{1}{\varepsilon_0} \int_V \rho \, dV$$

where *S* is a closed surface enclosing the volume *V*. This integral form can be converted to differential form using the divergence theorem:

$$\nabla \cdot \varepsilon_0 \mathbf{E} = \rho$$

Taking the time derivative of both sides and reversing the order of differentiation on the left-hand side gives:

$$\nabla \cdot \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} = \frac{\partial \rho}{\partial t}$$

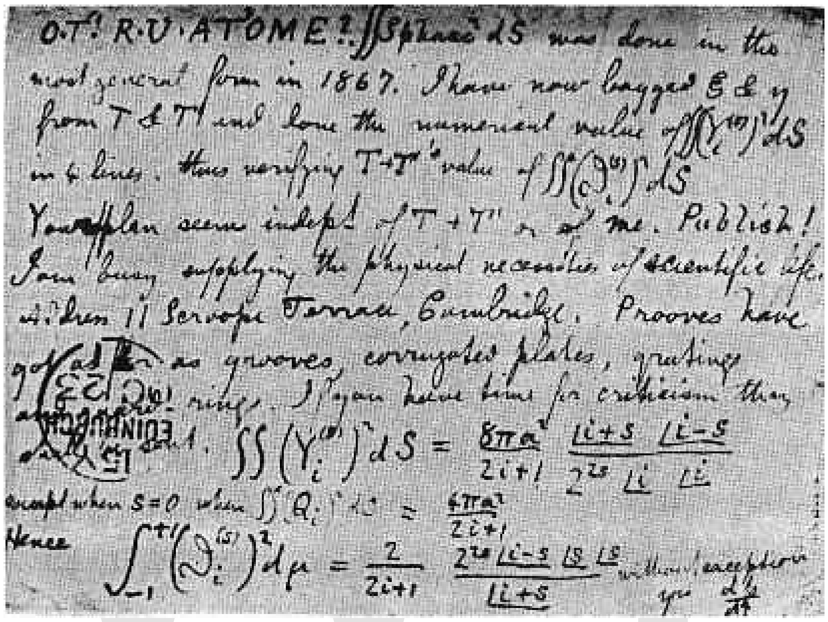
This last result, along with Ampère's circuital law and the conservation of charge equation, suggests that there are actually *two* origins of the magnetic field: the current density **j**, as Ampère had already established, and the so-called **displacement current**:

$$\frac{\partial \mathbf{D}}{\partial t} = \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

So the corrected form of Ampère's circuital law becomes:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

Maxwell's hypothesis that light is an electromagnetic wave


O.T.! R.V. ATOME?
was done in the most general form in 1867. I have now bagged ξ & η from T & T' and done the numerical value of
in 4 lines. thus verifying T+T' value of
Your plan seems indep^t of T+T' or of me. Publish!
I am busy supplying the physical necessities of scientific life within 11 Scroope Terrace, Cambridge. Grooves have
got as far as grooves, corrugated plates, gratings
rings. If you have time for criticism then
$$\iint \left(Y_i^{(s)}\right)^2 dS = \frac{8\pi a^2}{2i+1}$$
except when s=0
Hence
$$\int_{-1}^{+1} \left(\vartheta_i^{(s)}\right)^2 d\mu = \frac{2}{2i+1}$$

A postcard from Maxwell to Peter Tait

In his 1864 paper titled A Dynamical Theory of the Electromagnetic Field, Maxwell utilized the correction to Ampère's circuital law that he had made in part III of his 1861 paper On Physical Lines of Force. In PART VI of his 1864 paper which is entitled 'ELECTROMAGNETIC THEORY OF LIGHT', Maxwell combined displacement current with some of the other equations of electromagnetism and he obtained a wave equation with a speed equal to the speed of light. He commented:

> *The agreement of the results seems to show that light and magnetism are affections of the same substance, and that light is an electromagnetic disturbance propagated through the field according to electromagnetic laws.*

Maxwell's derivation of the electromagnetic wave equation has been replaced in modern physics by a much less cumbersome method involving combining the corrected version of Ampère's circuital law with Faraday's law of induction.

To obtain the electromagnetic wave equation in a vacuum using the modern method, we begin with the modern 'Heaviside' form of Maxwell's equations. In a vacuum and charge free space, these equations are:

$$\nabla \cdot \mathbf{E} = 0$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\nabla \cdot \mathbf{B} = 0$$

$$\nabla \times \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

Taking the curl of the curl equations gives:

$$\nabla \times \nabla \times \mathbf{E} = -\frac{\partial}{\partial t} \nabla \times \mathbf{B} = -\mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}$$

$$\nabla \times \nabla \times \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial}{\partial t} \nabla \times \mathbf{E} = -\mu_o \varepsilon_o \frac{\partial^2 \mathbf{B}}{\partial t^2}$$

By using the vector identity

$$\nabla \times (\nabla \times \mathbf{V}) = \nabla (\nabla \cdot \mathbf{V}) - \nabla^2 \mathbf{V}$$

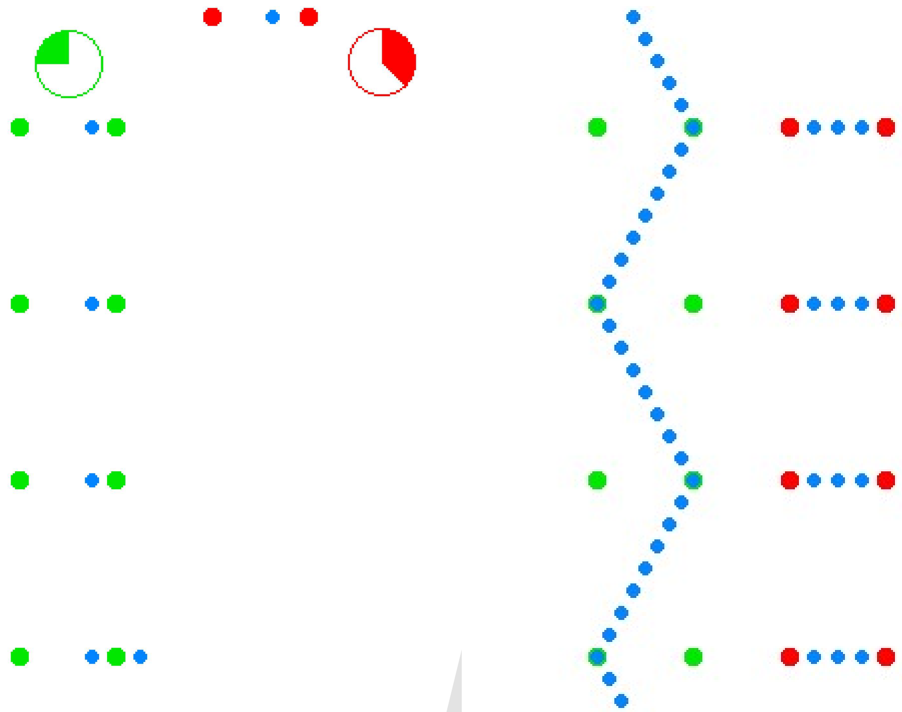
where $\mathbf{V}$ is any vector function of space, it turns into the wave equations:

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} - {c_0}^2 \cdot \nabla^2 \mathbf{E} = 0$$

$$\frac{\partial^2 \mathbf{B}}{\partial t^2} - {c_0}^2 \cdot \nabla^2 \mathbf{B} = 0$$

where $c_0 = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = 2.99792458 \times 10^8$ m/s is the speed of light in free space.

Covariant form of the homogeneous wave equation



Time dilation in transversal motion. The requirement that the speed of light is constant in every inertial reference frame leads to the theory of Special Relativity

These relativistic equations can be written in contravariant form as

$$\Box A^{\mu} = 0$$

where the electromagnetic four-potential is

$$A^{\mu} = (\varphi, \mathbf{A}c)$$

with the Lorenz gauge condition:

$$\partial_{\mu} A^{\mu} = 0.$$

Where

$$\Box = \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$$ is the d'Alembertian operator. (The square box is not a typographical error; it is the correct symbol for this operator.)

Homogeneous wave equation in curved spacetime

The electromagnetic wave equation is modified in two ways, the derivative is replaced with the covariant derivative and a new term that depends on the curvature appears.

$$-A^{\alpha;\beta}{}_{;\beta} + R^{\alpha}{}_{\beta} A^{\beta} = 0$$

where $R^{\alpha}{}_{\beta}$ is the Ricci curvature tensor and the semicolon indicates covariant differentiation.

The generalization of the Lorenz gauge condition in curved spacetime is assumed:

$$A^{\mu}{}_{;\mu} = 0.$$

Inhomogeneous electromagnetic wave equation

Localized time-varying charge and current densities can act as sources of electromagnetic waves in a vacuum. Maxwell's equations can be written in the form of a wave equation with sources. The addition of sources to the wave equations makes the partial differential equations inhomogeneous.

Solutions to the homogeneous electromagnetic wave equation

The general solution to the electromagnetic wave equation is a linear superposition of waves of the form

$$\mathbf{E}(\mathbf{r}, t) = g(\phi(\mathbf{r}, t)) = g(\omega t - \mathbf{k} \cdot \mathbf{r})$$

and

$$\mathbf{B}(\mathbf{r}, t) = g(\phi(\mathbf{r}, t)) = g(\omega t - \mathbf{k} \cdot \mathbf{r})$$

for virtually *any* well-behaved function *g* of dimensionless argument φ, where

ω is the angular frequency (in radians per second), and
$\mathbf{k} = (k_x, k_y, k_z)$ is the wave vector (in radians per meter).

Although the function *g* can be and often is a monochromatic sine wave, it does not have to be sinusoidal, or even periodic. In practice, *g* cannot have infinite periodicity because any real electromagnetic wave must always have a finite extent in time and space. As a result, and based on the theory of Fourier decomposition, a real wave must consist of the superposition of an infinite set of sinusoidal frequencies.

In addition, for a valid solution, the wave vector and the angular frequency are not independent; they must adhere to the dispersion relation:

$$k = |\mathbf{k}| = \frac{\omega}{c} = \frac{2\pi}{\lambda}$$

where *k* is the wavenumber and λ is the wavelength.

Monochromatic, sinusoidal steady-state

The simplest set of solutions to the wave equation result from assuming sinusoidal waveforms of a single frequency in separable form:

$$\mathbf{E}(\mathbf{r}, t) = \mathrm{Re}\{\mathbf{E}(\mathbf{r})e^{i\omega t}\}$$

where

- i is the imaginary unit,
- $\omega = 2\pi f$ **is the angular frequency in radians per second,**
- f is the **frequency in hertz, and**
- $e^{i\omega t} = \cos(\omega t) + i\sin(\omega t)$ is Euler's formula.

Plane wave solutions

Consider a plane defined by a unit normal vector

$$\mathbf{n} = \frac{\mathbf{k}}{k}.$$

Then planar traveling wave solutions of the wave equations are

$$\mathbf{E}(\mathbf{r}) = E_0 e^{-i\mathbf{k}\cdot\mathbf{r}}$$

and

$$\mathbf{B}(\mathbf{r}) = B_0 e^{-i\mathbf{k}\cdot\mathbf{r}}$$

where

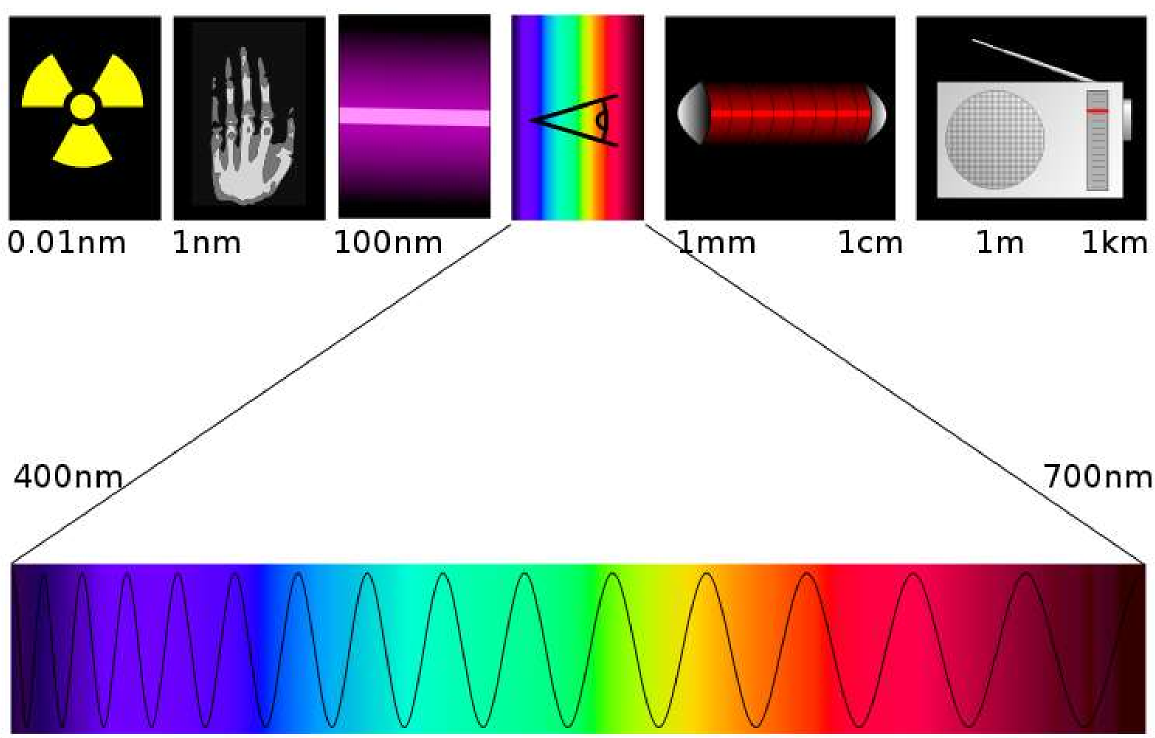
$\mathbf{r} = (x, y, z)$ is the position vector (in meters).

These solutions represent planar waves traveling in the direction of the normal vector $\mathbf{n}$. If we define the z direction as the direction of $\mathbf{n}$ and the x direction as the direction of $\mathbf{E}$, then by Faraday's Law the magnetic field lies in the y direction and is related to the electric field by the relation $c^2 \frac{\partial B}{\partial z} = \frac{\partial E}{\partial t}$. Because the divergence of the electric and magnetic fields are zero, there are no fields in the direction of propagation.

This solution is the linearly polarized solution of the wave equations. There are also circularly polarized solutions in which the fields rotate about the normal vector.

Spectral decomposition

Because of the linearity of Maxwell's equations in a vacuum, solutions can be decomposed into a superposition of sinusoids. This is the basis for the Fourier transform method for the solution of differential equations. The sinusoidal solution to the electromagnetic wave equation takes the form



Electromagnetic spectrum illustration

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \cos(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi_0)$$

and

$$\mathbf{B}(\mathbf{r}, t) = \mathbf{B}_0 \cos(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi_0)$$

where

t is time (in seconds),
ω is the angular frequency (in radians per second),
$\mathbf{k} = (k_x, k_y, k_z)$ is the wave vector (in radians per meter), and
ϕ_0 is the phase angle (in radians).

The wave vector is related to the angular frequency by

$$k = |\mathbf{k}| = \frac{\omega}{c} = \frac{2\pi}{\lambda}$$

where k is the wavenumber and λ is the wavelength.

The electromagnetic spectrum is a plot of the field magnitudes (or energies) as a function of wavelength.

Other solutions

Spherically symmetric and cylindrically symmetric analytic solutions to the electromagnetic wave equations are also possible.

In cylindrical coordinates the wave equation can be written as follows:

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{s}\mathbf{E}_0 \cos(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi_0)$$

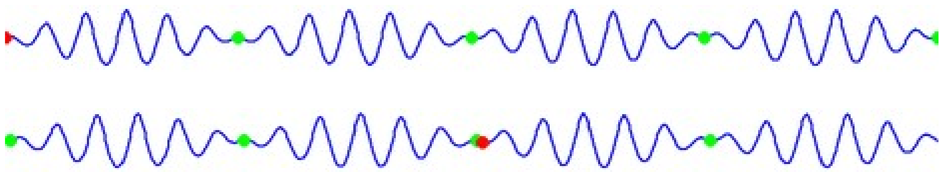
and

$$\mathbf{B}(\mathbf{r}, t) = \frac{1}{s}\mathbf{B}_0 \cos(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi_0)$$

Chapter 2

Phase Velocity and Group Velocity

Phase Velocity



Frequency dispersion in groups of gravity waves on the surface of deep water. The red dot moves with the phase velocity, and the green dots propagate with the group velocity.

The **phase velocity** of a wave is the rate at which the phase of the wave propagates in space. This is the speed at which the phase of any one frequency component of the wave travels. For such a component, any given phase of the wave (for example, the crest) will appear to travel at the phase velocity. The phase velocity is given in terms of the wavelength λ (lambda) and period T as

$$v_{\mathrm{p}} = \frac{\lambda}{T}.$$

Or, equivalently, in terms of the wave's angular frequency ω and wavenumber k by

$$v_{\mathrm{p}} = \frac{\omega}{k}.$$

In a dispersive medium, the phase velocity varies with frequency and is not necessarily the same as the group velocity of the wave, which is the rate that changes in amplitude (known as the *envelope* of the wave) propagate.

The phase velocity of electromagnetic radiation may, under certain circumstances, (for example anomalous dispersion) exceed the speed of light in a vacuum, but this does not indicate any superluminal information or energy transfer. It was theoretically described by physicists such as Arnold Sommerfeld and Léon Brillouin.

Matter wave phase

In quantum mechanics, particles also behave as waves with complex phases. By the de Broglie hypothesis, we see that

$$v_{\mathrm{p}} = \frac{\omega}{k} = \frac{E/\hbar}{p/\hbar} = \frac{E}{p}.$$

Using relativistic relations for energy and momentum, we have

$$v_{\mathrm{p}} = \frac{E}{p} = \frac{\gamma mc^2}{\gamma mv} = \frac{c^2}{v} = \frac{c}{\beta}$$

where E is the total energy of the particle (i.e. rest energy plus kinetic energy in kinematic sense), p the momentum, γ the Lorentz factor, c the speed of light, and β the speed as a fraction of c. The variable v can either be taken to be the speed of the particle or the group velocity of the corresponding matter wave. Since the particle speed $v < c$ for any particle that has mass (according to special relativity), the phase velocity of matter waves always exceeds c, i.e.

$$v_{\mathrm{p}} > c,$$

and as we can see, it approaches c when the particle speed is in the relativistic range. The superluminal phase velocity does not violate special relativity, as it carries no information.

Group Velocity

The **group velocity** of a wave is the velocity with which the overall shape of the wave's amplitudes — known as the *modulation* or *envelope* of the wave — propagates through space.

For example, imagine what happens if a stone is thrown into the middle of a very still pond. When the stone hits the surface of the water, a circular pattern of waves appears. It soon turns into a circular ring of waves with a quiescent center. The ever expanding ring of waves is the **wave group**, within which one can discern individual wavelets of differing wavelengths traveling at different speeds. The longer waves travel faster than the group as a whole, but they die out as they approach the leading edge. The shorter waves travel slower and they die out as they emerge from the trailing boundary of the group.

Definition and interpretation

Definition

The group velocity v_g is defined by the equation

$$v_g \equiv \frac{\partial \omega}{\partial k}$$

where:

ω is the wave's angular frequency;
k is the wave number.

The function $\omega(k)$, which gives ω as a function of k, is known as the dispersion relation. If ω is directly proportional to k, then the group velocity is exactly equal to the phase velocity. Otherwise, the envelope of the wave will become distorted as it propagates. This "group velocity dispersion" is an important effect in the propagation of signals through optical fibers and in the design of high-power, short-pulse lasers.

Note: The above definition of group velocity is only useful for wavepackets, which is a pulse that is localized in both real space and frequency space. Because waves at different frequencies propagate at differing phase velocities in dispersive media, for a large frequency range (a narrow envelope in space) the observed pulse would change shape while traveling, making group velocity an unclear or useless quantity.

Physical interpretation

The group velocity is often thought of as the velocity at which energy or information is conveyed along a wave. In most cases this is accurate, and the group velocity can be thought of as the signal velocity of the waveform. However, if the wave is travelling through an absorptive medium, this does not always hold. Since the 1980s, various experiments have verified that it is possible for the group velocity of laser light pulses sent through specially prepared materials to significantly exceed the speed of light in vacuum. However, superluminal communication is not possible in this case, since the signal velocity remains less than the speed of light. It is also possible to reduce the group velocity to zero, stopping the pulse, or have negative group velocity, making the pulse appear to propagate backwards. However, in all these cases, photons continue to propagate at the expected speed of light in the medium.

Anomalous dispersion happens in areas of rapid spectral variation with respect to the refractive index. Therefore, negative values of the group velocity will occur in these areas. Anomalous dispersion plays a fundamental role in achieving backward propagating and superluminal light. Anomalous dispersion can also be used to produce group and phase velocities that are in different directions. Materials that exhibit large anomalous dispersion allow the group velocity of the light to exceed c and/or become negative.

History

The idea of a group velocity distinct from a wave's phase velocity was first proposed by W.R. Hamilton in 1839, and the first full treatment was by Rayleigh in his "Theory of Sound" in 1877.

Matter-wave group velocity

Albert Einstein first explained the wave–particle duality of light in 1905. Louis de Broglie hypothesized that any particle should also exhibit such a duality. The velocity of a particle, he concluded then (but may be questioned today, see above), should always equal the group velocity of the corresponding wave. De Broglie deduced that if the duality equations already known for light were the same for any particle, then his hypothesis would hold. This means that

$$v_g = \frac{\partial \omega}{\partial k} = \frac{\partial (E/\hbar)}{\partial (p/\hbar)} = \frac{\partial E}{\partial p}$$

where

E is the total energy of the particle,
p is its momentum,
$\hbar$ is the reduced Planck constant.

For a free non-relativistic particle it follows that

$$\begin{aligned} v_g &= \frac{\partial E}{\partial p} = \frac{\partial}{\partial p}\left(\frac{1}{2}\frac{p^2}{m}\right) \\ &= \frac{p}{m} \\ &= v. \end{aligned}$$

where

m is the mass of the particle and
v its velocity.

Also in special relativity we find that

$$\begin{aligned} v_g = \frac{\partial E}{\partial p} &= \frac{\partial}{\partial p}\left(\sqrt{p^2c^2 + m^2c^4}\right) \\ &= \frac{pc^2}{\sqrt{p^2c^2 + m^2c^4}} \\ &= \frac{p}{m\sqrt{(p/(mc))^2 + 1}} \\ &= \frac{p}{m\gamma} \\ &= \frac{mv\gamma}{m\gamma} \\ &= v. \end{aligned}$$

where

> *m* is the rest mass of the particle,
> *c* is the speed of light in a vacuum,
> γ is the Lorentz factor.
> and *v* is the velocity of the particle regardless of wave behavior.

Group velocity (equal to an electron's speed) should not be confused with phase velocity (equal to the product of the electron's frequency multiplied by its wavelength).

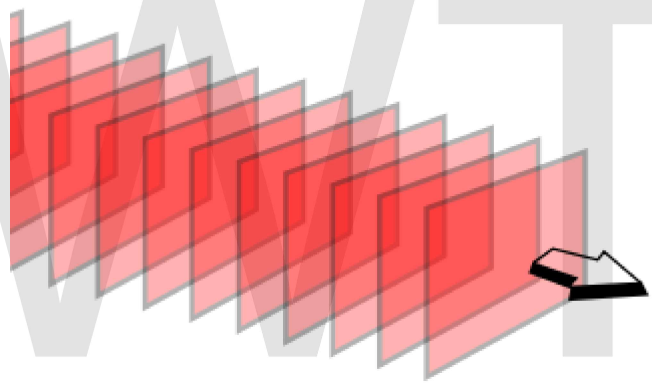
Both in relativistic and non-relativistic quantum physics, we can identify the group velocity of a particle's wave function with the particle velocity. Quantum mechanics has very accurately demonstrated this hypothesis, and the relation has been shown explicitly for particles as large as molecules.

Chapter 3

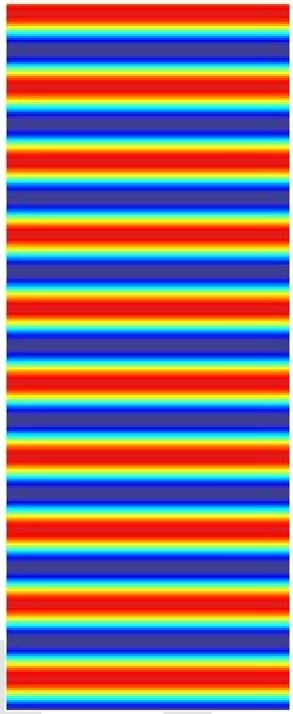
Plane Waves and Standing Waves

Plane Waves

In the physics of wave propagation, a **plane wave** (also spelled **planewave**) is a constant-frequency wave whose wavefronts (surfaces of constant phase) are infinite parallel planes of constant amplitude normal to the phase velocity vector.



The wavefronts of a plane wave traveling in 3-space



A pseudo-colored upward-traveling plane wave in 2-space

By extension, the term is also used to describe waves that are approximately plane waves in a localized region of space. For example, a localized source such as an antenna produces a field that is approximately a plane wave in its far-field region. Equivalently, for propagation in a homogeneous medium over lengthscales much longer than the wavelength, the "rays" in the limit where ray optics is valid correspond locally to approximate plane waves.

Mathematically, a plane wave is a wave of the following form:

$$u(\mathbf{x},t) = Ae^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$$

where i is the imaginary unit, **k** is the wave vector, ω is the angular frequency, and A is the (complex) amplitude. This form of the plane wave uses the physics time convention; in the engineering time convention, $-j$ is used instead of $+i$ in the exponent. The physical solution is found by taking the real part of this expression:

$$Re[u(\mathbf{x},t)] = |A|\cos(\mathbf{k}\cdot\mathbf{x} - \omega t + \arg A)$$

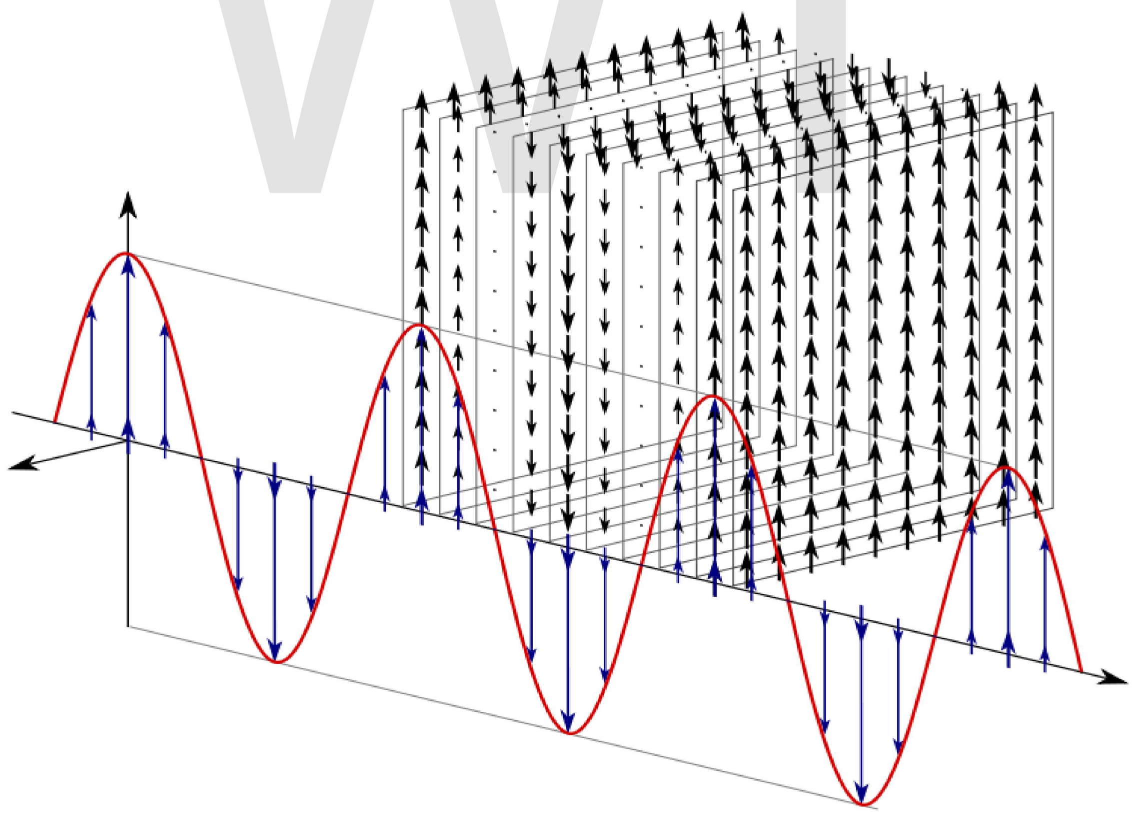
This is the solution for a scalar wave equation in a homogeneous medium. For vector wave equations, such as the ones describing electromagnetic radiation or waves in an elastic solid, the solution for a homogeneous medium is similar: the *scalar* amplitude A is replaced by a constant *vector* **A**. For example, in electromagnetism **A** is typically the vector for the electric field, magnetic field, or vector potential. A transverse wave is one in which the amplitude vector is orthogonal to **k**, which is the case for electromagnetic waves in an isotropic medium. By contrast, a longitudinal wave is one in which the amplitude vector is parallel to **k**, such as for acoustic waves in a gas or fluid.

In this equation, the function $\omega(\mathbf{k})$ is the dispersion relation of the medium, with the ratio $\omega/|\mathbf{k}|$ giving the magnitude of the phase velocity and $d\omega/d\mathbf{k}$ giving the group velocity. For electromagnetism in an isotropic medium with index of refraction n, the phase velocity is c/n, which equals the group velocity only if the index is not frequency-dependent.

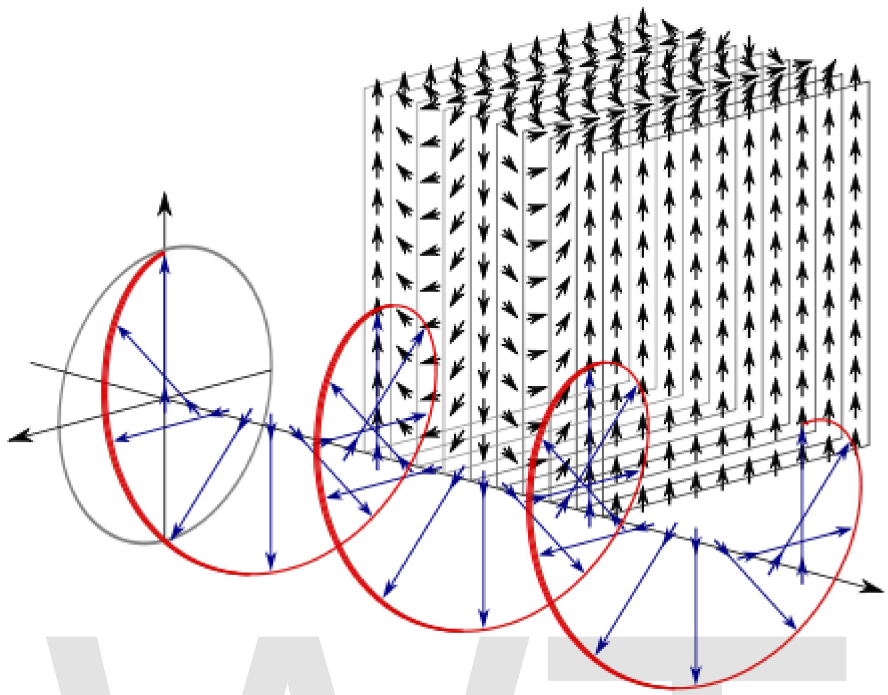
Generally, a wave solution can be expressed as a superposition of plane waves. This approach is known as the Angular spectrum method. The form of the planewave solution is actually a general consequence of translational symmetry. More generally, for periodic structures having discrete translational symmetry, the solutions take the form of Bloch waves, most famously in crystalline atomic materials but also in photonic crystals and other periodic wave equations. As another generalization, for structures that are only uniform along one direction x (such as a waveguide along the x direction), the solutions (waveguide modes) are of the form $\exp[i(kx-\omega t)]$ multiplied by some amplitude function $a(y,z)$. This is a special case of a separable partial differential equation.

The term is used in the same way for telecommunication, e.g. in Federal Standard 1037C and MIL-STD-188.

Polarized electromagnetic plane waves



Linearly polarized light



Circularly polarized light

The blocks of vectors represent how the magnitude and direction of the electric field is constant for an entire plane perpendicular to the direction of travel.

Represented in the first illustration toward the right is a linearly polarized, electromagnetic wave. Because this is a plane wave, each blue vector, indicating the perpendicular displacement from a point on the axis out to the sine wave, represents the magnitude and direction of the electric field for an entire plane that is perpendicular to the axis.

Represented in the second illustration is a circularly polarized, electromagnetic plane wave. Each blue vector indicating the perpendicular displacement from a point on the axis out to the helix, also represents the magnitude and direction of the electric field for an entire plane perpendicular to the axis.

In both illustrations, along the axes is a series of shorter blue vectors which are scaled down versions of the longer blue vectors. These shorter blue vectors are extrapolated out into the block of black vectors which fill a volume of space. Notice that for a given plane, the black vectors are identical, indicating that the magnitude and direction of the electric field is constant along that plane.

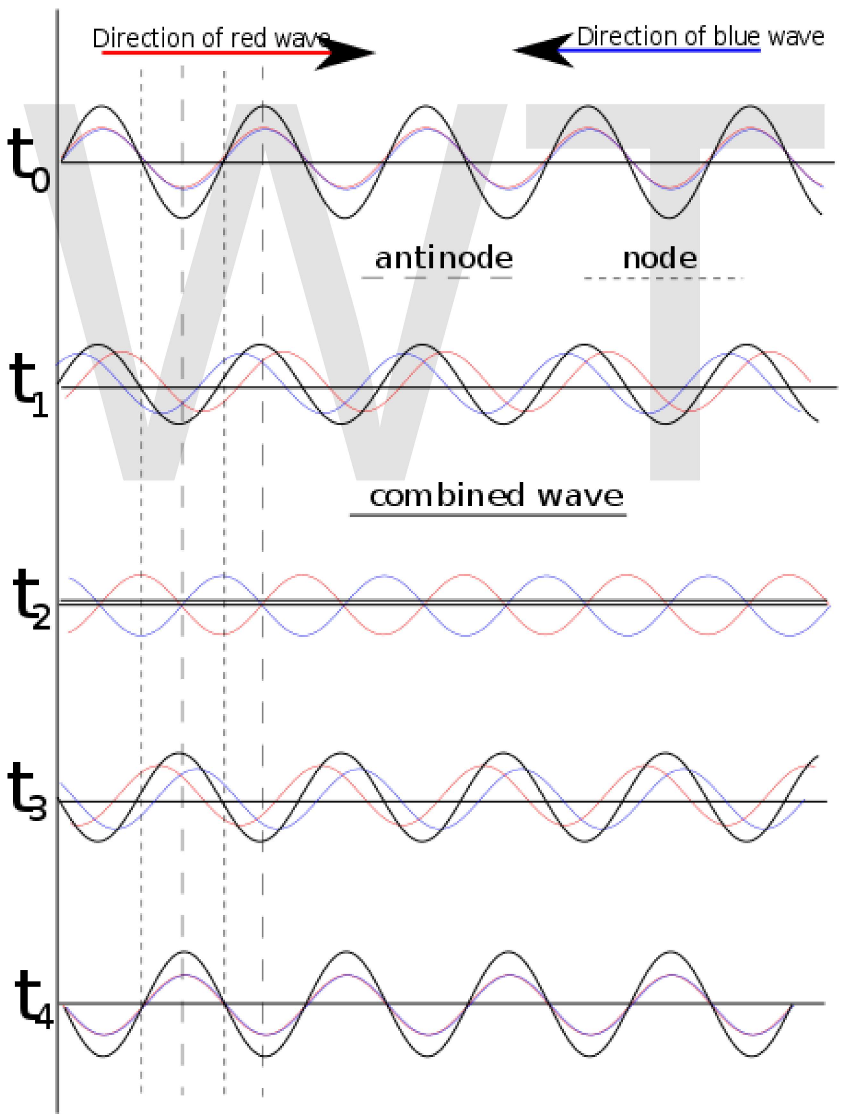
In the case of the linearly polarized light, the field strength from plane to plane varies from a maximum in one direction, down to zero, and then back up to a maximum in the opposite direction.

In the case of the circularly polarized light, the field strength remains constant from plane to plane but its direction steadily changes in a rotary type manner.

Not indicated in either illustration is the electric field's corresponding magnetic field which is proportional in strength to the electric field at each point in space but is at a right angle to it. Illustrations of the magnetic field vectors would be virtually identical to these except all the vectors would be rotated 90 degrees perpendicular to the direction of propagation.

Standing Waves

A **standing wave**, also known as a **stationary wave**, is a wave that remains in a constant position.



Two opposing waves combine to form a standing wave

This phenomenon can occur because the medium is moving in the opposite direction to the wave, or it can arise in a stationary medium as a result of interference between two waves traveling in opposite directions. In the second case, for waves of equal amplitude traveling in opposing directions, there is on average no net propagation of energy.

Standing waves in resonators are one of the causes of the phenomenon known as resonance.

Moving medium

As an example of the first type, under certain meteorological conditions standing waves form in the atmosphere in the lee of mountain ranges. Such waves are often exploited by glider pilots.

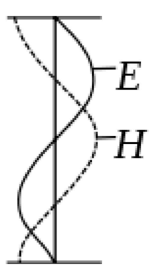
Standing waves and hydraulic jumps also form on fast flowing river rapids and tidal currents such as the Saltstraumen maelstrom. Many standing river waves are popular river surfing breaks.

Opposing waves

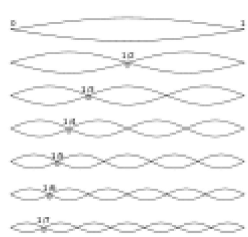
Standing waves

Standing wave in stationary medium. The red dots represent the wave nodes.

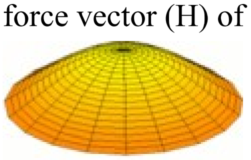
A standing wave (black) depicted as the sum of two propagating waves traveling in opposite directions (red and blue).

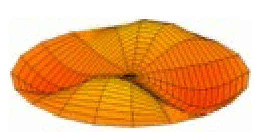


Electric force vector (E) and magnetic force vector (H) of a standing wave.



Standing waves in a string — the fundamental mode and the first 6 overtones.





A two-dimensional standing wave on a disk; this is the fundamental mode

A higher harmonic standing wave on a disk with two nodal lines crossing at the center.

As an example of the second type, a *standing wave* in a transmission line is a wave in which the distribution of current, voltage, or field strength is formed by the superposition of two waves of the same frequency propagating in opposite directions. The effect is a series of nodes (zero displacement) and anti-nodes (maximum displacement) at fixed points along the transmission line. Such a standing wave may be formed when a wave is transmitted into one end of a transmission line and is reflected from the other end by an impedance mismatch, *i.e.*, discontinuity, such as an open circuit or a short. The failure of the line to transfer power at the standing wave frequency will usually result in attenuation distortion.

Another example is standing waves in the open ocean formed by waves with the same wave period moving in opposite directions. These may form near storm centres, or from reflection of a swell at the shore, and are the source of microbaroms and microseisms.

In practice, losses in the transmission line and other components mean that a perfect reflection and a pure standing wave are never achieved. The result is a *partial standing wave*, which is a superposition of a standing wave and a traveling wave. The degree to which the wave resembles either a pure standing wave or a pure traveling wave is measured by the standing wave ratio (SWR).

Mathematical description

In one dimension, two waves with the same frequency, wavelength and amplitude traveling in opposite directions will interfere and produce a standing wave or stationary wave. For example: a wave traveling to the right along a taut string and hitting the end will reflect back in the other direction along the string, and the two waves will superpose to produce a standing wave. The reflective wave has to have the same amplitude and frequency as the incoming wave.

If the string is held at both ends, forcing zero movement at the ends, the ends become zeroes or *nodes* of the wave. The length of the string then becomes a measure of which waves the string will entertain: the longest wavelength is called the *fundamental*. Half a wavelength of the fundamental fits on the string. Shorter wavelengths also can be supported as long as multiples of half a wavelength fit on the string. The frequencies of these waves all are multiples of the fundamental, and are called *harmonics* or *overtones*. For example, a guitar player can select an overtone by putting a finger on a string to force a node at the proper position between the ends of the string, suppressing all harmonics that do not share this node.

Harmonic waves travelling in opposite directions can be represented by the equations below:

$$y_1 \; = \; y_0 \, \sin(kx - \omega t)$$

and

$$y_2 \; = \; y_0 \, \sin(kx + \omega t)$$

where:

- y_0 is the amplitude of the wave,
- ω (called angular frequency, measured in *radians per second*) is 2π times the frequency (in *hertz*),
- k (called the wave number and measured in *radians per metre*) is 2π divided by the wavelength λ (in *metres*), and
- x and t are variables for longitudinal position and time, respectively.

So the resultant wave y equation will be the sum of y_1 and y_2:

$$y \; = \; y_0 \, \sin(kx - \omega t) \; + \; y_0 \, \sin(kx + \omega t).$$

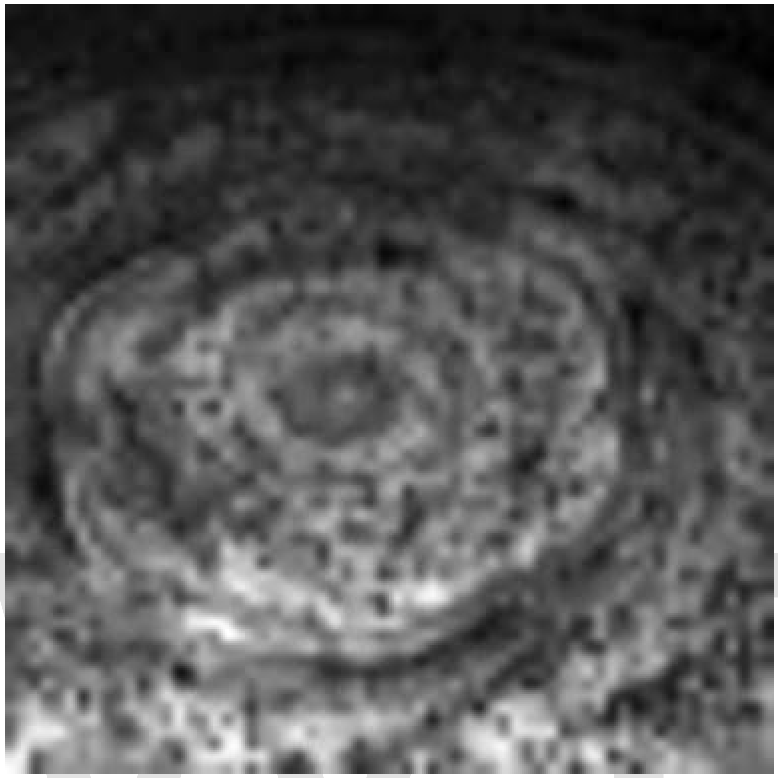
Using a trigonometric identity (the 'Sum to Product' identity for 'sin(u)+sin(v)') to simplify:

$$y \; = \; 2\, y_0 \, \cos(\omega t) \; \sin(kx).$$

This describes a wave that oscillates in time, but has a spatial dependence that is stationary: *sin(kx)*. At locations $x = 0, \lambda/2, \lambda, 3\lambda/2, ...$ called the nodes the amplitude is always zero, whereas at locations $x = \lambda/4, 3\lambda/4, 5\lambda/4, ...$ called the anti-nodes, the amplitude is maximum. The distance between two conjugative nodes or anti-nodes is $\lambda/2$.

Standing waves can also occur in two- or three-dimensional resonators. With standing waves on two dimensional membranes such as drumheads, illustrated in the animations above, the nodes become nodal lines, lines on the surface at which there is no movement, that separate regions vibrating with opposite phase. These nodal line patterns are called Chladni figures. In three-dimensional resonators, such as musical instrument sound boxes and microwave cavity resonators, there are nodal surfaces.

Physical waves



The hexagonal cloud feature at the north pole of Saturn is thought to be some sort of standing wave pattern.

Standing waves are also observed in physical media such as strings and columns of air. Any waves traveling along the medium will reflect back when they reach the end. This effect is most noticeable in musical instruments where, at various multiples of a vibrating string or air column's natural frequency, a standing wave is created, allowing harmonics to be identified. Nodes occur at fixed ends and anti-nodes at open ends. If fixed at only one end, only odd-numbered harmonics are available. At the open end of a pipe the anti-node will not be exactly at the end as it is altered by its contact with the air and so end correction is used to place it exactly. The density of a string will affect the frequency at which harmonics will be produced; the greater the density the lower the frequency needs to be to produce a standing wave of the same harmonic.

Optical waves

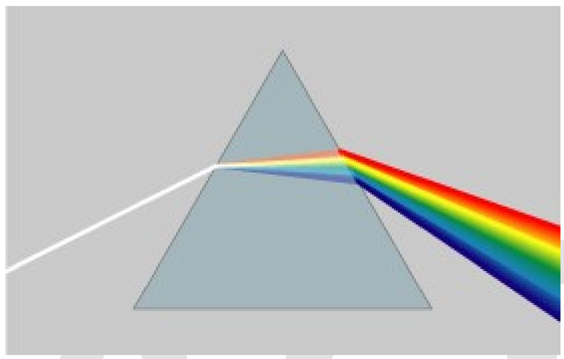
Standing waves are also observed in optical media such as optical wave guides, optical cavities, etc. In an optical cavity, the light wave from one end is made to reflect from the other. The transmitted and reflected waves superpose, and form a standing-wave pattern.

Mechanical waves

Standing waves can be mechanically induced into solid medium using resonance. One easy to understand example is two people shaking either end of a jump rope. If they shake in sync the rope will form a regular pattern with nodes and antinodes and appear to be stationary, hence the name standing wave. Similarly a cantilever beam can have a standing wave imposed on it by applying a base excitation. In this case the free end moves the greatest distance laterally compared to any location along the beam. Such a device can be used as a sensor to track changes in frequency or phase of the resonance of the fiber. One application is as a measurement device for dimensional metrology.

Chapter 4

Dispersion (optics)



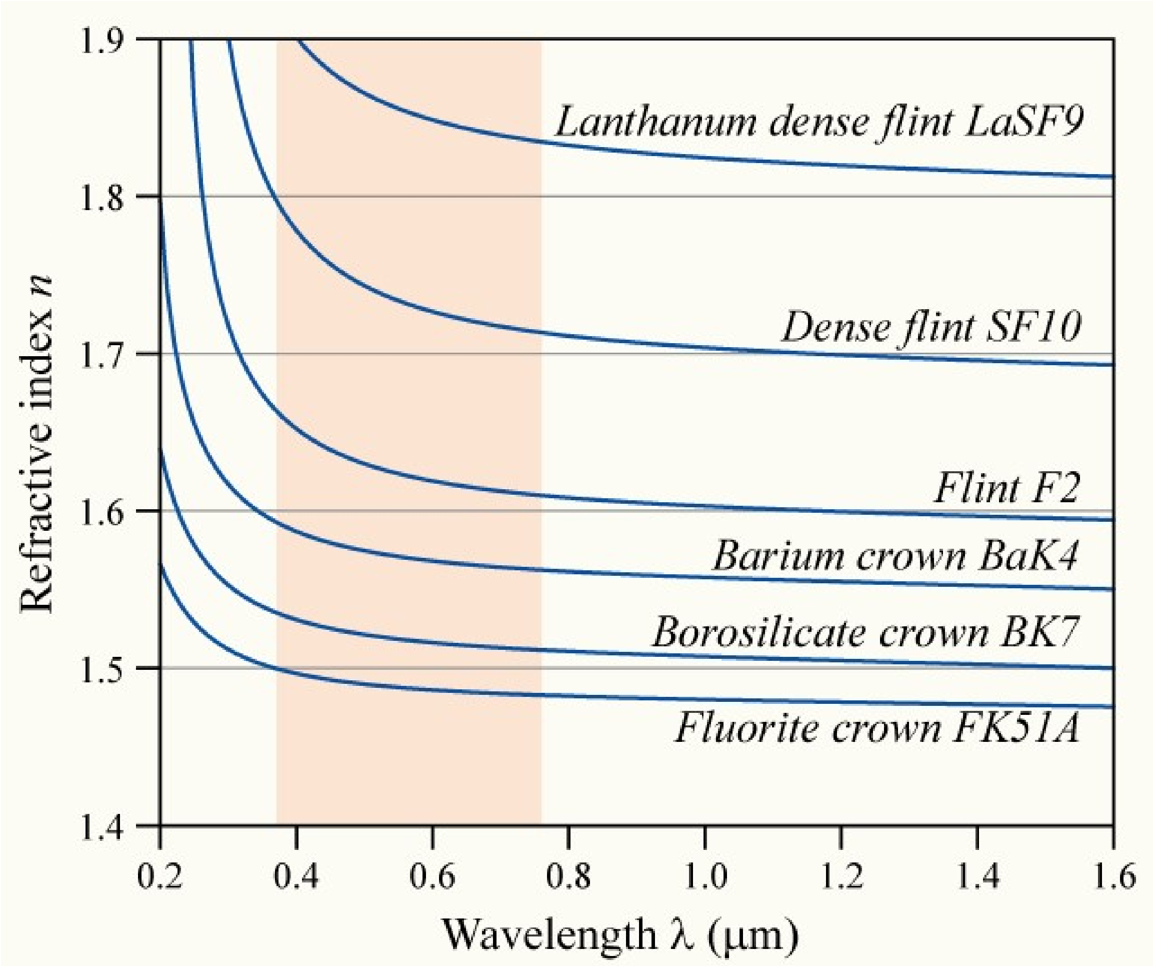
In a prism, material dispersion (a wavelength-dependent refractive index) causes different colors to refract at different angles, splitting white light into a rainbow.

In optics, **dispersion** is the phenomenon in which the phase velocity of a wave depends on its frequency, or alternatively when the group velocity depends on the frequency. Media having such a property are termed *dispersive media*. Dispersion is sometimes called ***chromatic*** **dispersion** to emphasize its wavelength-dependent nature, or **group-velocity dispersion** (**GVD**) to emphasize the role of the group velocity.

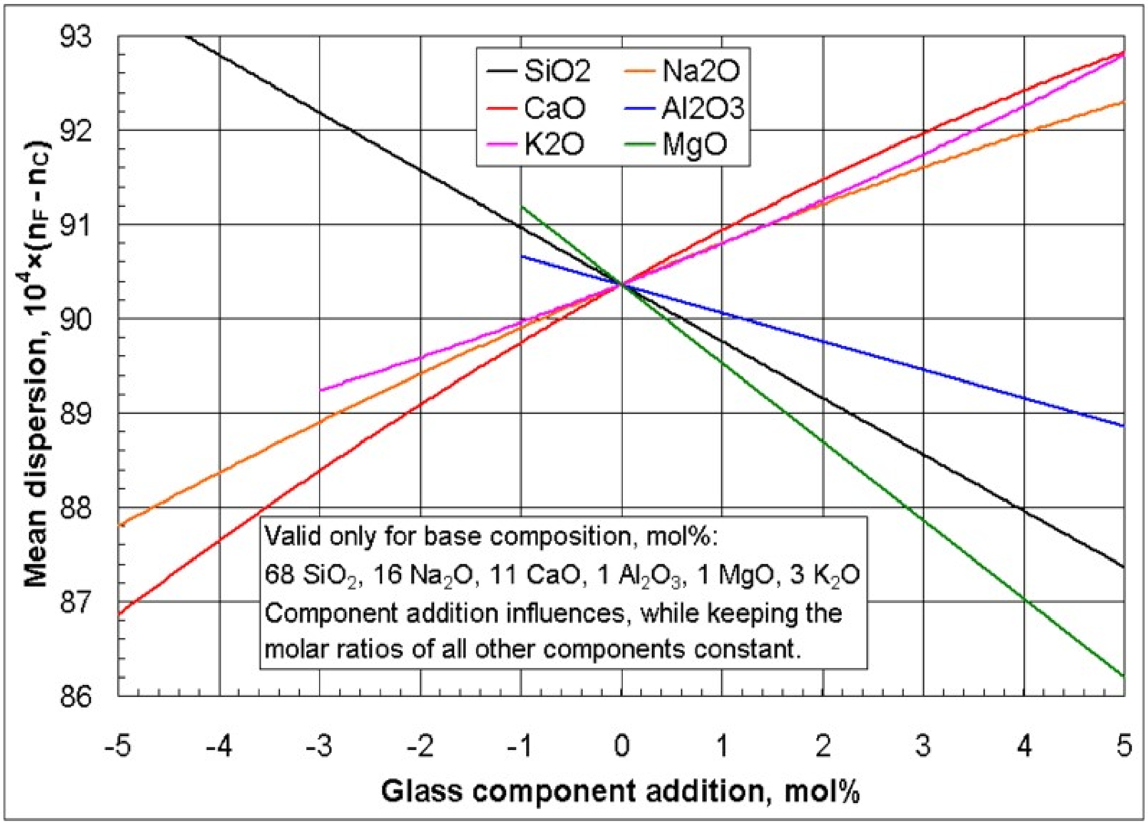
The most familiar example of dispersion is probably a rainbow, in which dispersion causes the spatial separation of a white light into components of different wavelengths (different colors). However, dispersion also has an effect in many other circumstances: for example, GVD causes pulses to spread in optical fibers, degrading signals over long distances; also, a cancellation between group-velocity dispersion and nonlinear effects leads to soliton waves. Dispersion is most often described for light waves, but it may occur for any kind of wave that interacts with a medium or passes through an inhomogeneous geometry (e.g., a waveguide), such as sound waves.

There are generally two sources of dispersion: **material dispersion** and **waveguide dispersion**. Material dispersion comes from a frequency-dependent response of a material to waves. For example, material dispersion leads to undesired chromatic aberration in a lens or the separation of colors in a prism. Waveguide dispersion occurs when the speed of a wave in a waveguide (such as an optical fiber) depends on its frequency for geometric reasons, independent of any frequency dependence of the materials from which it is constructed. More generally, "waveguide" dispersion can occur for waves propagating through any inhomogeneous structure (e.g., a photonic crystal), whether or not the waves are confined to some region. In general, *both* types of dispersion may be present, although they are not strictly additive. Their combination leads to signal degradation in optical fibers for telecommunications, because the varying delay in arrival time between different components of a signal "smears out" the signal in time.

Material dispersion in optics



The variation of refractive index vs. wavelength for various glasses. The wavelengths of visible light are shaded in red.



Influences of selected glass component additions on the mean dispersion of a specific base glass (n_F valid for λ = 486 nm (blue), n_C valid for λ = 656 nm (red))

Material dispersion can be a desirable or undesirable effect in optical applications. The dispersion of light by glass prisms is used to construct spectrometers and spectroradiometers. Holographic gratings are also used, as they allow more accurate discrimination of wavelengths. However, in lenses, dispersion causes chromatic aberration, an undesired effect that may degrade images in microscopes, telescopes and photographic objectives.

The *phase velocity*, *v*, of a wave in a given uniform medium is given by

$$v = \frac{c}{n}$$

where *c* is the speed of light in a vacuum and *n* is the refractive index of the medium.

In general, the refractive index is some function of the frequency *f* of the light, thus $n = n(f)$, or alternatively, with respect to the wave's wavelength $n = n(\lambda)$. The wavelength dependence of a material's refractive index is usually quantified by an empirical formula, the Cauchy or Sellmeier equations.

Because of the Kramers–Kronig relations, the wavelength dependence of the real part of the refractive index is related to the material absorption, described by the imaginary part of the refractive index (also called the extinction coefficient). In particular, for non-magnetic materials ($\mu = \mu_0$), the susceptibility χ that appears in the Kramers–Kronig relations is the electric susceptibility $\chi_e = n^2 - 1$.

The most commonly seen consequence of dispersion in optics is the separation of white light into a color spectrum by a prism. From Snell's law it can be seen that the angle of refraction of light in a prism depends on the refractive index of the prism material. Since that refractive index varies with wavelength, it follows that the angle that the light is refracted by will also vary with wavelength, causing an angular separation of the colors known as *angular dispersion*.

For visible light, most transparent materials (e.g., glasses) have:

$$1 < n(\lambda_{\text{red}}) < n(\lambda_{\text{yellow}}) < n(\lambda_{\text{blue}}) \ ,$$

or alternatively:

$$\frac{\mathrm{d}n}{\mathrm{d}\lambda} < 0,$$

that is, refractive index n decreases with increasing wavelength λ. In this case, the medium is said to have *normal dispersion*. Whereas, if the index increases with increasing wavelength the medium has *anomalous dispersion*.

At the interface of such a material with air or vacuum (index of ~1), Snell's law predicts that light incident at an angle θ to the normal will be refracted at an angle $\arcsin(\sin(\theta)/n)$. Thus, blue light, with a higher refractive index, will be bent more strongly than red light, resulting in the well-known rainbow pattern.

Group and phase velocity

Another consequence of dispersion manifests itself as a temporal effect. The formula $v = c / n$ calculates the *phase velocity* of a wave; this is the velocity at which the *phase* of any one frequency component of the wave will propagate. This is not the same as the *group velocity* of the wave, that is the rate at which changes in amplitude (known as the *envelope* of the wave) will propagate. For a homogeneous medium, the group velocity v_g is related to the phase velocity by (here λ is the wavelength in vacuum, not in the medium):

$$v_g = c\left(n - \lambda\frac{dn}{d\lambda}\right)^{-1}.$$

The group velocity v_g is often thought of as the velocity at which energy or information is conveyed along the wave. In most cases this is true, and the group velocity can be thought of as the *signal velocity* of the waveform. In some unusual circumstances, called cases of anomalous dispersion, the rate of change of the index of refraction with respect to the wavelength changes sign, in which case it is possible for the group velocity to exceed the speed of light ($v_g > c$). Anomalous dispersion occurs, for instance, where the wavelength of the light is close to an absorption resonance of the medium. When the dispersion is anomalous, however, group velocity is no longer an indicator of signal velocity. Instead, a signal travels at the speed of the wavefront, which is c irrespective of the index of refraction. Recently, it has become possible to create gases in which the group velocity is not only larger than the speed of light, but even negative. In these cases, a pulse can appear to exit a medium before it enters. Even in these cases, however, a signal travels at, or less than, the speed of light, as demonstrated by Stenner, et al.

The group velocity itself is usually a function of the wave's frequency. This results in **group velocity dispersion** (GVD), which causes a short pulse of light to spread in time as a result of different frequency components of the pulse travelling at different velocities. GVD is often quantified as the *group delay dispersion parameter* (again, this formula is for a uniform medium only):

$$D = -\frac{\lambda}{c}\,\frac{d^2n}{d\lambda^2}.$$

If D is less than zero, the medium is said to have *positive dispersion*. If D is greater than zero, the medium has *negative dispersion*. If a light pulse is propagated through a normally dispersive medium, the result is the higher frequency components travel slower than the lower frequency components. The pulse therefore becomes *positively chirped*, or *up-chirped*, increasing in frequency with time. Conversely, if a pulse travels through an anomalously dispersive medium, high frequency components travel faster than the lower ones, and the pulse becomes *negatively chirped*, or *down-chirped*, decreasing in frequency with time.

The result of GVD, whether negative or positive, is ultimately temporal spreading of the pulse. This makes dispersion management extremely important in optical communications systems based on optical fiber, since if dispersion is too high, a group of pulses representing a bit-stream will spread in time and merge together, rendering the bit-stream unintelligible. This limits the length of fiber that a signal can be sent down without regeneration. One possible answer to this problem is to send signals down the optical fibre at a wavelength where the GVD is zero (e.g., around 1.3–1.5 μm in silica fibres), so pulses at this wavelength suffer minimal spreading from dispersion—in practice, however, this approach causes more problems than it solves because zero GVD unacceptably amplifies other nonlinear effects (such as four wave mixing). Another possible option is to use soliton pulses in the regime of anomalous dispersion, a form of optical pulse which uses a nonlinear optical effect to self-maintain its shape—solitons have the practical problem, however, that they require a certain power level to be maintained in the pulse for the nonlinear effect to be of the correct strength. Instead, the solution that is currently used in practice is to perform dispersion compensation, typically by matching the fiber with another fiber of opposite-sign dispersion so that the dispersion effects cancel; such compensation is ultimately limited by nonlinear effects such as self-phase modulation, which interact with dispersion to make it very difficult to undo.

Dispersion control is also important in lasers that produce short pulses. The overall dispersion of the optical resonator is a major factor in determining the duration of the pulses emitted by the laser. A pair of prisms can be arranged to produce net negative dispersion, which can be used to balance the usually positive dispersion of the laser medium. Diffraction gratings can also be used to produce dispersive effects; these are often used in high-power laser amplifier systems. Recently, an alternative to prisms and gratings has been developed: chirped mirrors. These dielectric mirrors are coated so that different wavelengths have different penetration lengths, and therefore different group delays. The coating layers can be tailored to achieve a net negative dispersion.

Dispersion in waveguides

Optical fibers, which are used in telecommunications, are among the most abundant types of waveguides. Dispersion in these fibers is one of the limiting factors that determine how much data can be transported on a single fiber.

The transverse modes for waves confined laterally within a waveguide generally have different speeds (and field patterns) depending upon their frequency (that is, on the relative size of the wave, the wavelength) compared to the size of the waveguide.

In general, for a waveguide mode with an angular frequency $\omega(\beta)$ at a propagation constant β (so that the electromagnetic fields in the propagation direction *(z)* oscillate proportional to $e^{i(\beta z - \omega t)}$), the group-velocity dispersion parameter D is defined as:

$$D = -\frac{2\pi c}{\lambda^2}\frac{d^2\beta}{d\omega^2} = \frac{2\pi c}{v_g^2\lambda^2}\frac{dv_g}{d\omega}$$

where $\lambda = 2\pi c / \omega$ is the vacuum wavelength and $v_g = d\omega / d\beta$ is the group velocity. This formula generalizes the one in the previous section for homogeneous media, and includes both waveguide dispersion and material dispersion. The reason for defining the dispersion in this way is that $|D|$ is the (asymptotic) temporal pulse spreading Δt per unit bandwidth $\Delta\lambda$ per unit distance travelled, commonly reported in ps / nm km for optical fibers.

A similar effect due to a somewhat different phenomenon is modal dispersion, caused by a waveguide having multiple modes at a given frequency, each with a different speed. A special case of this is polarization mode dispersion (PMD), which comes from a superposition of two modes that travel at different speeds due to random imperfections that break the symmetry of the waveguide.

Higher-order dispersion over broad bandwidths

When a broad range of frequencies (a broad bandwidth) is present in a single wavepacket, such as in an ultrashort pulse or a chirped pulse or other forms of spread spectrum transmission, it may not be accurate to approximate the dispersion by a constant over the entire bandwidth, and more complex calculations are required to compute effects such as pulse spreading.

In particular, the dispersion parameter *D* defined above is obtained from only one derivative of the group velocity. Higher derivatives are known as *higher-order dispersion*. These terms are simply a Taylor series expansion of the dispersion relation $\beta(\omega)$ of the medium or waveguide around some particular frequency. Their effects can be computed via numerical evaluation of Fourier transforms of the waveform, via integration of higher-order slowly varying envelope approximations, by a split-step method (which can use the exact dispersion relation rather than a Taylor series), or by direct simulation of the full Maxwell's equations rather than an approximate envelope equation.

Dispersion in gemology

In the technical terminology of gemology, *dispersion* is the difference in the refractive index of a material at the B and G Fraunhofer wavelengths of 686.7 nm and 430.8 nm and is meant to express the degree to which a prism cut from the gemstone shows "fire", or color. Dispersion is a material property. Fire depends on the dispersion, the cut angles, the lighting environment, the refractive index, and the viewer.

Dispersion in imaging

In photographic and microscopic lenses, dispersion causes chromatic aberration, distorting the image, and various techniques have been developed to counteract it such as the use of multielement lenses with glasses with different dispersion characteristics: the net effect is to recombine (at least approximately) all colors.

Dispersion in pulsar timing

Pulsars are spinning neutron stars that emit pulses at very regular intervals ranging from milliseconds to seconds. Astronomers believe that the pulses are emitted simultaneously over a wide range of frequencies. However, as observed on Earth, the components of each pulse emitted at higher radio frequencies arrive before those emitted at lower frequencies. This dispersion occurs because of the ionised component of the interstellar medium, which makes the group velocity frequency dependent. The extra delay added at a frequency ν is

$$t = k_{\mathrm{DM}} \times \left(\frac{\mathrm{DM}}{\nu^2}\right)$$

where the dispersion constant k_{DM} is given by

$$k_{\mathrm{DM}} = \frac{e^2}{2\pi m_e c} \simeq 4.149 \mathrm{GHz}^2 \mathrm{pc}^{-1} \mathrm{cm}^3 \mathrm{ms},$$

and the dispersion measure *DM* is the free electron column density (total electron content) n_e integrated along the path traveled by the photon from the pulsar to the Earth, and is given by

$$\mathrm{DM} = \int_0^d n_e \, dl$$

with units of parsecs per cubic centimetre ($1\mathrm{pc/cm}^3 = 30.857\times10^{21}\ \mathrm{m}^{-2}$).

Typically for astronometric observations, this delay cannot be measured directly, since the emission time is unknown. What *can* be measured is the difference in arrival times at two different frequencies. The delay ΔT between a high frequency ν_{hi} and a low frequency ν_{lo} component of a pulse will be

$$\Delta t = k_{\mathrm{DM}} \times \mathrm{DM} \times \left(\frac{1}{\nu_{\mathrm{lo}}^2} - \frac{1}{\nu_{\mathrm{hi}}^2}\right)$$

Re-writing the above equation in terms of *DM* allows one to determine the *DM* by measuring pulse arrival times at multiple frequencies. This in turn can be used to study the interstellar medium, as well as allow for observations of pulsars at different frequencies to be combined.

Dispersion (water waves)

In fluid dynamics, **dispersion** of water waves generally refers to frequency dispersion, which means that waves of different wavelengths travel at different phase speeds. Water waves, in this context, are waves propagating on the water surface, and forced by gravity and surface tension. As a result, water with a free surface is generally considered to be a dispersive medium.

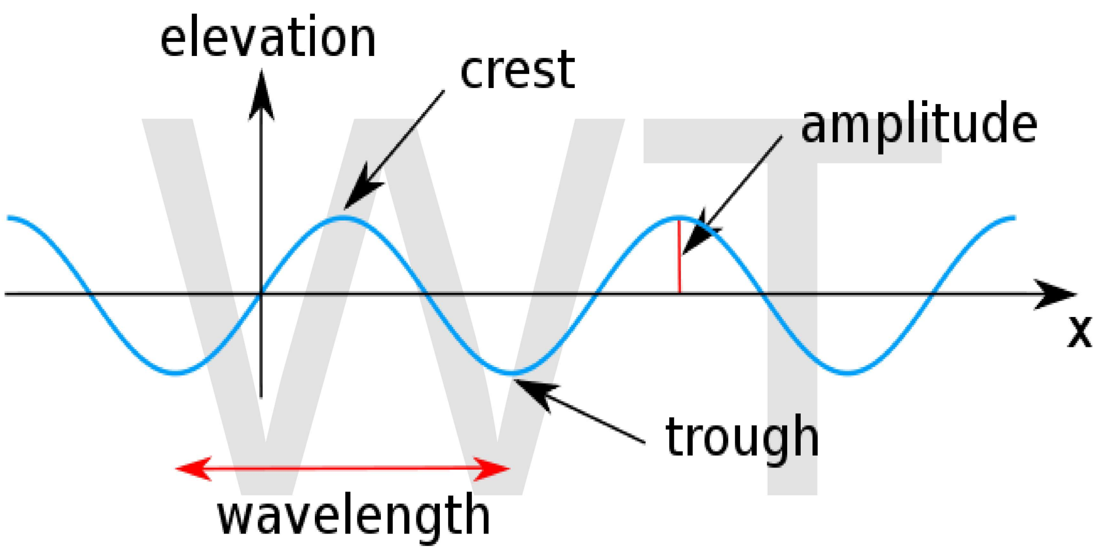
Surface gravity waves, moving under the forcing by gravity, propagate faster for increasing wavelength. For a certain wavelength, gravity waves in deeper water have a larger phase speed than in shallower water. In contrast with this, capillary waves only forced by surface tension, propagate faster for shorter wavelengths.

Besides frequency dispersion, water waves also exhibit amplitude dispersion. This is a nonlinear effect, by which waves of larger amplitude have a different phase speed from small-amplitude waves.

Frequency dispersion for surface gravity waves

This section is about frequency dispersion for waves on a fluid layer forced by gravity, and according to linear theory.

Wave propagation and dispersion



Sinusoidal wave

The simplest propagating wave of unchanging form is a sine wave. A sine wave with water surface elevation $\eta(x, t)$ (SI measure in metre) is given by:

$$\eta(x,t) = a \sin\left(\theta(x,t)\right),$$

where a is the amplitude (in metre) and $\theta = \theta\, x, t)$ is the phase function (in radians), depending on the horizontal position (x, in metre) and time (t, in seconds):

$$\theta = 2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right) = kx - \omega t, \quad \text{with} \quad k = \frac{2\pi}{\lambda} \quad \text{and} \quad \omega = \frac{2\pi}{T},$$

where:

- λ is the wavelength (in metre),
- T is the period (in seconds),
- k is the wavenumber (in radians per metre) and
- ω is the angular frequency (in radians per second).

Characteristic phases of a water wave are:

- the upward zero-crossing at $\theta = 0$,
- the wave crest at $\theta = ½\,\pi$,
- the downward zero-crossing at $\theta = \pi$ and
- the wave trough at $\theta = 1½\,\pi$.

A certain phase repeats itself after an integer m multiple of 2π: $\sin(\theta) = \sin(\theta + m \cdot 2\pi)$.

Essential for water waves, and other wave phenomena in physics, is that free propagating waves of non-zero amplitude only exist when the angular frequency ω and wavenumber k (or equivalently the wavelength λ and period T) satisfy a functional relationship: the frequency dispersion relation

$$\omega^2 = \Omega^2(k).$$

The dispersion relation has two solutions: $\omega = +\Omega(k)$ and $\omega = -\Omega(k)$, corresponding with waves travelling in the positive and negative x–direction. The dispersion relation will in general depend on several other parameters in addition to the wavenumber k. For gravity waves according to linear theory these are the acceleration by gravity and the water depth.

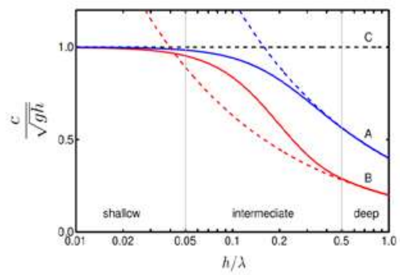
An initial wave phase $\theta = \theta_0$ propagates as a function of space and time. Its subsequent position is given by:

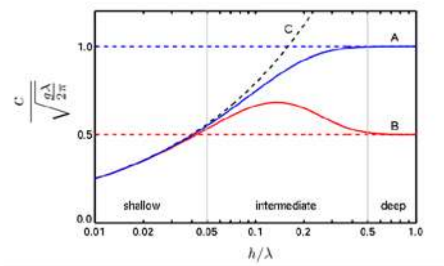
$$x = \frac{\lambda}{T}\,t + \frac{\lambda}{2\pi}\,\theta_0 = \frac{\omega}{k}\,t + \frac{\theta_0}{k}.$$

This shows that the phase moves with the velocity:

$$c_p = \frac{\lambda}{T} = \frac{\omega}{k} = \frac{\Omega(k)}{k},$$

which is called the phase velocity.





Dispersion of gravity waves on a fluid surface. Phase and group velocity divided by shallow-water phase velocity $\sqrt{(gh)}$ as a function of relative depth h / λ.
Blue lines (A): phase velocity; Red lines (B): group velocity; Black dashed line (C): phase and group velocity $\sqrt{(gh)}$ valid in shallow water.
Drawn lines: dispersion relation valid in arbitrary depth.
Dashed lines (blue and red): deep water limits.

Dispersion of gravity waves on a fluid surface. Phase and group velocity divided by deep-water phase velocity $\sqrt{(½\, g\lambda / \pi)}$ as a function of relative depth h / λ.
Blue lines (A): phase velocity; Red lines (B): group velocity; Black dashed line (C): phase and group velocity $\sqrt{(gh)}$ valid in shallow water.
Drawn lines: dispersion relation valid in arbitrary depth.
Dashed lines (blue and red): deep water limits.

Phase velocity

A sinusoidal wave, of small surface-elevation amplitude and with a constant wavelength, propagates with the phase velocity, also called celerity or phase speed. While the phase velocity has in general a direction, celerity and phase speed only provide the magnitude of the velocity and not its direction. According to linear theory for waves forced by gravity, the phase speed depends on the wavelength and the water depth. For a fixed water depth, long waves (with large wavelength) propagate faster than shorter waves.

In the left figure, it can be seen that shallow water waves, with wavelengths λ much larger than the water depth h, travel with the phase velocity

$$c_p = \sqrt{gh} \qquad \text{(shallow water)},$$

with g the acceleration by gravity and c_p the phase speed. Since this shallow-water phase speed is independent of the wavelength, shallow water waves do not have frequency dispersion.

Using another normalization for the same frequency dispersion relation, the figure on the right shows that in deep water, with water depth h larger than half the wavelength λ (so for $h/\lambda > 0.5$), the phase velocity c_p is independent of the water depth:

$$c_p = \sqrt{\frac{g\lambda}{2\pi}} = \frac{g}{2\pi}T \qquad \text{(deep water)},$$

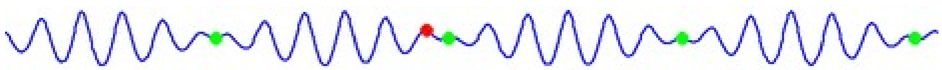
with T the wave period (the reciprocal of the frequency f, $T=1/f$). So in deep water the phase speed increases with the wavelength, and with the period.

Since the phase speed satisfies $c_p = \lambda/T = \lambda f$, wavelength and period (or frequency) are related. For instance in deep water:

$$\lambda = \frac{g}{2\pi}T^2 \qquad \text{(deep water)}.$$

The dispersion characteristics for intermediate depth are given below.

Group velocity



Frequency dispersion in bichromatic groups of gravity waves on the surface of deep water. The red dot moves with the phase velocity, and the green dots propagate with the group velocity.

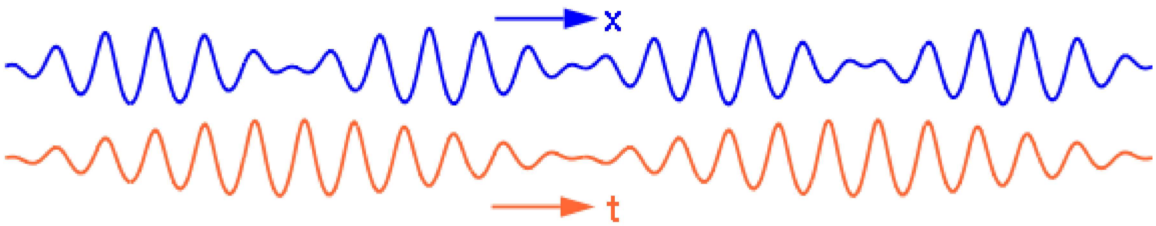
Interference of two sinusoidal waves with slightly different wavelengths, but the same amplitude and propagation direction, results in a beat pattern, called a wave group. As can be seen in the animation, the group moves with a group velocity c_g different from the phase velocity c_p, due to frequency dispersion.

The group velocity is depicted by the red lines (marked *B*) in the two figures above. In shallow water, the group velocity is equal to the shallow-water phase velocity. This is because shallow water waves are not dispersive. In deep water, the group velocity is equal to half the phase velocity: $c_g = ½ c_p$.

The group velocity also turns out the be the energy transport velocity. This is the velocity with which the mean wave energy is transported horizontally in a narrow-band wave field.

In case of a the group velocity different from the phase velocity, a consequence is that the number of waves counted in a wave group is different when counted from a snapshot in space at a certain moment, from when counted in time from the measured surface elevation at a fixed position. Consider a wave group of length Λ_g and group duration of τ_g. The group velocity is:

$$c_g = \frac{\Lambda_g}{\tau_g}.$$



The number of waves per group as observed in space at a certain moment (upper blue line), is different from the number of waves per group seen in time at a fixed position (lower orange line), due to frequency dispersion.

North Pacific storm waves as seen from the NOAA M/V Noble Star, Winter 1989

The number of waves in a wave group, measured in space at a certain moment is: Λ_g / λ. While measured at a fixed location in time, the number of waves in a group is: τ_g / T. So the ratio of the number of waves measured in space to those measured in time is:

$$\frac{\text{No. of waves in space}}{\text{No. of waves in time}} = \frac{\Lambda_g/\lambda}{\tau_g/T} = \frac{\Lambda_g}{\tau_g} \cdot \frac{T}{\lambda} = \frac{c_g}{c_p}.$$

So in deep water, with $c_g = \frac{1}{2} c_p$, a wave group has twice as many waves in time as it has in space.

The water surface elevation $\eta(x,t)$, as a function of horizontal position x and time t, for a bichromatic wave group of full modulation can be mathematically formulated as:

$$\eta = a \sin \left(k_1 x - \omega_1 t \right) + a \sin \left(k_2 x - \omega_2 t \right),$$

with:

- a the wave amplitude of each frequency component in metres,
- k_1 and k_2 the wave number of each wave component, in radians per metre, and
- ω_1 and ω_2 the angular frequency of each wave component, in radians per second.

Both ω_1 and k_1, as well as ω_2 and k_2, have to satisfy the dispersion relation:

$$\omega_1^2 = \Omega^2(k_1) \text{ and } \omega_2^2 = \Omega^2(k_2).$$

Using trigonometric identities, the surface elevation is written as:

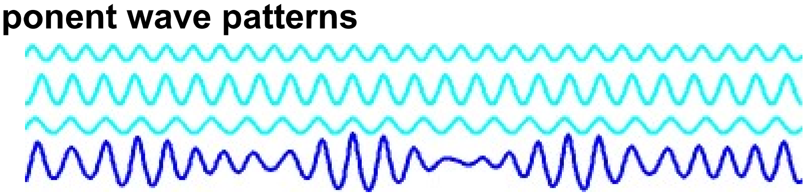
$$\eta = \left[2\, a\, \cos \left(\frac{k_1 - k_2}{2} x - \frac{\omega_1 - \omega_2}{2} t \right) \right] \cdot \sin \left(\frac{k_1 + k_2}{2} x - \frac{\omega_1 + \omega_2}{2} t \right).$$

The part between square brackets is the slowly-varying amplitude of the group, with group wave number $\frac{1}{2}(k_1 - k_2)$ and group angular frequency $\frac{1}{2}(\omega_1 - \omega_2)$. As a result, the group velocity is, for the limit $k_1 \rightarrow k_2$:

$$c_g = \lim_{k_1 \rightarrow k_2} \frac{\omega_1 - \omega_2}{k_1 - k_2} = \lim_{k_1 \rightarrow k_2} \frac{\Omega(k_1) - \Omega(k_2)}{k_1 - k_2} = \frac{\mathrm{d}\Omega(k)}{\mathrm{d}k}.$$

Wave groups can only be discerned in case of a narrow-banded signal, with the wave-number difference $k_1 - k_2$ small compared to the mean wave number $\frac{1}{2}(k_1 + k_2)$.

Multi-component wave patterns



Frequency dispersion of surface gravity waves on deep water. The superposition (dark blue line) of three sinusoidal wave components (light blue lines) is shown.

The effect of frequency dispersion is that the waves travel as a function of wavelength, so that spatial and temporal phase properties of the propagating wave are constantly changing. For example, under the action of gravity, water waves with a longer wavelength travel faster than those with a shorter wavelength.

While two superimposed sinusoidal waves, called a bichromatic wave, have an envelope which travels unchanged, three or more sinusoidal wave components result in a changing pattern of the waves and their envelope. A sea state — that is: real waves on the sea or ocean — can be described as a superposition of many sinusoidal waves with different wavelengths, amplitudes, initial phases and propagation directions. Each of these components travels with its own phase velocity, in accordance with the dispersion relation. The statistics of such a surface can be described by its power spectrum.

Dispersion relation

In the table below, the dispersion relation $\omega^2 = [\Omega(k)]^2$ between angular frequency $\omega = 2\pi / T$ and wave number $k = 2\pi / \lambda$ is given, as well as the phase and group speeds.

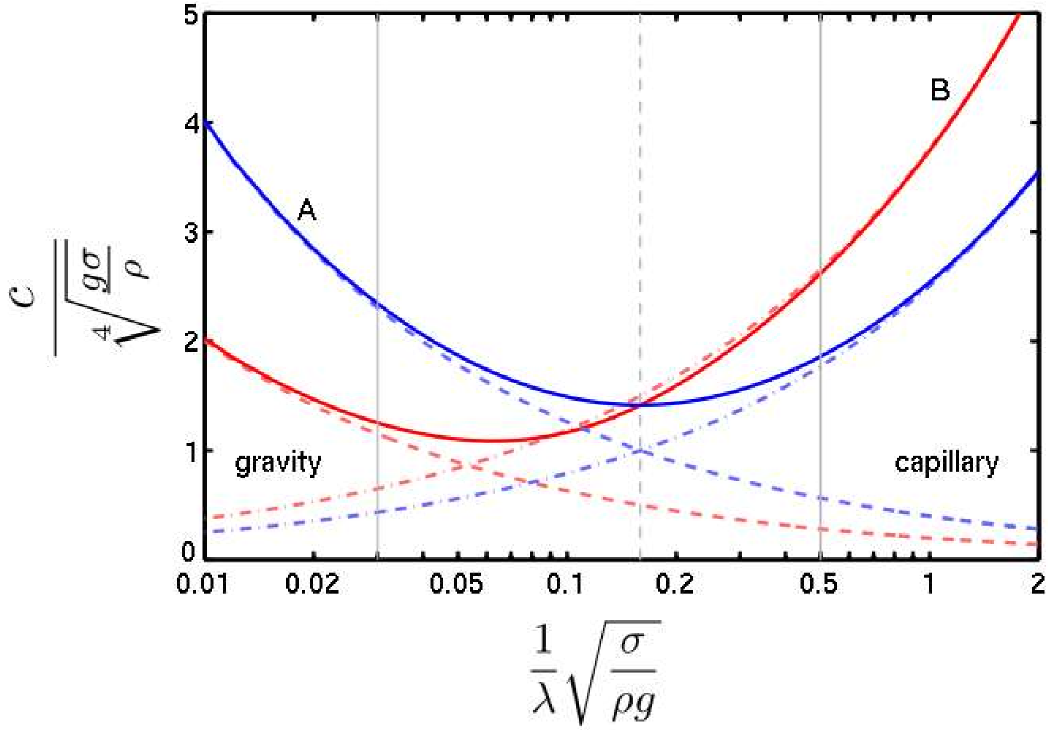
Deep water corresponds with water depths larger than half the wavelength, which is the common situation in the ocean. In deep water, longer period waves propagate faster and transport their energy faster. The deep-water group velocity is half the phase velocity. In shallow water, for wavelengths larger than twenty times the water depth, as found quite often near the coast, the group velocity is equal to the phase velocity.

History

The full linear dispersion relation was first found by Pierre-Simon Laplace, although there were some errors in his solution for the linear wave problem. The complete theory for linear water waves, including dispersion, was derived by George Biddell Airy and published in about 1840. A similar equation was also found by Philip Kelland at around the same time (but making some mistakes in his derivation of the wave theory).

The shallow water (with small h / λ) limit, $\omega^2 = gh\, k^2$, was derived by Joseph Louis Lagrange.

Surface tension effects



Dispersion of gravity-capillary waves on the surface of deep water. Phase and group velocity divided by $\sqrt[4]{g\sigma/\rho}$ as a function of relative wavelength $\frac{1}{\lambda}\sqrt{\sigma/(\rho g)}$.
Blue lines (A): phase velocity, Red lines (B): group velocity.
Drawn lines: dispersion relation for gravity-capillary waves.
Dashed lines: dispersion relation for deep-water gravity waves.
Dash-dot lines: dispersion relation valid for deep-water capillary waves.

In case of gravity-capillary waves, where surface tension affects the waves, the dispersion relation becomes:

$$\omega^2 = \left(gk + \frac{\sigma}{\rho}k^3\right)\tanh(kh),$$

with σ the surface tension (in N/m).

Nonlinear effects

Shallow water

Amplitude dispersion effects appear for instance in the solitary wave (or soliton): a single hump of water traveling with constant velocity in shallow water with a horizontal bed. The single soliton solution of the Korteweg–de Vries equation, of wave height H in water depth h far away from the wave crest, travels with the velocity:

$$c_p = c_g = \sqrt{g(h+H)}.$$

So for this nonlinear gravity wave it is the total water depth under the wave crest that determines the speed, with higher waves traveling faster than lower waves. Note that soliton solutions only exist for positive values of H, solitary gravity waves of depression do not exist.

Deep water

The linear dispersion relation — unaffected by wave amplitude — is for nonlinear waves also correct at the second order of the perturbation theory expansion, with the orders in terms of the wave steepness $k\,A$ (where A is wave amplitude). To the third order, and for deep water, the dispersion relation is

$$\omega^2 = gk\left[1 + (kA)^2\right].$$

This implies that large waves travel faster than small ones of the same frequency. This is only noticeable when the wave steepness $k\,A$ is large.

Waves on a mean current: Doppler shift

Water waves on a mean flow (so a wave in a moving medium) experience a Doppler shift. Suppose the dispersion relation for a non-moving medium is:

$$\omega^2 = \Omega^2(k),$$

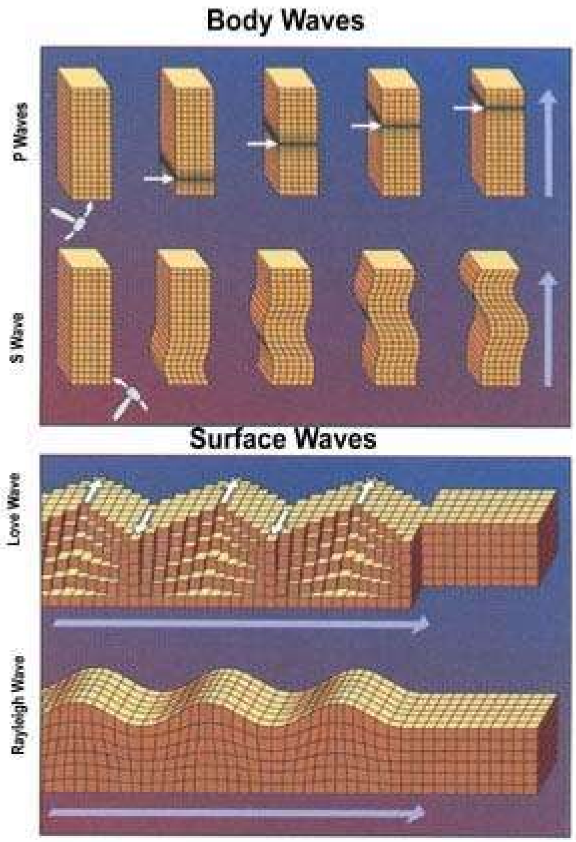
with k the wavenumber. Then for a medium with mean velocity vector **V**, the dispersion relationship with Doppler shift becomes:

$$(\omega - \mathbf{k}\cdot\mathbf{V})^2 = \Omega^2(k),$$

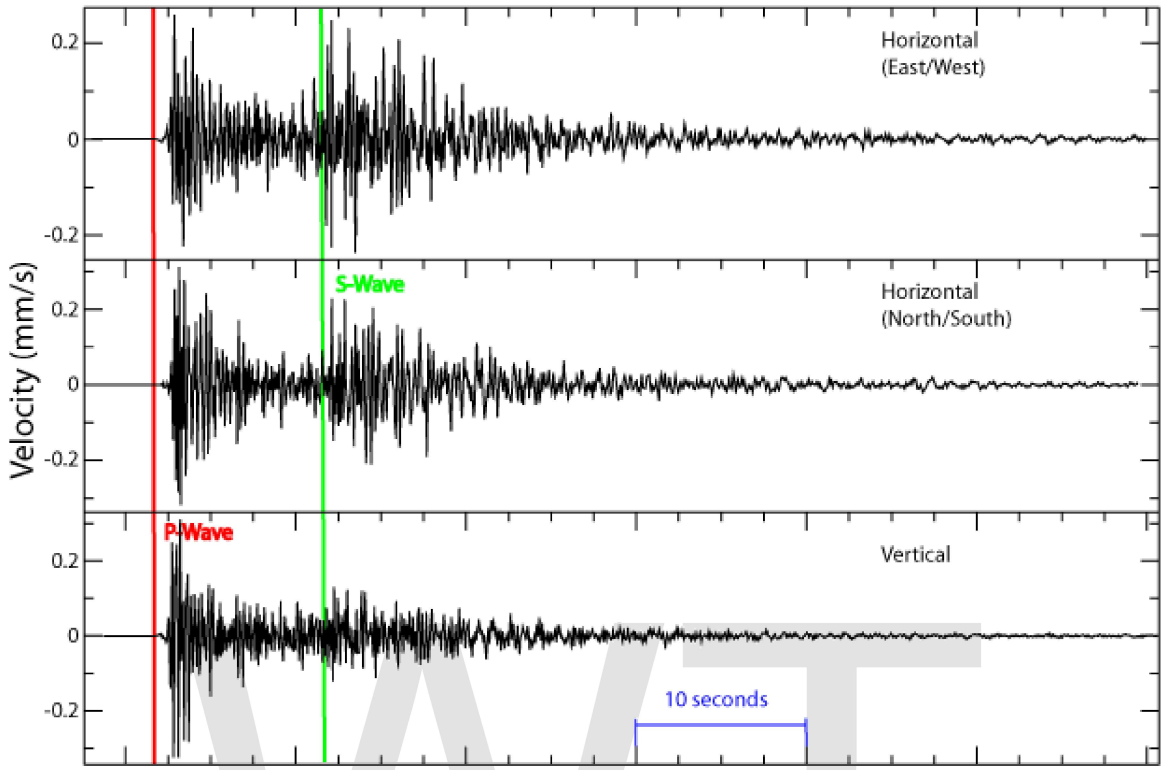
where **k** is the wavenumber vector, related to k as: $k = |\mathbf{k}|$. The inner product $\mathbf{k}\bullet\mathbf{V}$ is equal to: $\mathbf{k}\bullet\mathbf{V} = kV \cos \alpha$, with V the length of the mean velocity vector **V**: $V = |\mathbf{V}|$. And α the angle between the wave propagation direction and the mean flow direction. For waves and current in the same direction, $\mathbf{k} \cdot \mathbf{V} = kV$.

Chapter 5

Seismic Wave



Body waves and surface waves



p-wave and s-wave from seismograph

Seismic waves are waves of energy that travel through the earth or other elastic bodies, for example as a result of an earthquake, explosion, or some other process that imparts low-frequency acoustic energy. Seismic waves are studied by seismologists and geophysicists. Seismic wavefields are measured by a seismograph, geophone, hydrophone (in water), or accelerometer.

The propagation velocity of the waves depends on density and elasticity of the medium. Velocity tends to increase with depth, and ranges from approximately 2 to 8 km/s in the Earth's crust up to 13 km/s in the deep mantle.

Earthquakes create various types of waves with different velocities; when reaching seismic observatories, their different travel time enables the scientists to locate the epicenter. In geophysics the refraction or reflection of seismic waves is used for research of the Earth's interior, and artificial vibrations to investigate subsurface structures.

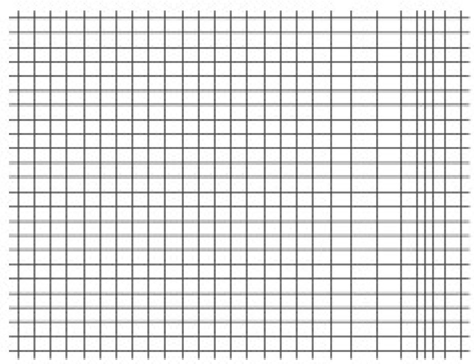
Types of seismic waves

There are two types of seismic waves, *body waves* and *surface waves*. Other modes of wave propagation exist than those described in this article, but they are of comparatively minor importance for earth-borne waves, although they are important in the case of asteroseismology, especially helioseismology.

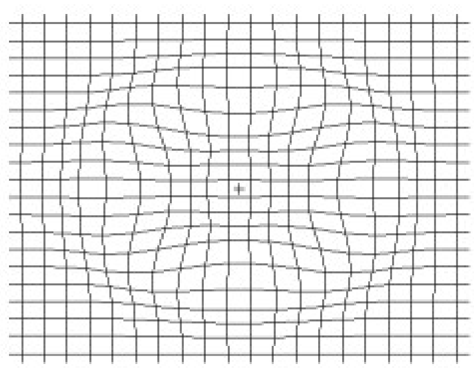
Body waves

Body waves travel through the interior of the Earth. They follow raypaths refracted by the varying density and modulus (stiffness) of the Earth's interior. The density and modulus, in turn, vary according to temperature, composition, and phase. This effect is similar to the refraction of light waves.

P-waves



Plane P-wave



Representation of the propagation of a P-wave on a 2d grid (empirical shape)

P-waves are type of elastic wave, also called seismic waves, that can travel through gases (as sound waves), solids and liquids, including the Earth. P-waves are produced by earthquakes and recorded by seismometers. The name P-wave stands either for **primary wave**, as it has the highest velocity and is therefore the first to be recorded, or **pressure wave**, as it is formed from alternating compressions and rarefactions.

In isotropic and homogeneous solids, the polarization of a P-wave is always longitudinal; thus, the particles in the solid have vibrations along or parallel to the travel direction of the wave energy.

Velocity

The velocity of P-waves in a homogeneous isotropic medium is given by

$$v_p = \sqrt{\frac{K + \frac{4}{3}\mu}{\rho}} = \sqrt{\frac{\lambda + 2\mu}{\rho}}$$

where *K* is the bulk modulus (the modulus of incompressibility), μ is the shear modulus (modulus of rigidity, sometimes denoted as *G* and also called the second Lamé parameter), ρ is the density of the material through which the wave propagates, and λ is the first Lamé parameter.

Of these, density shows the least variation, so the velocity is mostly controlled by *K* and μ.

The elastic moduli P-wave modulus, *M*, is defined so that $M = K + 4\mu / 3$ and thereby

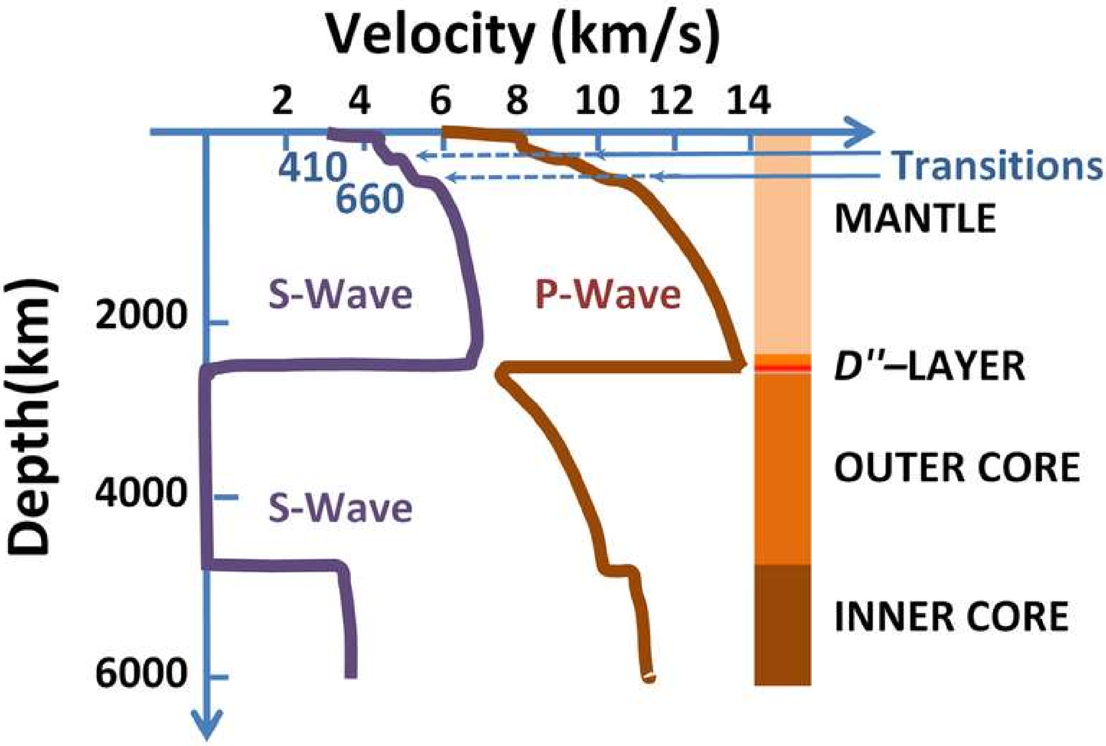
$$v_p = \sqrt{M/\rho}.$$

Typical values for P-wave velocity in earthquakes are in the range 5 to 8 km/s.. The precise speed varies according to the region of the Earth's interior, from less than 6 km/s in the Earth's crust to 13 km/s through the core .

Polarization

In isotropic and homogeneous solids, the polarization of P-waves is always longitudinal. This means that the particles in the body have vibrations along or parallel to the direction of travel of the wave energy.

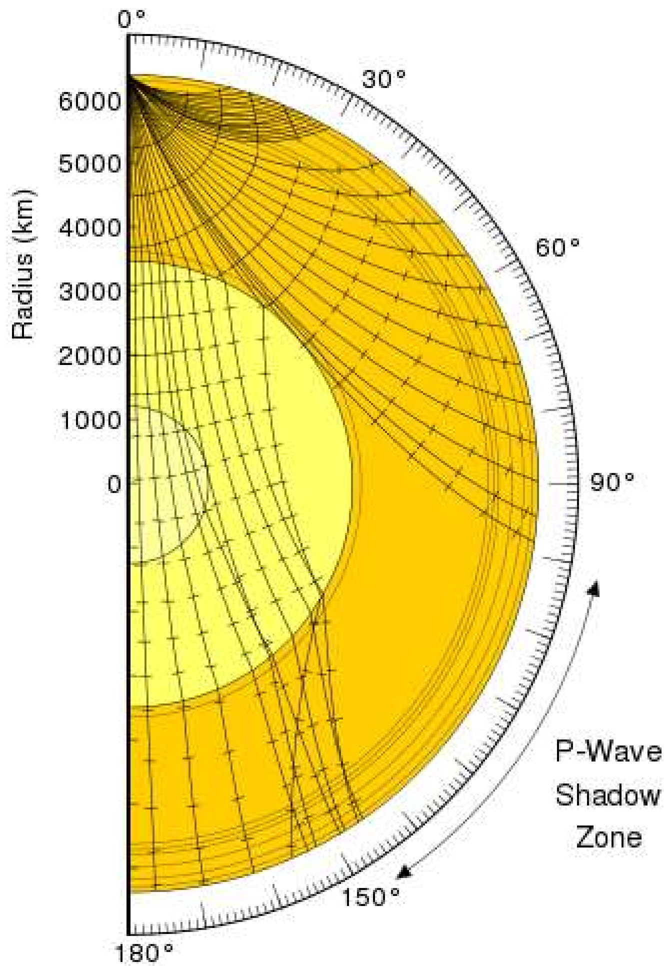
Seismic waves in the Earth



Velocity of seismic waves in the Earth versus depth. The negligible *S*-wave velocity in the outer core occurs because it is liquid, while in the solid inner core the *S*-wave velocity is non-zero.

The seismic waves of both *P*-type and *S*-type in the Earth are monitored to probe the interior structure of the Earth. Discontinuities in velocity as a function of depth are indicative of changes in phase or composition. Differences in arrival times of waves originating in a seismic event like an earthquake as a result of waves taking different paths allow mapping of the Earth's inner structure.

P-wave shadow zone



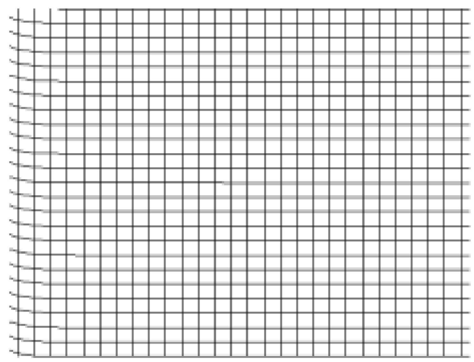
P-wave shadow zone (from USGS)

Almost all the information available on the structure of the Earth's deep interior is derived from observations of the travel times, reflections, refractions and phase transitions of seismic body waves, or normal modes. Body waves travel through the fluid layers of the Earth's interior, but P-waves are refracted slightly when they pass through the transition between the semisolid mantle and the liquid outer core. As a result, there is a P-wave "shadow zone" between 104° and 140° from the earthquake's focus, where the initial P-waves are not registered on seismometers. In contrast, S-waves do not travel through liquids, rather, they are attenuated.

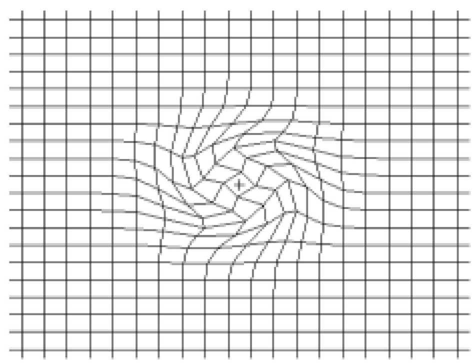
As an earthquake warning

Earthquake advance warning is possible by detecting the non-destructive primary waves that travel more quickly through the Earth's crust than do the destructive secondary and Rayleigh waves. The amount of advance warning depends on the delay between the arrival of the P-wave and other destructive waves, generally on the order of seconds up to about a minute maximum for deep, distant, large quakes. The effectiveness of advance warning depends on accurate detection of the P-waves and compensation for ground vibrations caused by local activity (such as trucks or construction work).

S-waves



Plane shear wave



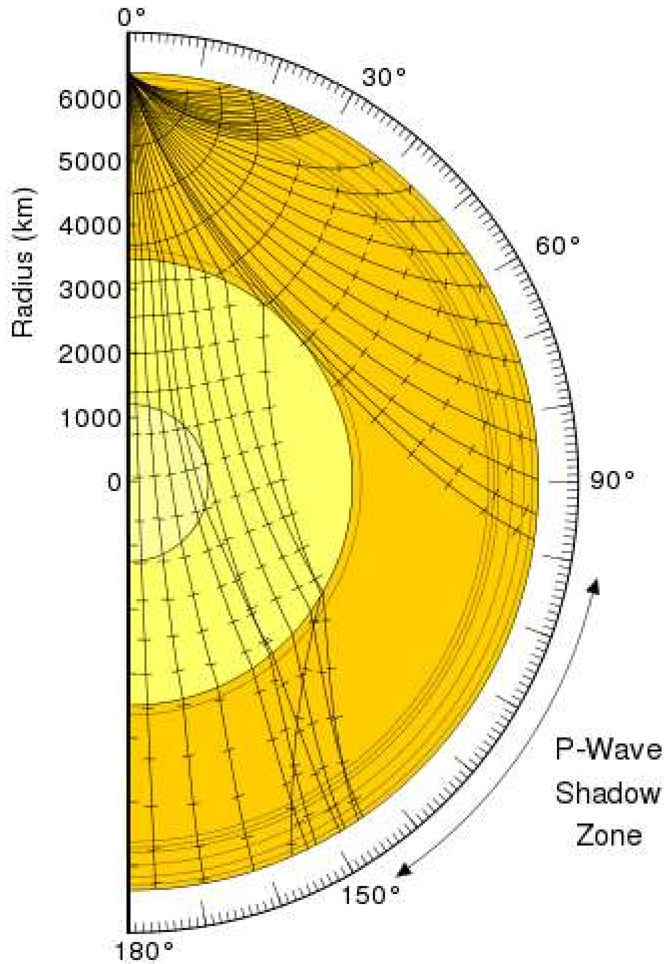
Propagation of a spherical S-wave in a 2d grid (empirical model)

S-wave can also refer to the lowest energy electronic wavefunction in atomic physics.

A type of seismic wave, the **S-wave**, **secondary wave**, or **shear wave** (sometimes called an **elastic S-wave**) is one of the two main types of elastic body waves, so named because they move through the body of an object, unlike surface waves.

The S-wave move as a shear or transverse wave, so motion is perpendicular to the direction of wave propagation: S-waves, like waves in a rope, as opposed to waves moving through a slinky, the P-wave. The wave moves through elastic media, and the main restoring force comes from shear effects. These waves are divergenceless and obey the continuity equation for incompressible media:

$$\nabla \cdot \mathbf{u} = 0.$$



The shadow zone of a P-wave. S-waves don't penetrate the outer core, so they're shadowed everywhere more than 104° away from the epicenter (from USGS)

Its name, S for secondary, comes from the fact that it is the second direct arrival on an earthquake seismogram, after the compressional primary wave, or P-wave, because S-waves travel slower in rock. Unlike the P-wave, the S-wave cannot travel through the molten outer core of the Earth, and this causes a shadow zone for S-waves opposite to where they originate. They can still appear in the solid inner core: when a P-wave strikes the boundary of molten and solid cores, called the Lehmann discontinuity, S-waves will then propagate in the solid medium. And when the S-waves hit the boundary again they will in turn create P-waves. This property allows seismologists to determine the nature of the inner core.

As transverse waves, S-waves exhibit properties, such as polarization and birefringence, much like other transverse waves. S-waves polarized in the horizontal plane are classified as SH-waves. If polarized in the vertical plane, they are classified as SV-waves. When an S- or P-wave strikes an interface at an angle other than 90 degrees, a phenomenon known as mode conversion occurs. As described above, if the interface is between a solid and liquid, S becomes P or vice versa. However, even if the interface is between two solid media, mode conversion results. If a P-wave strikes an interface, four propagation modes may result: reflected and transmitted P and reflected and transmitted SV. Similarly, if an SV-wave strikes an interface, the same four modes occur in different proportions. The exact amplitudes of all these waves are described by the Zoeppritz equations, which in turn are solutions to the wave equation. S-waves are slower than P-waves.

Theory

The prediction of S-waves came out of theory in the 1800s. Starting with the stress-strain relationship for an isotropic solid:

$$\tau_{ij} = \lambda \delta_{ij} e_{kk} + 2\mu e_{ij}$$

where τ is the stress, λ and μ are the Lamé parameters (with μ as the shear modulus), δ_{ij} is the Kronecker delta, and the strain tensor is defined

$$e_{ij} = \frac{1}{2}\left(\partial_i u_j + \partial_j u_i\right)$$

for strain displacement u. Plugging the latter into the former yields:

$$\tau_{ij} = \lambda \delta_{ij} \partial_k u_k + \mu\left(\partial_i u_j + \partial_j u_i\right).$$

Newton's 2nd law in this situation gives the *homogeneous equation of motion* for seismic wave propagation:

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \partial_j \tau_{ij}$$

where ρ is the mass density. Plugging in what the stress tensor is above gives:

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \partial_i \lambda \partial_k u_k + \partial_j \mu\left(\partial_i u_j + \partial_j u_i\right) + \lambda \partial_i \partial_k u_k + \mu \partial_i \partial_j u_j + \mu \partial_j \partial_j u_i.$$

Applying vector identities and making certain approximations gives the seismic wave equation in homogeneous media:

$$\rho \ddot{\boldsymbol{u}} = (\lambda + 2\mu)\nabla(\nabla \cdot \boldsymbol{u}) - \mu \nabla \times (\nabla \times \boldsymbol{u})$$

where Newton's notation has been used for the time derivative. Taking the curl of this equation and applying vector identities eventually gives:

$$\nabla^2(\nabla \times \boldsymbol{u}) - \frac{1}{\beta^2}\frac{\partial^2(\nabla \times \boldsymbol{u})}{\partial t^2} = 0$$

which is simply the wave equation applied to the curl of u with a velocity β satisfying

$$\beta^2 = \frac{\mu}{\rho}.$$

This describes S-wave propagation. Taking the divergence of seismic wave equation in homogeneous media instead of the curl, yields an equation describing P-wave propgation.

Surface waves

Diving grebe creates surface waves

In physics, a **surface wave** is a mechanical wave that propagates along the interface between differing media, usually two fluids with different densities. A surface wave can also be an electromagnetic wave guided by a refractive index gradient. In radio transmission, a **ground wave** is a surface wave that propagates close to the surface of the Earth.

Mechanical waves

In seismology, several types of surface waves are encountered. Surface waves, in this mechanical sense, are commonly known as either **Love waves** (L waves) or **Rayleigh waves**. A seismic wave is a wave that *travels through the Earth, often as the result of an earthquake or explosion.* Love waves have transverse motion (movement is perpendicular to the direction of travel, like light waves), whereas Rayleigh waves have both longitudinal (movement parallel to the direction of travel, like sound waves) and transverse motion. Seismic waves are studied by seismologists and measured by a seismograph or seismometer. Surface waves span a wide frequency range, and the period of waves that are most damaging is usually 10 seconds or longer. Surface waves can travel around the globe many times from the largest earthquakes.

The term "surface wave" can describe waves over an ocean, even when they are approximated by Airy functions and are more properly called creeping waves. Examples are the waves at the surface of water and air (ocean surface waves), or ripples in the sand at the interface with water or air. Another example is internal waves, which can be transmitted along the interface of two water masses of different densities.

Electromagnetic waves

Ground waves refer to the propagation of radio waves close to or at the surface of the Earth. These *surface waves* are also known loosely as **Norton surface waves**, **Zenneck waves**, **Sommerfeld waves**, or **gliding waves**.

Radio propagation

Lower frequencies, especially AM broadcasts in the mediumwave (sometimes called "medium frequency") and long wave bands (and other types of radio frequencies below that), travel efficiently as a surface wave. This is because they are more efficiently diffracted by the figure of the Earth due to their low frequencies. Ionospheric reflection is taken into consideration as well. The ionosphere reflects frequencies in a certain band, which often changes due to solar conditions. The Earth has one refractive index and the atmosphere has another, thus constituting an interface that supports the surface wave transmission.

Conductivity of the surface affects the propagation of ground waves, with more conductive surfaces such as water providing better propagation. Increasing the conductivity in a surface results in less dissipation. The refractive indices are subject to spatial and temporal changes. Since the ground is not a perfect electrical conductor, ground waves are attenuated as they follow the earth's surface.

Most long-distance LF "longwave" radio communication (between 30 kHz and 300 kHz) is a result of groundwave propagation. Mediumwave radio transmissions (frequencies between 300 kHz and 3000 kHz) have the property of following the curvature of the earth (the groundwave) in the majority of occurrences. At low frequencies, ground losses are low and become lower at lower frequencies. The VLF and LF frequencies are mostly used for military communications, especially with ships and submarines.

Surface waves have been used in over-the-horizon radar. In the development of radio, surface waves were used extensively. Early commercial and professional radio services relied exclusively on long wave, low frequencies and ground-wave propagation. To prevent interference with these services, amateur and experimental transmitters were restricted to the higher (HF) frequencies, felt to be useless since their ground-wave range was limited. Upon discovery of the other propagation modes possible at medium wave and short wave frequencies, the advantages of HF for commercial and military purposes became apparent. Amateur experimentation was then confined only to authorized frequencies in the range.

Mediumwave and shortwave reflect off the ionosphere at night, which is known as skywave. During daylight hours, the lower "D" layer of the ionosphere forms and absorbs lower frequency energy. This prevents skywave propagation from being very effective on mediumwave frequencies in daylight hours. At night, when the "D" layer dissipates, mediumwave transmissions travel better by skywave. Ground waves *do not* include ionospheric and tropospheric waves.

Microwave field theory

Within microwave field theory, the refractive index of many cavities constitute an interface that supports "surface wave transmission". Surface waves have been studied as part of transmission lines and some may be considered as single-wire transmission line.

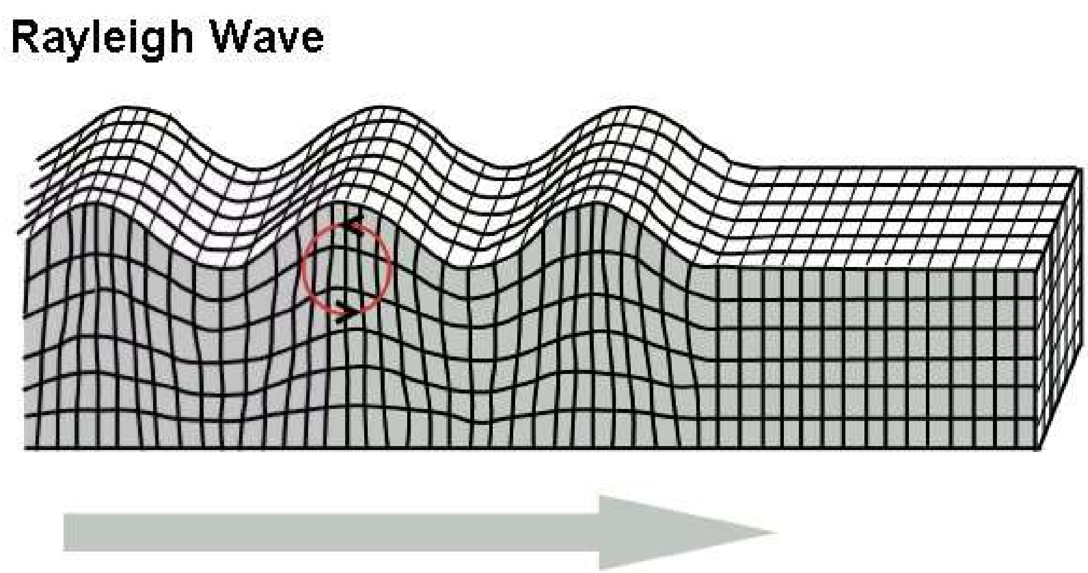
Characteristics and utilizations of the electrical surface wave phenomena include:

- The field components of the wave diminish with distance from the interface.
- Optical energy is not converted from the surface wave field to another form of energy and the wave does not have a component directed normal to the interface surface.
- In optical fiber transmission, evanescent waves are surface waves.
- In coaxial cable in addition to the TEM mode there also exists a transverse-magnetic (TM) mode which propagates as a surface wave in the region around the central conductor. For coax of common impedance this mode is effectively suppressed but in high impedance coax and on a single central conductor without any outer shield, low attenuation and very broadband propagation is supported. Transmission line operating in this mode is called E-Line.

Rayleigh waves

Rayleigh waves are a type of surface acoustic wave that travels on solids. They are produced on the Earth by earthquakes, in which case they are also known as "ground roll", or by other sources of seismic energy such as an explosion or even a sledgehammer impact. They can also be produced in materials by many mechanisms, including by piezo-electric transducers, and are frequently used in non-destructive testing for detecting defects. When guided in layers they are referred to as Lamb waves, Rayleigh–Lamb waves, or generalized Rayleigh waves.

Characteristics



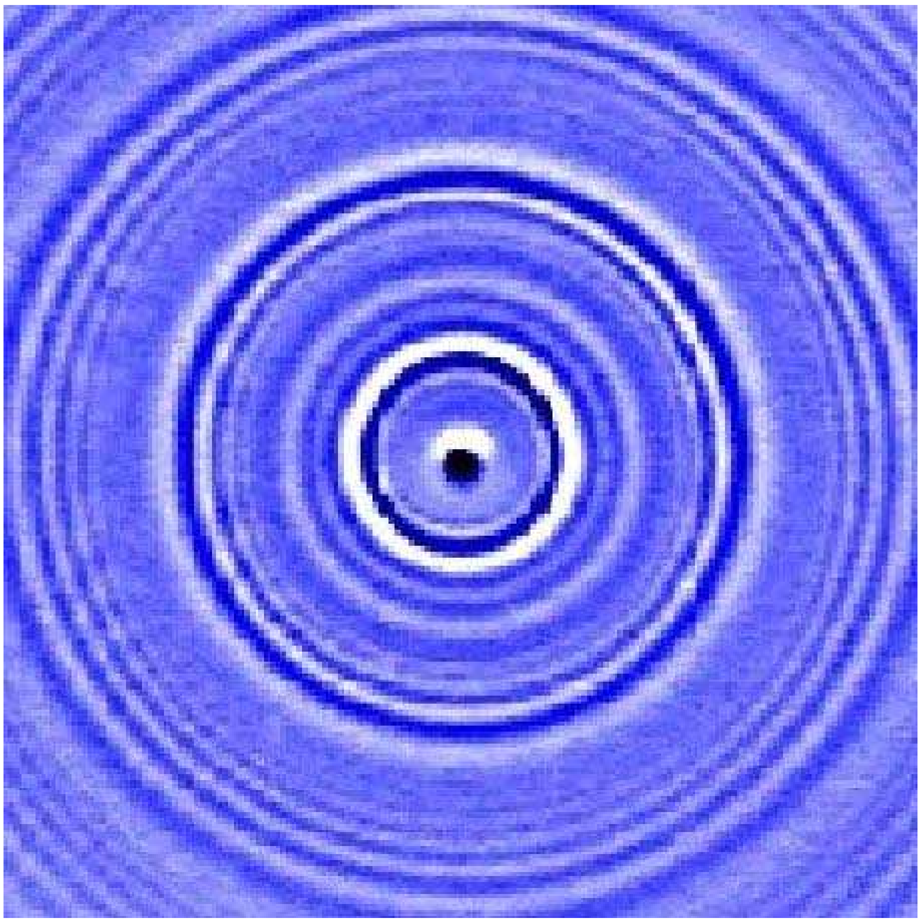
Picture of a Rayleigh wave

Rayleigh waves travel across surfaces. In isotropic solids the surface particles move in ellipses in planes normal to the surface and parallel to the direction of propagation. At the surface and at shallow depths this motion is retrograde. Particles deeper in the material move in smaller ellipses with an eccentricity that changes with depth. At greater depths the particle motion becomes prograde. The depth of significant displacement[] in the solid is approximately equal to the acoustic wavelength. Rayleigh waves are distinct from other types of acoustic waves such as Love waves or Lamb waves, both being types of guided wave in a layer, or longitudinal and shear waves, that travel in the bulk.

Since Rayleigh waves are confined near the surface, their in-plane amplitude when generated by a point source decays only as $1/\sqrt{r}$, where *r* is the radial distance. Surface waves therefore decay more slowly with distance than do bulk waves, which spread out in three dimensions from a point source. The speed of Rayleigh waves on bulk solids, of the order of 2–5 km/s, is slightly less than the shear velocity.

The existence of Rayleigh waves was predicted in 1885 by Lord Rayleigh, after whom they were named.

Rayleigh wave dispersion



Dispersion of Rayleigh waves in a thin gold film on glass

Rayleigh waves on ideal, homogeneous and flat elastic solids show no dispersion. However, if a solid or structure has a density or sound velocity that varies with depth, Rayleigh waves become dispersive. One example is Rayleigh waves on the Earth's surface: those waves with a higher frequency travel more slowly than those with a lower frequency. This occurs because a Rayleigh wave of lower frequency has a relatively long wavelength. The displacement of long wavelength waves penetrates more deeply into the Earth than short wavelength waves. Since the speed of waves in the Earth increases with increasing depth, the longer wavelength (low frequency) waves can travel faster than the shorter wavelength (high frequency) waves. Rayleigh waves thus often appear spread out on seismograms recorded at distant earthquake recording stations. It is also possible to observe Rayleigh wave dispersion in thin films or multilayed structures.

Rayleigh waves in non-destructive testing

Rayleigh waves are widely used for materials characterization at different length scales because they are easily generated and detected on the free surface of solid objects. Since they are confined in the vicinity of the free surface within a depth (~ the wavelength) linked to the frequency of the wave, different frequencies can be used for characterization at different length scales.

Rayleigh waves in the ultrasonic frequency range are used in non-destructive testing applications to help find cracks and other imperfections in materials.

Rayleigh waves in geophysics

Rayleigh waves from earthquakes

Since Rayleigh waves are surface waves, the amplitude of such waves generated by an earthquake generally decreases exponentially with the depth of the hypocenter. However, Large earthquakes may generate Rayleigh waves that travel around the Earth several times before dissipating.

In seismology longitudinal and shear waves are known as P-waves and S-waves, respectively, and are termed body waves. Rayleigh waves are generated by the interaction of P- and S- waves at the surface of the earth, and travel with a velocity that is lower than the P-, S-, and Love wave velocities. Rayleigh waves emanating outward from the epicenter of an earthquake travel along the surface of the earth at about 10 times the speed of sound (0.340 km/s), in air, that is ~3 km/s.

Due to their higher speed, the P- and S-waves generated by an earthquake arrive before the surface waves. However, the particle motion of surface waves is larger than that of body waves, so the surface waves tend to cause more damage. In the case of Rayleigh waves, the motion is of a rolling nature, similar to an ocean surface wave. The intensity of Rayleigh wave shaking at a particular location is dependent on several factors:

- The size of the earthquake.
- The distance to the earthquake.
- The depth of the earthquake.
- The geologic structure of the crust.
- The focal mechanism of the earthquake.
- The rupture directivity of the earthquake.

Local geologic structure can serve to focus or defocus Rayleigh waves, leading to significant differences in shaking over short distances.

Rayleigh waves in seismology

Low frequency Rayleigh waves generated during earthquakes are used in seismology to characterise the Earth's interior. In intermediate ranges, Rayleigh waves are used in geophysics and geotechnical engineering for the characterisation of oil deposits. These application are based on the geometric dispersion of Rayleigh waves and on the solution of an inverse problem on the basis of seismic data collected on the ground surface using active sources (falling weights, hammers or small explosions, for example) or by recording microtremors.

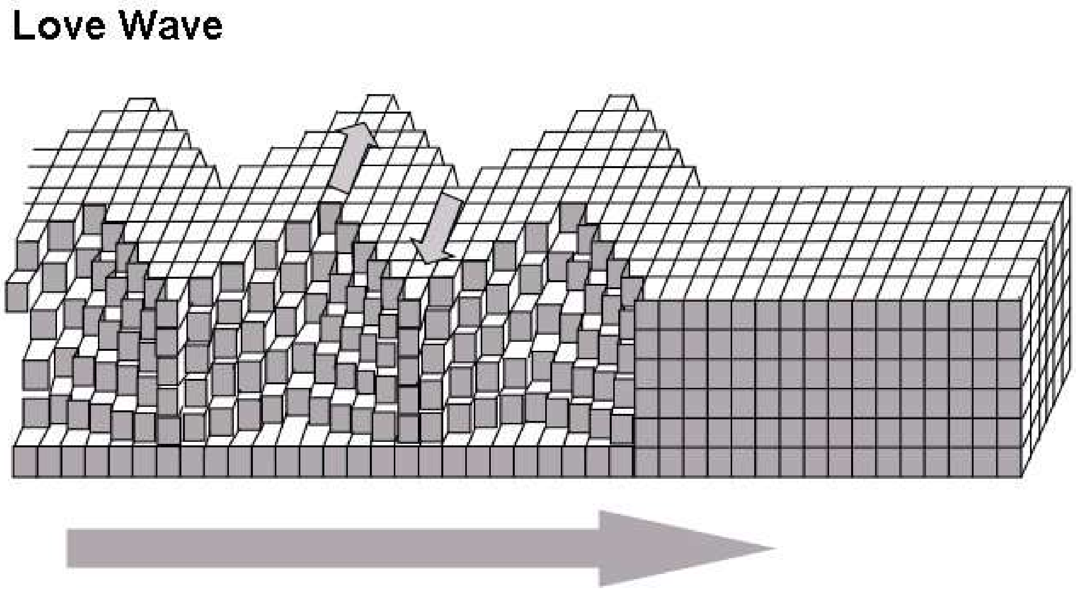
Other manifestations

Animals

Low frequency (< 20 Hz) Rayleigh waves are inaudible, yet they can be detected by many mammals, birds, insects and spiders. Human beings should be able to detect such Rayleigh waves through their Pacinian corpuscles, which are in the joints, although people do not seem to consciously respond to the signals. Some animals seem to use Rayleigh waves to communicate. In particular, some biologists theorize that elephants may use vocalizations to generate Rayleigh waves. Since Rayleigh waves decay slowly, they should be detectable over long distances. Note that these Rayleigh waves have a much higher frequency than Rayleigh waves generated by earthquakes.

After the 2004 Indian Ocean Earthquake, some people have speculated that Rayleigh waves served as a warning to animals to seek higher ground, allowing them to escape the more slowly-traveling tsunami. At this time, evidence for this is mostly anecdotal. Another animal early warning systems may rely on an ability to sense infrasonic waves traveling through the air.

L waves or Love waves



How love waves work

In elastodynamics, **Love waves** are essentially horizontally polarized shear waves (SH waves) guided by an elastic layer, which is "welded" to an elastic half space on one side while bordering a vacuum on the other side. In seismology, **Love waves** (also named **Q waves** (**Q**uer: German for lateral)) are surface seismic waves that cause horizontal shifting of the earth during an earthquake. A.E.H. Love predicted the existence of Love waves mathematically in 1911; the name comes from him (Chapter 11 from Love's book "Some problems of geodynamics", first published in 1911). They form a distinct class, different from other types of seismic waves, such as P-waves and S-waves (both body waves), or Rayleigh waves (another type of surface wave). Love waves travel with a slower velocity than P- or S- waves, but faster than Rayleigh waves.

Description

The particle motion of a Love wave forms a horizontal line perpendicular to the direction of propagation (i.e. are transverse waves). Moving deeper into the material, motion can decrease to a "node" and then alternately increase and decrease as one examines deeper layers of particles. The amplitude, or maximum particle motion, often decreases rapidly with depth.

Since Love waves travel on the Earth's surface, the ability (or amplitude) of the waves decrease exponentially with the depth of an earthquake. However, given their confinement to the surface, their amplitude decays only as $\frac{1}{\sqrt{r}}$, where *r* represents the distance the wave has traveled from the earthquake. Surface waves therefore decay more slowly with distance than do body waves, which travel in three dimensions. Large earthquakes may generate Love waves that travel around the Earth several times before dissipating.

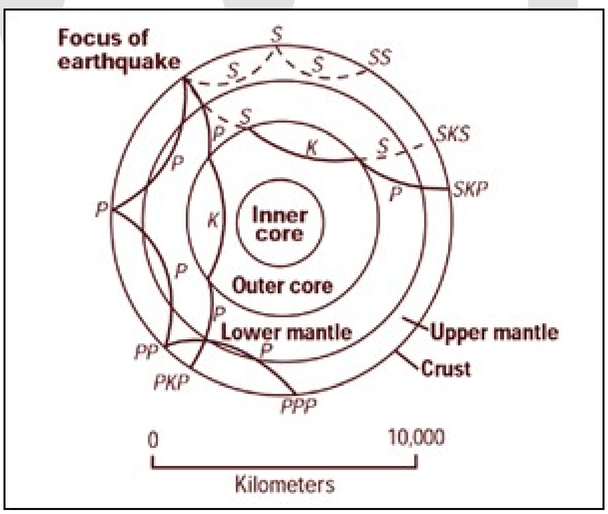
Since they decay so slowly, Love waves are the most destructive outside the immediate area of the focus or epicentre of an earthquake. They are what most people feel directly during an earthquake.

In the past, it was often thought that animals like cats and dogs could predict an earthquake before it happened. However, they are simply more sensitive to ground vibrations than humans and able to detect the subtler waves that precede Love waves, like the P-waves and the S-waves.

P and S waves in Earth's mantle and core

When an earthquake occurs, seismographs near the epicenter are able to record both P and S waves, but those at a greater distance no longer detect the high frequencies of the first S wave. Since shear waves cannot pass through liquids, this phenomenon was original evidence for the now well-established observation that the Earth has a liquid outer core, as demonstrated by Richard Dixon Oldham. This kind of observation has also been used to argue, by seismic testing, that the Moon has a solid core, although recent geodetic studies suggest the core is still molten.

Notation



Earthquake wave paths

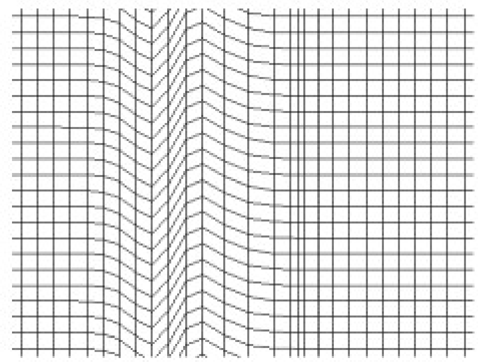
The path that a wave takes between the focus and the observation point is often drawn as a ray diagram. An example of this is shown in a figure above. When reflections are taken into account there are an infinite number of paths that a wave can take. Each path is denoted by a set of letters that describe the trajectory and phase through the Earth. In general an upper case denotes a transmitted wave and a lower case denotes a reflected wave. The two exceptions to this seem to be "g" and "n". The notation is taken from and.

c the wave reflects off the outer core
d a wave that has been reflected off a discontinuity at depth d
g a wave that only travels through the crust
i a wave that reflects off the inner core
I a P-wave in the inner core
h a reflection off a discontinuity in the inner core
J an S wave in the inner core
K a P-wave in the outer core
L a Love wave sometimes called LT-Wave (Both caps, while an Lt is different)
n a wave that travels along the boundary between the crust and mantle
P a P wave in the mantle
p a P wave ascending to the surface from the focus
R a Rayleigh wave
S an S wave in the mantle
s an S wave ascending to the surface from the focus
w the wave reflects off the bottom of the ocean
No letter is used when the wave reflects off of the surface

For example:

- **ScP** is a wave that begins traveling towards the center of the Earth as an S wave. Upon reaching the outer core the wave reflects as a P wave.
- **sPKIKP** is a wave path that begins traveling towards the surface as an S-wave. At the surface it reflects as a P-wave. The P-wave then travels through the outer core, the inner core, the outer core, and the mantle.

Usefulness of P and S waves in locating an event



P- and S-waves sharing with the propagation

In the case of local or nearby earthquakes, the difference in the arrival times of the P and S waves can be used to determine the distance to the event. In the case of earthquakes that have occurred at global distances, four or more geographically diverse observing stations (using a common clock) recording P-wave arrivals permits the computation of a unique time and location on the planet for the event. Typically, dozens or even hundreds of P-wave arrivals are used to calculate hypocenters. The misfit generated by a hypocenter calculation is known as "the residual". Residuals of 0.5 second or less are typical for distant events, residuals of 0.1-0.2 s typical for local events, meaning most reported P arrivals fit the computed hypocenter that well. Typically a location program will start by assuming the event occurred at a depth of about 33 km; then it minimizes the residual by adjusting depth. Most events occur at depths shallower than about 40 km, but some occur as deep as 700 km.

A quick way to determine the distance from a location to the origin of a seismic wave less than 200 km away is to take the difference in arrival time of the P wave and the S wave in seconds and multiply by 8 kilometers per second. Modern seismic arrays use more complicated earthquake location techniques.

At teleseismic distances, the first arriving P waves have necessarily travelled deep into the mantle, and perhaps have even refracted into the outer core of the planet, before travelling back up to the Earth's surface where the seismographic stations are located. The waves travel more quickly than if they had traveled in a straight line from the earthquake. This is due to the appreciably increased velocities within the planet, and is termed Huygens' Principle. Density in the planet increases with depth, which would slow the waves, but the modulus of the rock increases much more, so deeper means faster. Therefore, a longer route can take a shorter time.

The travel time must be calculated very accurately in order to compute a precise hypocenter. Since P waves move at many kilometers per second, being off on travel-time calculation by even a half second can mean an error of many kilometers in terms of distance. In practice, P arrivals from many stations are used and the errors cancel out, so the computed epicenter likely to be quite accurate, on the order of 10–50 km or so around the world. Dense arrays of nearby sensors such as those that exist in California can provide accuracy of roughly a kilometer, and much greater accuracy is possible when timing is measured directly by cross-correlation of seismogram waveforms.

Chapter 6

Electromagnetic Radiation

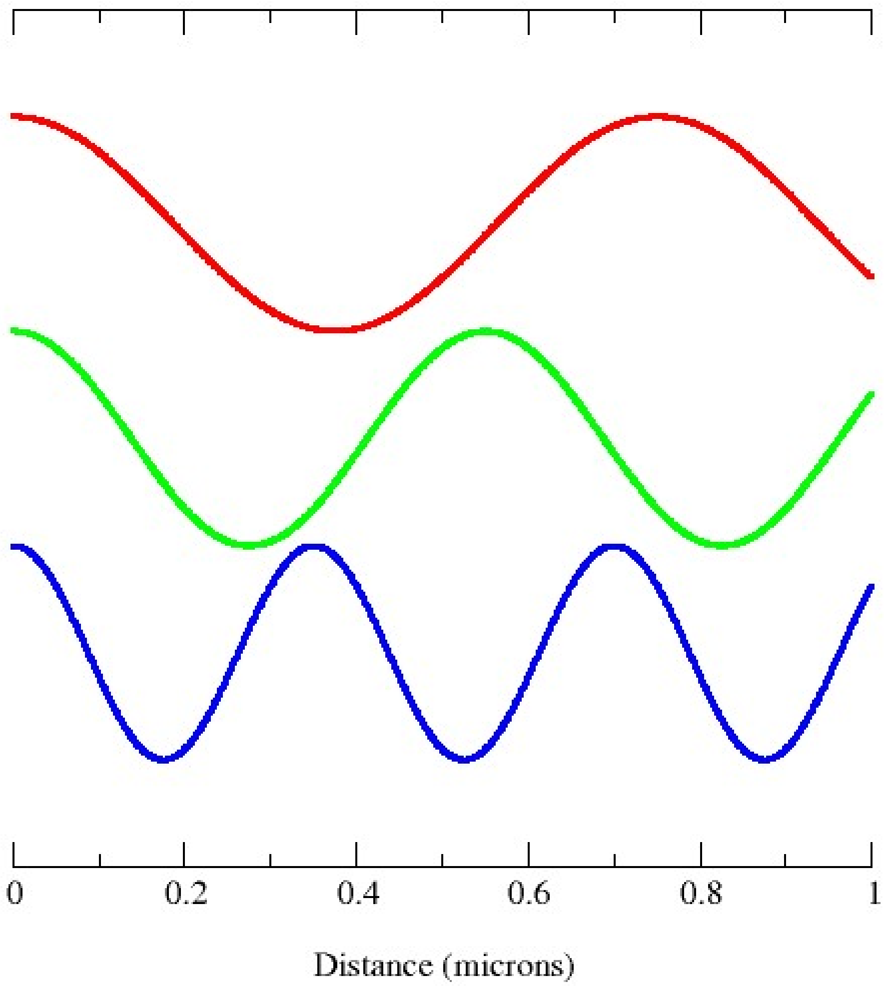
Electromagnetic radiation (often abbreviated **E-M radiation** or **EMR**) is a phenomenon that takes the form of self-propagating waves in a vacuum or in matter. It comprises electric and magnetic field components, which oscillate in phase perpendicular to each other and perpendicular to the direction of energy propagation.

Electromagnetic radiation is classified into several types according to the frequency of its wave; these types include (in order of increasing frequency and decreasing wavelength): radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays and gamma rays. A small and somewhat variable window of frequencies is sensed by the eyes of various organisms; this is what is called the visible spectrum. The photon is the quantum of the electromagnetic interaction and the basic "unit" of light and all other forms of electromagnetic radiation and is also the force carrier for the electromagnetic force.

EM radiation carries energy and momentum that may be imparted to matter with which it interacts.

Physics

Theory



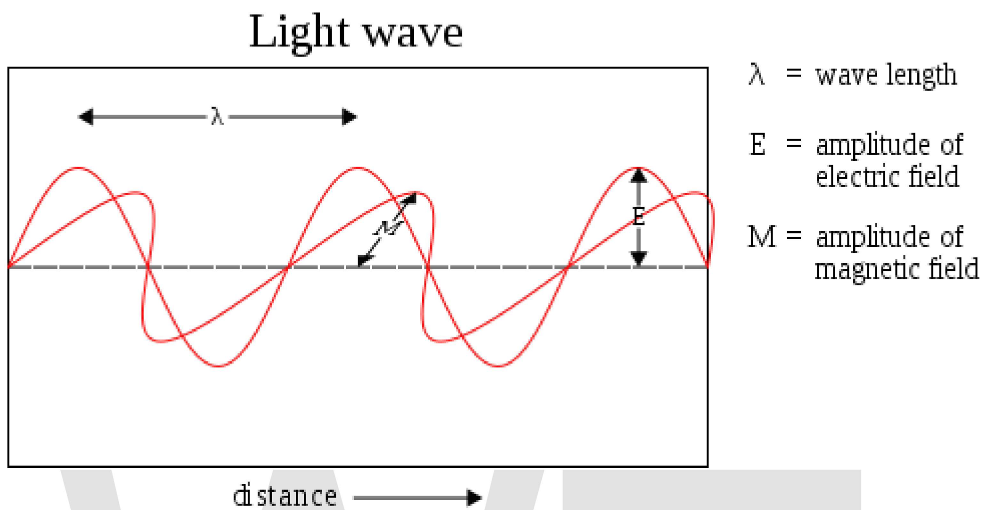
Shows three electromagnetic modes (blue, green and red) with a distance scale in micrometres along the x-axis.

Electromagnetic waves were first postulated by James Clerk Maxwell and subsequently confirmed by Heinrich Hertz. Maxwell derived a wave form of the electric and magnetic equations, revealing the wave-like nature of electric and magnetic fields, and their symmetry. Because the speed of EM waves predicted by the wave equation coincided with the measured speed of light, Maxwell concluded that light itself is an EM wave.

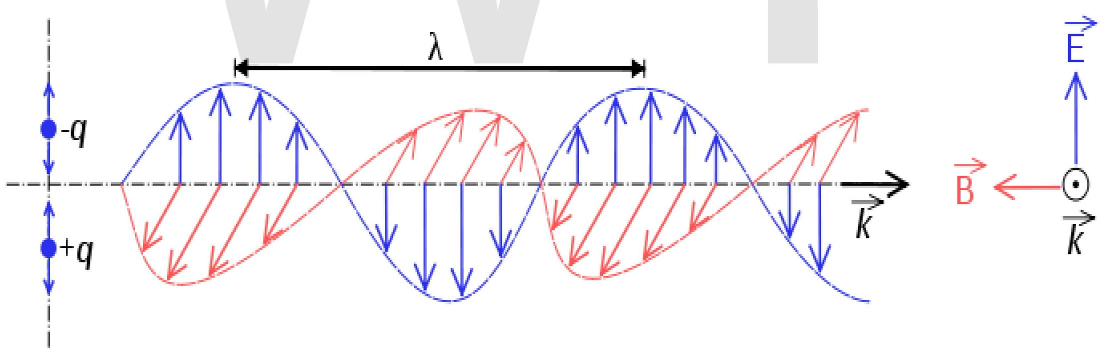
According to Maxwell's equations, a spatially-varying electric field generates a time-varying magnetic field and *vice versa*. Therefore, as an oscillating electric field generates an oscillating magnetic field, the magnetic field in turn generates an oscillating electric field, and so on. These oscillating fields together form an electromagnetic wave.

A quantum theory of the interaction between electromagnetic radiation and matter such as electrons is described by the theory of quantum electrodynamics.

Properties



Electromagnetic waves can be imagined as a self-propagating transverse oscillating wave of electric and magnetic fields. This diagram shows a plane linearly polarized wave propagating from right to left. The electric field is in a vertical plane and the magnetic field in a horizontal plane.



The physics of electromagnetic radiation is electrodynamics. Electromagnetism is the physical phenomenon associated with the theory of electrodynamics. Electric and magnetic fields obey the properties of superposition so that a field due to any particular particle or time-varying electric or magnetic field will contribute to the fields present in the same space due to other causes: as they are vector fields, all magnetic and electric field vectors add together according to vector addition. For instance, a travelling EM wave incident on an atomic structure induces oscillation in the atoms of that structure, thereby causing them to emit their own EM waves, emissions which alter the impinging wave through interference. These properties cause various phenomena including refraction and diffraction.

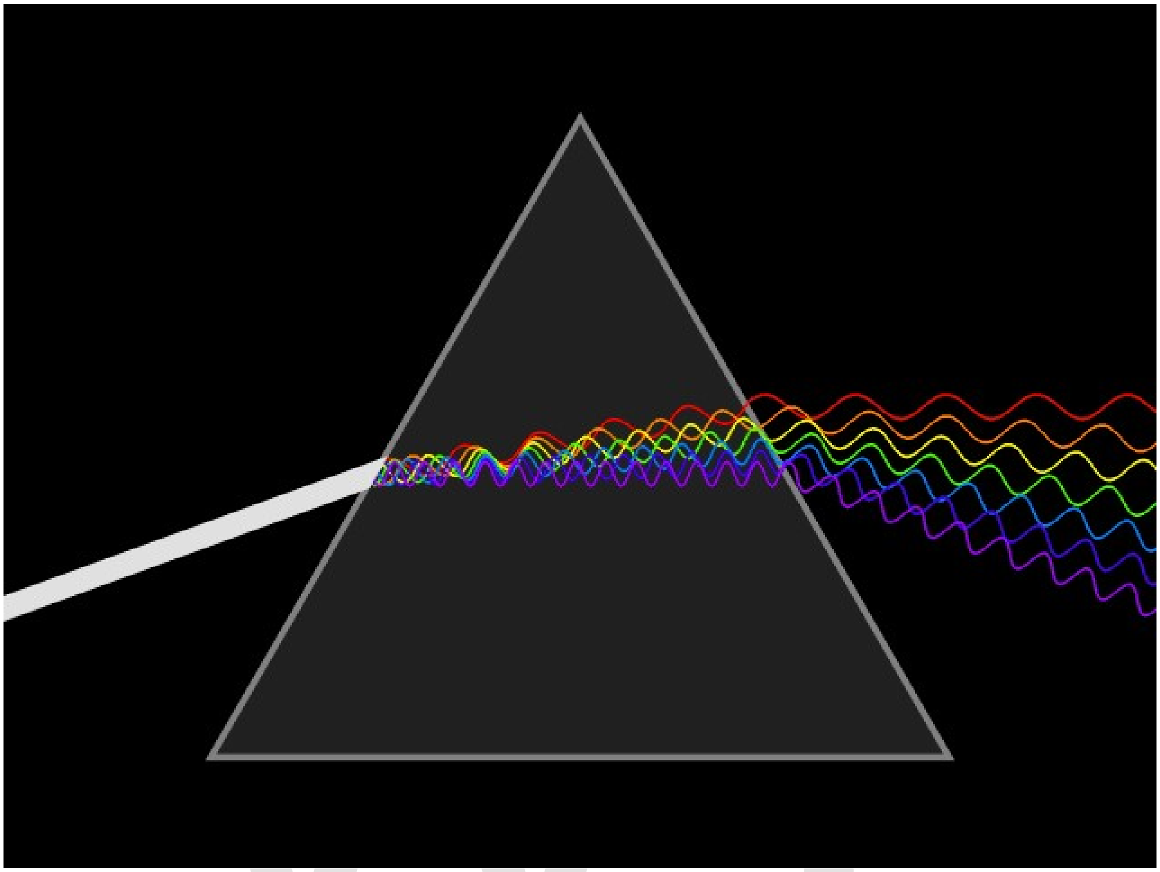
Since light is an oscillation it is not affected by travelling through static electric or magnetic fields in a linear medium such as a vacuum. However in nonlinear media, such as some crystals, interactions can occur between light and static electric and magnetic fields — these interactions include the Faraday effect and the Kerr effect.

In refraction, a wave crossing from one medium to another of different density alters its speed and direction upon entering the new medium. The ratio of the refractive indices of the media determines the degree of refraction, and is summarized by Snell's law. Light disperses into a visible spectrum as light is shone through a prism because of the wavelength dependent refractive index of the prism material (Dispersion).

EM radiation exhibits both wave properties and particle properties at the same time. Both wave and particle characteristics have been confirmed in a large number of experiments. Wave characteristics are more apparent when EM radiation is measured over relatively large timescales and over large distances while particle characteristics are more evident when measuring small timescales and distances. For example, when electromagnetic radiation is absorbed by matter, particle-like properties will be more obvious when the average number of photons in the cube of the relevant wavelength is much smaller than 1. Upon absorption of light, it is not too difficult to experimentally observe non-uniform deposition of energy. Strictly speaking, however, this alone is not evidence of "particulate" behavior of light, rather it reflects the quantum nature of *matter*.

There are experiments in which the wave and particle natures of electromagnetic waves appear in the same experiment, such as the self-interference of a single photon. *True* single-photon experiments (in a quantum optical sense) can be done today in undergraduate-level labs. When a single photon is sent through an interferometer, it passes through both paths, interfering with itself, as waves do, yet is detected by a photomultiplier or other sensitive detector only once.

Wave model



White light being separated into its components

Electromagnetic radiation is a transverse wave meaning that the oscillations of the waves are perpendicular to the direction of energy transfer and travel. An important aspect of the nature of light is frequency. The frequency of a wave is its rate of oscillation and is measured in hertz, the SI unit of frequency, where one hertz is equal to one oscillation per second. Light usually has a spectrum of frequencies which sum together to form the resultant wave. Different frequencies undergo different angles of refraction.

A wave consists of successive troughs and crests, and the distance between two adjacent crests or troughs is called the wavelength. Waves of the electromagnetic spectrum vary in size, from very long radio waves the size of buildings to very short gamma rays smaller than atom nuclei. Frequency is inversely proportional to wavelength, according to the equation:

$$v = f\lambda$$

where *v* is the speed of the wave (*c* in a vacuum, or less in other media), *f* is the frequency and λ is the wavelength. As waves cross boundaries between different media, their speeds change but their frequencies remain constant.

Interference is the superposition of two or more waves resulting in a new wave pattern. If the fields have components in the same direction, they constructively interfere, while opposite directions cause destructive interference.

The energy in electromagnetic waves is sometimes called radiant energy.

Particle model

Because energy of an EM wave is quantized, in the particle model of EM radiation, a wave consists of discrete packets of energy, or quanta, called photons. The frequency of the wave is proportional to the particle's energy. Because photons are emitted and absorbed by charged particles, they act as transporters of energy. The energy per photon can be calculated from the Planck–Einstein equation:

$$E = hf$$

where E is the energy, h is Planck's constant, and f is frequency. It is commonly expressed in the unit of electronvolt (eV). This photon-energy expression is a particular case of the energy levels of the more general *electromagnetic oscillator*, whose average energy, which is used to obtain Planck's radiation law, can be shown to differ sharply from that predicted by the equipartition principle at low temperature, thereby establishes a failure of equipartition due to quantum effects at low temperature.

As a photon is absorbed by an atom, it excites the atom, elevating an electron to a higher energy level. If the energy is great enough, so that the electron jumps to a high enough energy level, it may escape the positive pull of the nucleus and be liberated from the atom in a process called photoionisation. Conversely, an electron that descends to a lower energy level in an atom emits a photon of light equal to the energy difference. Since the energy levels of electrons in atoms are discrete, each element emits and absorbs its own characteristic frequencies.

Together, these effects explain the emission and absorption spectra of light. The dark bands in the absorption spectrum are due to the atoms in the intervening medium absorbing different frequencies of the light. The composition of the medium through which the light travels determines the nature of the absorption spectrum. For instance, dark bands in the light emitted by a distant star are due to the atoms in the star's atmosphere. These bands correspond to the allowed energy levels in the atoms. A similar phenomenon occurs for emission. As the electrons descend to lower energy levels, a spectrum is emitted that represents the jumps between the energy levels of the electrons. This is manifested in the emission spectrum of nebulae. Today, scientists use this phenomenon to observe what elements a certain star is composed of. It is also used in the determination of the distance of a star, using the red shift.

Speed of propagation

Any electric charge which accelerates, or any changing magnetic field, produces electromagnetic radiation. Electromagnetic information about the charge travels at the speed of light. Accurate treatment thus incorporates a concept known as retarded time (as opposed to advanced time, which is unphysical in light of causality), which adds to the expressions for the electrodynamic electric field and magnetic field. These extra terms are responsible for electromagnetic radiation. When any wire (or other conducting object such as an antenna) conducts alternating current, electromagnetic radiation is propagated at the same frequency as the electric current. At the quantum level, electromagnetic radiation is produced when the wavepacket of a charged particle oscillates or otherwise accelerates. Charged particles in a stationary state do not move, but a superposition of such states may result in oscillation, which is responsible for the phenomenon of radiative transition between quantum states of a charged particle.

Depending on the circumstances, electromagnetic radiation may behave as a wave or as particles. As a wave, it is characterized by a velocity (the speed of light), wavelength, and frequency. When considered as particles, they are known as photons, and each has an energy related to the frequency of the wave given by Planck's relation $E = hv$, where E is the energy of the photon, $h = 6.626 \times 10^{-34}$ J·s is Planck's constant, and v is the frequency of the wave.

One rule is always obeyed regardless of the circumstances: EM radiation in a vacuum always travels at the speed of light, *relative to the observer*, regardless of the observer's velocity. (This observation led to Albert Einstein's development of the theory of special relativity.)

In a medium (other than vacuum), velocity factor or refractive index are considered, depending on frequency and application. Both of these are ratios of the speed in a medium to speed in a vacuum.

Thermal radiation and electromagnetic radiation as a form of heat

The basic structure of matter involves charged particles bound together in many different ways. When electromagnetic radiation is incident on matter, it causes the charged particles to oscillate and gain energy. The ultimate fate of this energy depends on the situation. It could be immediately re-radiated and appear as scattered, reflected, or transmitted radiation. It may also get dissipated into other microscopic motions within the matter, coming to thermal equilibrium and manifesting itself as thermal energy in the material. With a few exceptions such as fluorescence, harmonic generation, photochemical reactions and the photovoltaic effect, absorbed electromagnetic radiation simply deposits its energy by heating the material. This happens both for infrared and non-infrared radiation. Intense radio waves can thermally burn living tissue and can cook food. In addition to infrared lasers, sufficiently intense visible and ultraviolet lasers can also easily set paper afire. Ionizing electromagnetic radiation can create high-speed electrons in a material and break chemical bonds, but after these electrons collide many times with other atoms in the material eventually most of the energy gets downgraded to thermal energy, this whole process happening in a tiny fraction of a second. That infrared radiation is a form of heat and other electromagnetic radiation is not, is a widespread misconception in physics. *Any* electromagnetic radiation can heat a material when it is absorbed.

The inverse or time-reversed process of absorption is responsible for thermal radiation. Much of the thermal energy in matter consists of random motion of charged particles, and this energy can be radiated away from the matter. The resulting radiation may subsequently be absorbed by another piece of matter, with the deposited energy heating the material. Radiation is an important mechanism of heat transfer.

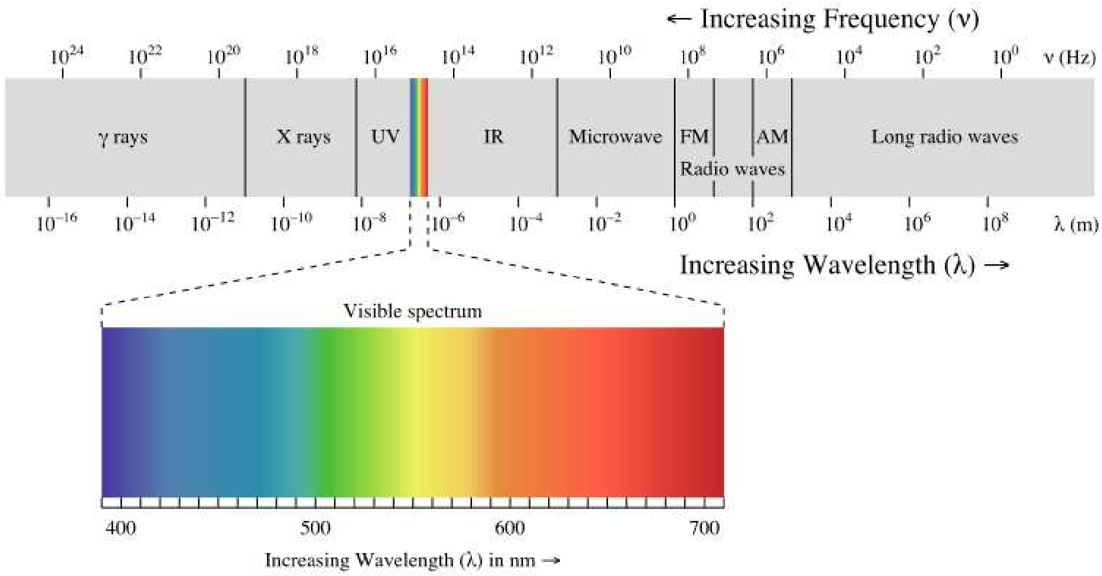
The electromagnetic radiation in an opaque cavity at thermal equilibrium is effectively a form of thermal energy, having maximum radiation entropy. The thermodynamic potentials of electromagnetic radiation can be well-defined as for matter. Thermal radiation in a cavity has energy density of

$$\frac{U}{V} = \frac{8\pi^5(kT)^4}{15(hc)^3},$$

Differentiating the above with respect to temperature, we may say that the electromagnetic radiation field has an effective volumetric heat capacity given by

$$\frac{32\pi^5 k^4 T^3}{15(hc)^3},$$

Electromagnetic spectrum



Electromagnetic spectrum with light highlighted

CLASS	FREQUENCY	WAVELENGTH	ENERGY
Y	300 EHz	1 pm	1.24 MeV
HX	30 EHz	10 pm	124 keV
	3 EHz	100 pm	12.4 keV
SX	300 PHz	1 nm	1.24 keV
EUV	30 PHz	10 nm	124 eV
NUV	3 PHz	100 nm	12.4 eV
NIR	300 THz	1 μm	1.24 eV
MIR	30 THz	10 μm	124 meV
FIR	3 THz	100 μm	12.4 meV
EHF	300 GHz	1 mm	1.24 meV
SHF	30 GHz	1 cm	124 μeV
UHF	3 GHz	1 dm	12.4 μeV
VHF	300 MHz	1 m	1.24 μeV
HF	30 MHz	10 m	124 neV
MF	3 MHz	100 m	12.4 neV
LF	300 kHz	1 km	1.24 neV
VLF	30 kHz	10 km	124 peV
VF/ULF	3 kHz	100 km	12.4 peV
SLF	300 Hz	1 Mm	1.24 peV
ELF	30 Hz	10 Mm	124 feV
	3 Hz	100 Mm	12.4 feV

Legend:
γ = Gamma rays
HX = Hard X-rays
SX = Soft X-Rays
EUV = Extreme ultraviolet
NUV = Near ultraviolet
Visible light
NIR = Near infrared
MIR = Moderate infrared
FIR = Far infrared

Radio waves:
EHF = Extremely high frequency (Microwaves)
SHF = Super high frequency (Microwaves)
UHF = Ultrahigh frequency

VHF = Very high frequency
HF = High frequency
MF = Medium frequency
LF = Low frequency
VLF = Very low frequency
VF = Voice frequency
ULF = Ultra low frequency
SLF = Super low frequency
ELF = Extremely low frequency

Generally, EM radiation (the designation 'radiation' excludes static electric and magnetic and near fields) is classified by wavelength into radio, microwave, infrared, the visible region we perceive as light, ultraviolet, X-rays and gamma rays. Arbitrary electromagnetic waves can always be expressed by Fourier analysis in terms of sinusoidal monochromatic waves which can be classified into these regions of the spectrum.

The behavior of EM radiation depends on its wavelength. Higher frequencies have shorter wavelengths, and lower frequencies have longer wavelengths. When EM radiation interacts with single atoms and molecules, its behavior depends on the amount of energy per quantum it carries. Spectroscopy can detect a much wider region of the EM spectrum than the visible range of 400 nm to 700 nm. A common laboratory spectroscope can detect wavelengths from 2 nm to 2500 nm. Detailed information about the physical properties of objects, gases, or even stars can be obtained from this type of device. It is widely used in astrophysics. For example, hydrogen atoms emit radio waves of wavelength 21.12 cm.

Soundwaves are not electromagnetic radiation. At the lower end of the electromagnetic spectrum, about 20 Hz to about 20kHz, are frequencies that might be considered in the audio range, however, electromagnetic waves cannot be directly perceived by human ears. Sound waves are the osscilating compression of molecules. To be heard, electromagnetic radiation must be converted to air pressure waves, or if the ear is submerged, water pressure waves.

Light

EM radiation with a wavelength between approximately 400 nm and 700 nm is directly detected by the human eye and perceived as visible light. Other wavelengths, especially nearby infrared (longer than 700 nm) and ultraviolet (shorter than 400 nm) are also sometimes referred to as light, especially when visibility to humans is not relevant.

If radiation having a frequency in the visible region of the EM spectrum reflects off of an object, say, a bowl of fruit, and then strikes our eyes, this results in our visual perception of the scene. Our brain's visual system processes the multitude of reflected frequencies into different shades and hues, and through this not-entirely-understood psychophysical phenomenon, most people perceive a bowl of fruit.

At most wavelengths, however, the information carried by electromagnetic radiation is not directly detected by human senses. Natural sources produce EM radiation across the spectrum, and our technology can also manipulate a broad range of wavelengths. Optical fiber transmits light which, although not suitable for direct viewing, can carry data that can be translated into sound or an image. The coding used in such data is similar to that used with radio waves.

Radio waves

Radio waves can be made to carry information by varying a combination of the amplitude, frequency and phase of the wave within a frequency band.

When EM radiation impinges upon a conductor, it couples to the conductor, travels along it, and induces an electric current on the surface of that conductor by exciting the electrons of the conducting material. This effect (the skin effect) is used in antennas. EM radiation may also cause certain molecules to absorb energy and thus to heat up; this is exploited in microwave ovens.

Derivation

Electromagnetic waves as a general phenomenon were predicted by the classical laws of electricity and magnetism, known as Maxwell's equations. If you inspect Maxwell's equations without sources (charges or currents) then you will find that, along with the possibility of nothing happening, the theory will also admit nontrivial solutions of changing electric and magnetic fields. Beginning with Maxwell's equations for free space:

$$\nabla \cdot \mathbf{E} = 0 \qquad (1)$$
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad (2)$$
$$\nabla \cdot \mathbf{B} = 0 \qquad (3)$$
$$\nabla \times \mathbf{B} = \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \qquad (4)$$

where
∇ is a vector differential operator.

One solution,

$$\mathbf{E} = \mathbf{B} = \mathbf{0},$$

is trivial.

To see the more interesting one, we utilize vector identities, which work for any vector, as follows:

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$$

To see how we can use this take the curl of equation (2):

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t}\right) \quad (5)$$

Evaluating the left hand side:

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla^2 \mathbf{E} \quad (6)$$

where we simplified the above by using equation (1).

Evaluate the right hand side:

$$\nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t}\right) = -\frac{\partial}{\partial t} (\nabla \times \mathbf{B}) = -\mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad (7)$$

Equations (6) and (7) are equal, so this results in a vector-valued differential equation for the electric field, namely

$$\boxed{\nabla^2 \mathbf{E} = \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}}$$

Applying a similar pattern results in similar differential equation for the magnetic field:

$$\boxed{\nabla^2 \mathbf{B} = \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{B}}{\partial t^2}.}$$

These differential equations are equivalent to the wave equation:

$$\nabla^2 f = \frac{1}{{c_0}^2} \frac{\partial^2 f}{\partial t^2}$$

where

c_0 is the speed of the wave in free space and

f describes a displacement

Or more simply:

$$\Box f = 0$$

where $\Box$ is d'Alembertian:

$$\Box = \nabla^2 - \frac{1}{{c_0}^2}\frac{\partial^2}{\partial t^2} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{{c_0}^2}\frac{\partial^2}{\partial t^2}$$

Notice that in the case of the electric and magnetic fields, the speed is:

$$c_0 = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$$

Which, as it turns out, is the speed of light in free space. Maxwell's equations have unified the permittivity of free space ε_0, the permeability of free space μ_0, and the speed of light itself, c_0. Before this derivation it was not known that there was such a strong relationship between light and electricity and magnetism.

But these are only two equations and we started with four, so there is still more information pertaining to these waves hidden within Maxwell's equations. Let's consider a generic vector wave for the electric field.

$$\mathbf{E} = \mathbf{E}_0 f\left(\hat{\mathbf{k}} \cdot \mathbf{x} - c_0 t\right)$$

Here $\mathbf{E}_0$is the constant amplitude, f is any second differentiable function, $\hat{\mathbf{k}}$is a unit vector in the direction of propagation, and $\mathbf{x}$is a position vector. We observe that $f\left(\hat{\mathbf{k}} \cdot \mathbf{x} - c_0 t\right)$is a generic solution to the wave equation. In other words

$$\nabla^2 f\left(\hat{\mathbf{k}} \cdot \mathbf{x} - c_0 t\right) = \frac{1}{{c_0}^2}\frac{\partial^2}{\partial t^2} f\left(\hat{\mathbf{k}} \cdot \mathbf{x} - c_0 t\right),$$

for a generic wave traveling in the $\hat{\mathbf{k}}$direction.

This form will satisfy the wave equation, but will it satisfy all of Maxwell's equations, and with what corresponding magnetic field?

$$\nabla \cdot \mathbf{E} = \hat{\mathbf{k}} \cdot \mathbf{E}_0 f'\left(\hat{\mathbf{k}} \cdot \mathbf{x} - c_0 t\right) = 0$$
$$\mathbf{E} \cdot \hat{\mathbf{k}} = 0$$

The first of Maxwell's equations implies that electric field is orthogonal to the direction the wave propagates.

$$\nabla \times \mathbf{E} = \hat{\mathbf{k}} \times \mathbf{E}_0 f'\left(\hat{\mathbf{k}} \cdot \mathbf{x} - c_0 t\right) = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\mathbf{B} = \frac{1}{c_0}\hat{\mathbf{k}} \times \mathbf{E}$$

The second of Maxwell's equations yields the magnetic field. The remaining equations will be satisfied by this choice of $\mathbf{E}, \mathbf{B}$.

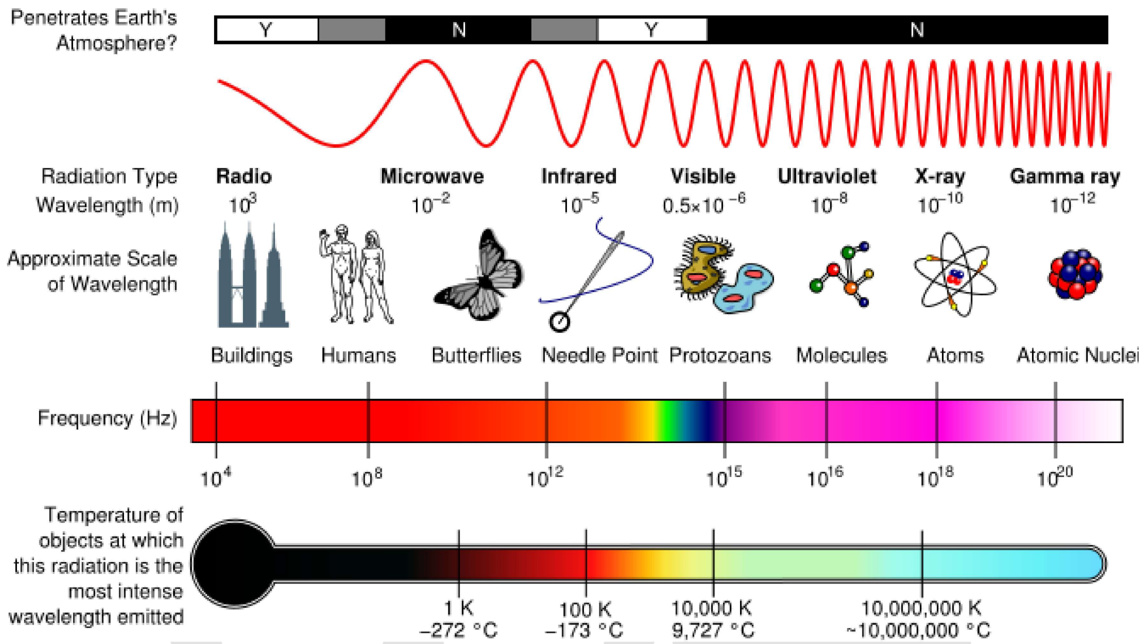
Not only are the electric and magnetic field waves traveling at the speed of light, but they have a special restricted orientation and proportional magnitudes, $E_0 = c_0 B_0$, which can be seen immediately from the Poynting vector. The electric field, magnetic field, and direction of wave propagation are all orthogonal, and the wave propagates in the same direction as $\mathbf{E} \times \mathbf{B}$.

From the viewpoint of an electromagnetic wave traveling forward, the electric field might be oscillating up and down, while the magnetic field oscillates right and left; but this picture can be rotated with the electric field oscillating right and left and the magnetic field oscillating down and up. This is a different solution that is traveling in the same direction. This arbitrariness in the orientation with respect to propagation direction is known as polarization. On a quantum level, it is described as photon polarization.

Electromagnetic Spectrum

The **electromagnetic spectrum** is the range of all possible frequencies of electromagnetic radiation. The "electromagnetic spectrum" of an object is the characteristic distribution of electromagnetic radiation emitted or absorbed by that particular object.

The electromagnetic spectrum extends from low frequencies used for modern radio to gamma radiation at the short-wavelength end, covering wavelengths from thousands of kilometers down to a fraction of the size of an atom. The long wavelength limit is the size of the universe itself, while it is thought that the short wavelength limit is in the vicinity of the Planck length, although in principle the spectrum is infinite and continuous.



Although some radiations are marked as "*N*" for "*no*" in the diagram, some waves do in fact penetrate the atmosphere, although extremely minimally compared to the other radiations.

CLASS	FREQUENCY	WAVELENGTH	ENERGY
γ	300 EHz	1 pm	1.24 MeV
HX	30 EHz	10 pm	124 keV
SX	3 EHz	100 pm	12.4 keV
	300 PHz	1 nm	1.24 keV
EUV	30 PHz	10 nm	124 eV
NUV	3 PHz	100 nm	12.4 eV
NIR	300 THz	1 μm	1.24 eV
MIR	30 THz	10 μm	124 meV
FIR	3 THz	100 μm	12.4 meV
EHF	300 GHz	1 mm	1.24 meV
SHF	30 GHz	1 cm	124 μeV
UHF	3 GHz	1 dm	12.4 μeV
VHF	300 MHz	1 m	1.24 μeV
HF	30 MHz	10 m	124 neV
MF	3 MHz	100 m	12.4 neV
LF	300 kHz	1 km	1.24 neV
VLF	30 kHz	10 km	124 peV
VF/ULF	3 kHz	100 km	12.4 peV
SLF	300 Hz	1 Mm	1.24 peV
ELF	30 Hz	10 Mm	124 feV
	3 Hz	100 Mm	12.4 feV

Legend

γ= Gamma rays
HX= Hard X-rays
SX= Soft X-rays
EUV= Extreme ultraviolet
NUV= Near ultraviolet
Visible light
NIR= Near Infrared

MIR= Mid infrared
FIR= Far infrared
Radio waves
EHF= Extremely high freq.
SHF= Super high freq.
UHF= Ultra high freq.
VHF= Very high freq.

HF= High freq.
MF= Medium freq.
LF= Low freq.
VLF= Very low freq.
VF/ULF= Voice freq.
SLF= Super low freq.
ELF= Extremely low freq.
Freq=Frequency

Range of the spectrum

EM waves are typically described by any of the following three physical properties: the frequency *f*, wavelength λ, or photon energy *E*. Frequencies range from 2.4×10^{23} Hz (1 GeV gamma rays) down to the local plasma frequency of the ionized interstellar medium (~1 kHz). Wavelength is inversely proportional to the wave frequency, so gamma rays have very short wavelengths that are fractions of the size of atoms, whereas wavelengths can be as long as the universe. Photon energy is directly proportional to the wave frequency, so gamma rays have the highest energy (around a billion electron volts) and radio waves have very low energy (around femto electron volts). These relations are illustrated by the following equations:

$$f = \frac{c}{\lambda}, \quad \text{or} \quad f = \frac{E}{h}, \quad \text{or} \quad E = \frac{hc}{\lambda},$$

where:

- c = 299,792,458 m/s is the speed of light in vacuum and
- $h = 6.62606896(33)\times10^{-34}$ J s $= 4.13566733(10)\times10^{-15}$ eV s is Planck's constant.

Whenever electromagnetic waves exist in a medium with matter, their wavelength is decreased. Wavelengths of electromagnetic radiation, no matter what medium they are traveling through, are usually quoted in terms of the *vacuum wavelength*, although this is not always explicitly stated.

Generally, EM radiation is classified by wavelength into radio wave, microwave, infrared, the visible region we perceive as light, ultraviolet, X-rays and gamma rays. The behavior of EM radiation depends on its wavelength. When EM radiation interacts with single atoms and molecules, its behavior also depends on the amount of energy per quantum (photon) it carries.

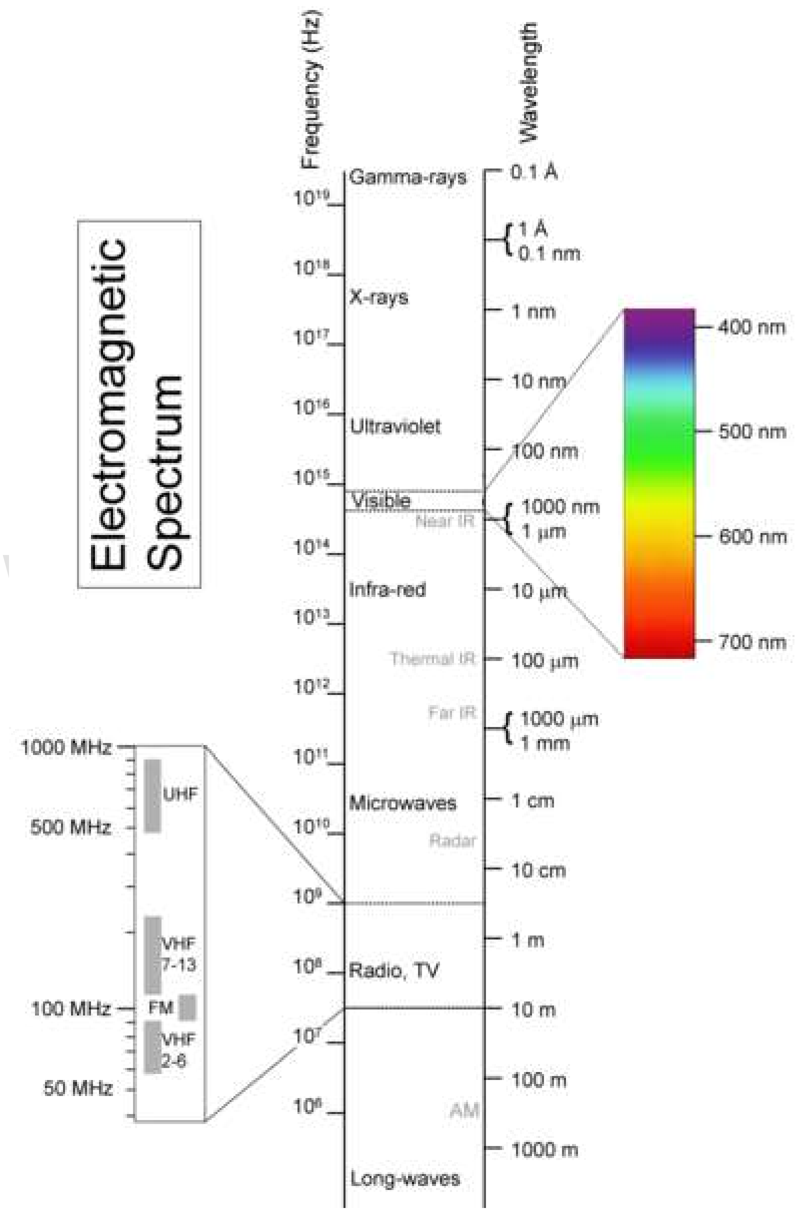
Spectroscopy can detect a much wider region of the EM spectrum than the visible range of 400 nm to 700 nm. A common laboratory spectroscope can detect wavelengths from 2 nm to 2500 nm. Detailed information about the physical properties of objects, gases, or even stars can be obtained from this type of device. Spectroscopes are widely used in astrophysics. For example, many hydrogen atoms emit a radio wave photon which has a wavelength of 21.12 cm. Also, frequencies of 30 Hz and below can be produced by and are important in the study of certain stellar nebulae and frequencies as high as 2.9×10^{27} Hz have been detected from astrophysical sources.

Rationale

Electromagnetic radiation interacts with matter in different ways in different parts of the spectrum. The types of interaction can be so different that it seems to be justified to refer to different types of radiation. At the same time, there is a continuum containing all these "different kinds" of electromagnetic radiation. Thus we refer to a spectrum, but divide it up based on the different interactions with matter.

Region of the spectrum	Main interactions with matter
Radio	Collective oscillation of charge carriers in bulk material (plasma oscillation). An example would be the oscillation of the electrons in an antenna.
Microwave through far infrared	Plasma oscillation, molecular rotation
Near infrared	Molecular vibration, plasma oscillation (in metals only)
Visible	Molecular electron excitation (including pigment molecules found in the human retina), plasma oscillations (in metals only)
Ultraviolet	Excitation of molecular and atomic valence electrons, including ejection of the electrons (photoelectric effect)
X-rays	Excitation and ejection of core atomic electrons, Compton scattering (for low atomic numbers)
Gamma rays	Energetic ejection of core electrons in heavy elements, Compton scattering (for all atomic numbers), excitation of atomic nuclei, including dissociation of nuclei
High energy gamma rays	Creation of particle-antiparticle pairs. At very high energies a single photon can create a shower of high energy particles and antiparticles upon interaction with matter.

Types of radiation



The electromagnetic spectrum

While the classification scheme is generally accurate, in reality there is often some overlap between neighboring types of electromagnetic energy. For example, SLF radio waves at 60 Hz may be received and studied by astronomers, or may be ducted along wires as electric power, although the latter is, strictly speaking, not electromagnetic radiation at all. The distinction between X and gamma rays is based on sources: gamma rays are the photons generated from nuclear decay or other nuclear and subnuclear/particle process, whereas X-rays are generated by electronic transitions involving highly energetic inner atomic electrons. Generally, nuclear transitions are much more energetic than electronic transitions, so usually, gamma-rays are more energetic than X-rays, but exceptions exist. By analogy to electronic transitions, muonic atom transitions are also said to produce X-rays, even though their energy may exceed 6 megaelectronvolts (0.96 pJ), whereas there are many (77 known to be less than 10 keV (1.6 fJ)) low-energy nuclear transitions (e.g. the 7.6 eV (1.22 aJ) nuclear transition of thorium-229), and despite being one million-fold less energetic than some muonic X-rays, the emitted photons are still called gamma rays due to their nuclear origin.

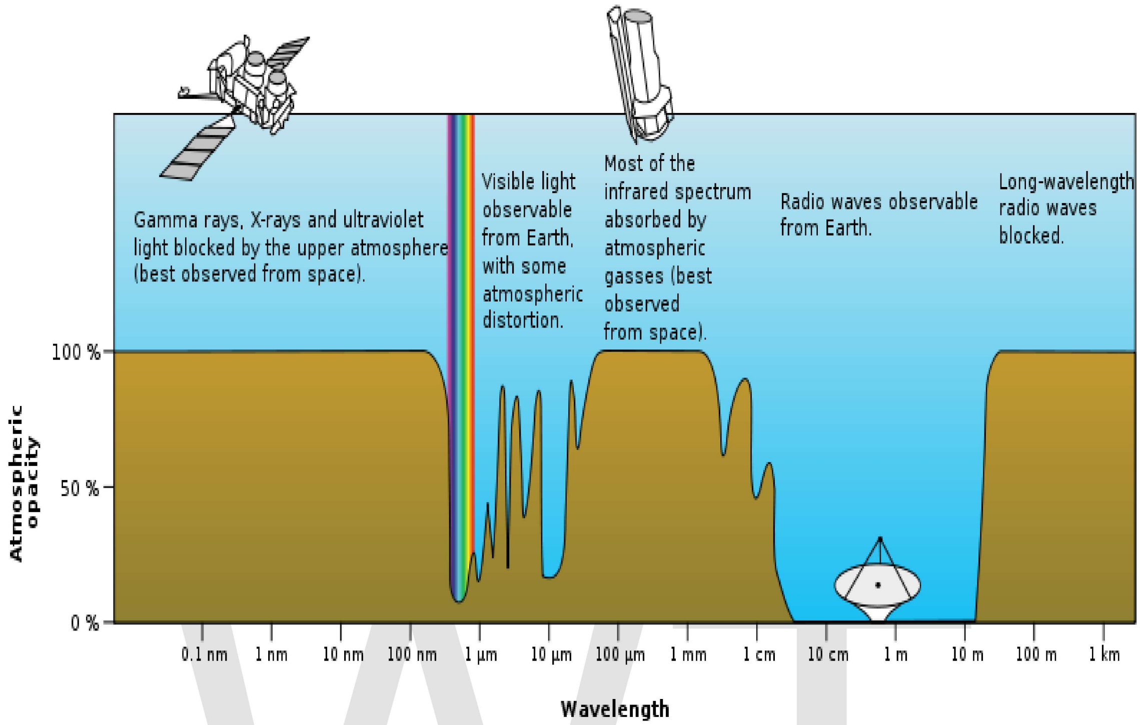
Also, the region of the spectrum of the particular electromagnetic radiation is reference-frame dependent (on account of the Doppler shift for light) so EM radiation which one observer would say is in one region of the spectrum could appear to an observer moving at a substantial fraction of the speed of light with respect to the first to be in another part of the spectrum. For example, consider the cosmic microwave background. It was produced, when matter and radiation decoupled, by the de-excitation of hydrogen atoms to the ground state. These photons were from Lyman series transitions, putting them in the ultraviolet (UV) part of the electromagnetic spectrum. Now this radiation has undergone enough cosmological red shift to put it into the microwave region of the spectrum for observers moving slowly (compared to the speed of light) with respect to the cosmos. However, for particles moving near the speed of light, this radiation will be blue-shifted in their rest frame. The highest energy cosmic ray protons are moving such that, in their rest frame, this radiation is blueshifted to high energy gamma rays which interact with the proton to produce bound quark-antiquark pairs (pions). This is the source of the GZK limit.

Radio frequency

Radio waves generally are utilized by antennas of appropriate size (according to the principle of resonance), with wavelengths ranging from hundreds of meters to about one millimeter. They are used for transmission of data, via modulation. Television, mobile phones, wireless networking and amateur radio all use radio waves. The use of the radio spectrum is regulated by many governments through frequency allocation.

Radio waves can be made to carry information by varying a combination of the amplitude, frequency and phase of the wave within a frequency band. When EM radiation impinges upon a conductor, it couples to the conductor, travels along it, and induces an electric current on the surface of that conductor by exciting the electrons of the conducting material. This effect (the skin effect) is used in antennas. **EM** radiation may also cause certain molecules to absorb energy and thus to heat up, causing thermal effects and sometimes burns. This is exploited in microwave ovens.

Microwaves



Plot of Earth's atmospheric transmittance (or opacity) to various wavelengths of electromagnetic radiation.

The super high frequency (SHF) and extremely high frequency (EHF) of microwaves come next up the frequency scale. Microwaves are waves which are typically short enough to employ tubular metal waveguides of reasonable diameter. Microwave energy is produced with klystron and magnetron tubes, and with solid state diodes such as Gunn and IMPATT devices. Microwaves are absorbed by molecules that have a dipole moment in liquids. In a microwave oven, this effect is used to heat food. Low-intensity microwave radiation is used in Wi-Fi, although this is at intensity levels unable to cause thermal heating.

Volumetric heating, as used by microwaves, transfer energy through the material electromagnetically, not as a thermal heat flux. The benefit of this is a more uniform heating and reduced heating time; microwaves can heat material in less than 1% of the time of conventional heating methods.

When active, the average microwave oven is powerful enough to cause interference at close range with poorly shielded electromagnetic fields such as those found in mobile medical devices and cheap consumer electronics.

Terahertz radiation

Terahertz radiation is a region of the spectrum between far infrared and microwaves. Until recently, the range was rarely studied and few sources existed for microwave energy at the high end of the band (sub-millimetre waves or so-called terahertz waves), but applications such as imaging and communications are now appearing. Scientists are also looking to apply terahertz technology in the armed forces, where high frequency waves might be directed at enemy troops to incapacitate their electronic equipment.

Infrared radiation

The infrared part of the electromagnetic spectrum covers the range from roughly 300 GHz (1 mm) to 400 THz (750 nm). It can be divided into three parts:

- **Far-infrared**, from 300 GHz (1 mm) to 30 THz (10 μm). The lower part of this range may also be called microwaves. This radiation is typically absorbed by so-called rotational modes in gas-phase molecules, by molecular motions in liquids, and by phonons in solids. The water in the Earth's atmosphere absorbs so strongly in this range that it renders the atmosphere effectively opaque. However, there are certain wavelength ranges ("windows") within the opaque range which allow partial transmission, and can be used for astronomy. The wavelength range from approximately 200 μm up to a few mm is often referred to as "sub-millimetre" in astronomy, reserving far infrared for wavelengths below 200 μm.
- **Mid-infrared**, from 30 to 120 THz (10 to 2.5 μm). Hot objects (black-body radiators) can radiate strongly in this range. It is absorbed by molecular vibrations, where the different atoms in a molecule vibrate around their equilibrium positions. This range is sometimes called the *fingerprint region* since the mid-infrared absorption spectrum of a compound is very specific for that compound.
- **Near-infrared**, from 120 to 400 THz (2,500 to 750 nm). Physical processes that are relevant for this range are similar to those for visible light.

Visible radiation (light)

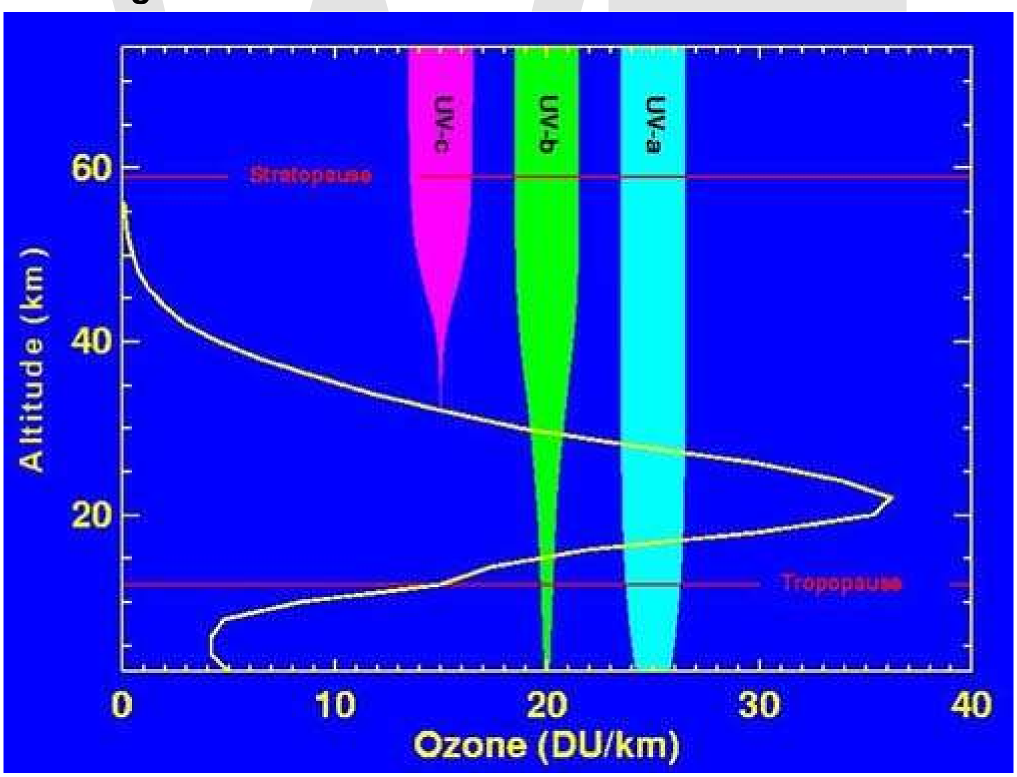
Above infrared in frequency comes visible light. This is the range in which the sun and stars similar to it emit most of their radiation. It is probably not a coincidence that the human eye is sensitive to the wavelengths that the sun emits most strongly. Visible light (and near-infrared light) is typically absorbed and emitted by electrons in molecules and atoms that move from one energy level to another. The light we see with our eyes is really a very small portion of the electromagnetic spectrum. A rainbow shows the optical (visible) part of the electromagnetic spectrum; infrared (if you could see it) would be located just beyond the red side of the rainbow with ultraviolet appearing just beyond the violet end.

Electromagnetic radiation with a wavelength between 380 nm and 760 nm (790–400 terahertz) is detected by the human eye and perceived as visible light. Other wavelengths, especially near infrared (longer than 760 nm) and ultraviolet (shorter than 380 nm) are also sometimes referred to as light, especially when the visibility to humans is not relevant.

If radiation having a frequency in the visible region of the EM spectrum reflects off an object, say, a bowl of fruit, and then strikes our eyes, this results in our visual perception of the scene. Our brain's visual system processes the multitude of reflected frequencies into different shades and hues, and through this not-entirely-understood psychophysical phenomenon, most people perceive a bowl of fruit.

At most wavelengths, however, the information carried by electromagnetic radiation is not directly detected by human senses. Natural sources produce EM radiation across the spectrum, and our technology can also manipulate a broad range of wavelengths. Optical fiber transmits light which, although not suitable for direct viewing, can carry data that can be translated into sound or an image. The coding used in such data is similar to that used with radio waves.

Ultraviolet light



The amount of penetration of UV relative to altitude in Earth's ozone

Next in frequency comes ultraviolet (UV). This is radiation whose wavelength is shorter than the violet end of the visible spectrum, and longer than that of an X-ray.

Being very energetic, UV can break chemical bonds, making molecules unusually reactive or ionizing them, in general changing their mutual behavior. Sunburn, for example, is caused by the disruptive effects of UV radiation on skin cells, which is the main cause of skin cancer, if the radiation irreparably damages the complex DNA molecules in the cells (UV radiation is a proven mutagen). The Sun emits a large amount of UV radiation, which could quickly turn Earth into a barren desert. However, most of it is absorbed by the atmosphere's ozone layer before reaching the surface.

X-rays

After UV come X-rays, which are also ionizing, but due to their higher energies they can also interact with matter by means of the Compton effect. Hard X-rays have shorter wavelengths than soft X-rays. As they can pass through most substances, X-rays can be used to 'see through' objects, most notably diagnostic X-ray images in medicine (a process known as radiography), as well as for high-energy physics and astronomy. Neutron stars and accretion disks around black holes emit X-rays, which enable us to study them. X-rays are given off by stars and are strongly emitted by some types of nebulae.

Gamma rays

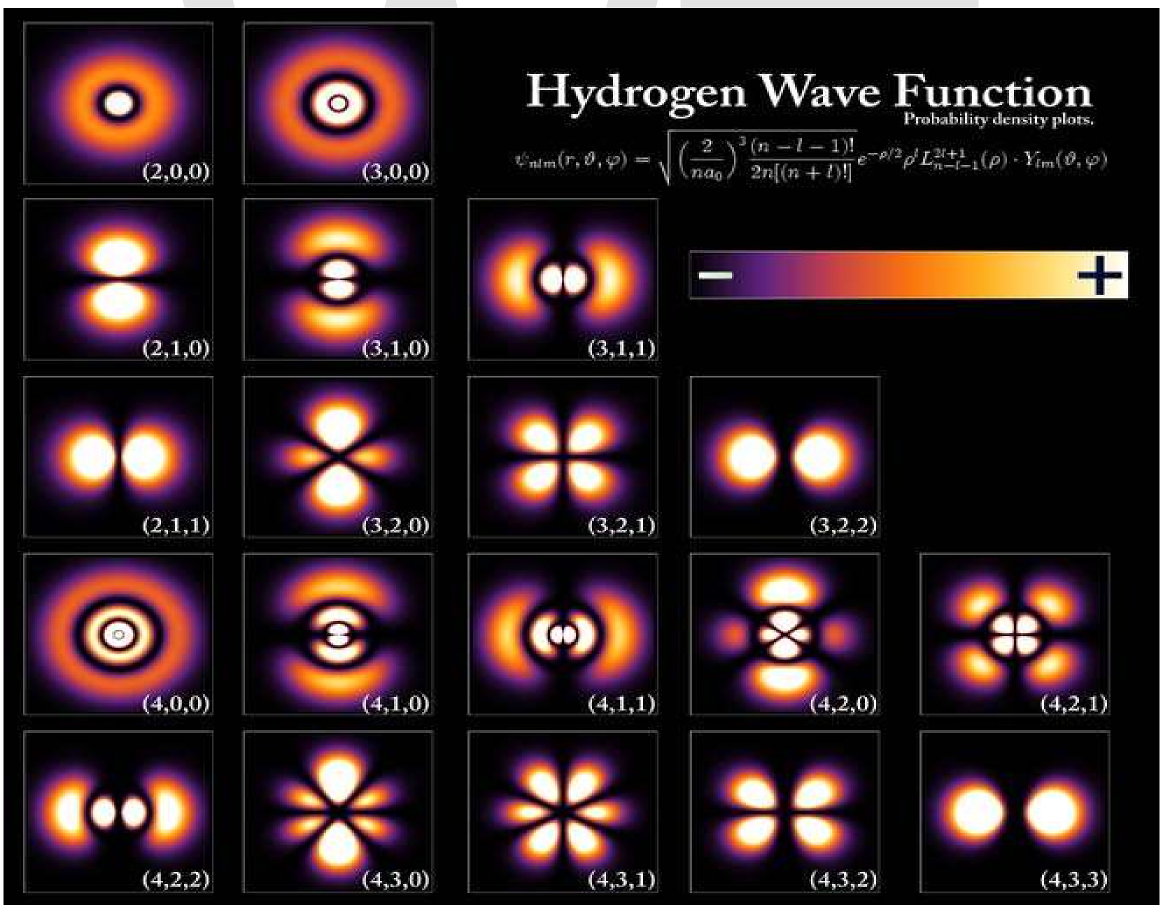
After hard X-rays come gamma rays, which were discovered by Paul Villard in 1900. These are the most energetic photons, having no defined lower limit to their wavelength. They are useful to astronomers in the study of high energy objects or regions, and find a use with physicists thanks to their penetrative ability and their production from radioisotopes. Gamma rays are also used for the irradiation of food and seed for sterilization, and in medicine they are used in radiation cancer therapy and some kinds of diagnostic imaging such as PET scans. The wavelength of gamma rays can be measured with high accuracy by means of Compton scattering.

Note that there are no precisely defined boundaries between the bands of the electromagnetic spectrum. Radiation of some types have a mixture of the properties of those in two regions of the spectrum. For example, red light resembles infrared radiation in that it can resonate some chemical bonds.

Chapter 7

Wave Function

A **wave function** or **wavefunction** is a mathematical tool used in quantum mechanics. It is a function typically of space or momentum or spin and possibly of time that returns the probability amplitude of a position or momentum for a subatomic particle. Mathematically, it is a function from a space that maps the possible states of the system into the complex numbers. The laws of quantum mechanics (the Schrödinger equation) describe how the wave function evolves over time.



The electron probability density for the first few hydrogen atom electron orbitals shown as cross-sections. These orbitals form an orthonormal basis for the wave function of the electron. Different orbitals are depicted with different scale.

It is commonly applied as a property of particles relating to their wave-particle duality, where it is denoted ψ(*position*,*time*) and where $|\psi|^2$ is equal to the chance of finding the subject at a certain time and position. For example, in an atom with a single electron, such as hydrogen or ionized helium, the wave function of the electron provides a complete description of how the electron behaves. It can be decomposed into a series of atomic orbitals which form a basis for the possible wave functions. For atoms with more than one electron (or any system with multiple particles), the underlying space is the possible configurations of all the electrons and the wave function describes the probabilities of those configurations.

A simple wave function is that for a particle in a box. Another simple example is a free particle (or a particle in a large box), whose wave function is a sinusoidal where, in the spirit of the uncertainty principle, the momentum is known but the position is not known.

Definition

The modern usage of the term *wave function* refers to a complex vector or function, i.e. an element in a complex Hilbert space. Typically, a wave function is either:

- a complex vector with finitely many components

$$\vec{\psi} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix},$$

- a complex vector with infinitely many components

$$\vec{\psi} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \\ \vdots \end{bmatrix},$$

- a complex function of one or more real variables (a *continuously indexed* complex vector)

$$\psi(x_1, \ldots, x_n).$$

In all cases, the wave function provides a complete description of the associated physical system. An element of a vector space can be expressed in different bases; and so the same applies to wave functions. The components of a wave function describing the same physical state take different complex values depending on the basis being used; however the wave function itself is not dependent on the basis chosen. In this respect they are like spatial vectors in ordinary space because choosing a new set of cartesian axes by rotation of the coordinate frame does not alter the vector itself, only the *representation* of the vector with respect to the coordinate frame. A basis in quantum mechanics is analogous to the coordinate frame in that choosing a new basis does not alter the wavefunction, only its representation, which is expressed as the values of the components above.

Because the probabilities that the system is in each possible state should add up to 1, the norm of the wave function must be 1.

Spatial interpretation

The physical interpretation of the wave function is context dependent. Several examples are provided below, followed by a detailed discussion of the three cases described above.

One particle in one spatial dimension

The spatial wave function associated with a particle in one dimension is a complex function $\psi(x)$ defined over the real line. The positive function $|\psi|^2$ is interpreted as the probability density associated with the particle's position. That is, the probability of a measurement of the particle's position yielding a value in the interval [*a*,*b*] is given by

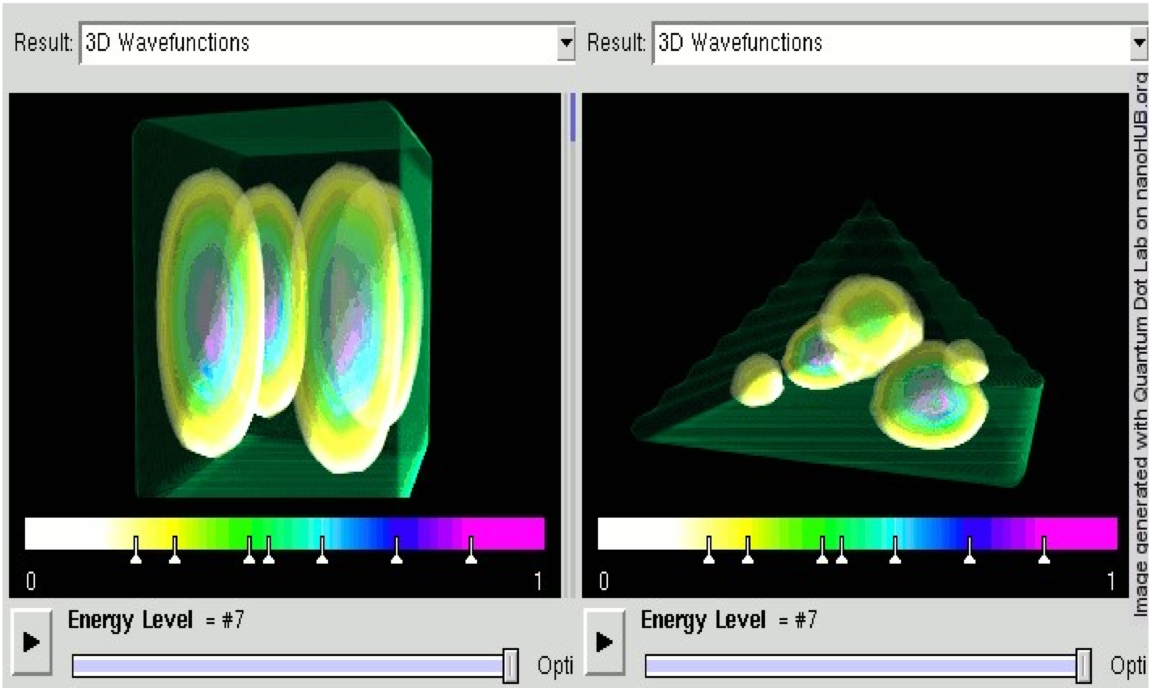
$$\mathbf{P}_{ab} = \int_a^b |\psi(x)|^2 \, dx .$$

This leads to the normalization condition

$$\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx = 1 .$$

since the probability of a measurement of the particle's position yielding a value in the range $(-\infty, \infty)$ is unity.

One particle in three spatial dimensions



3D confined electron wave functions in a Quantum Dot. Here, rectangular and triangular-shaped quantum dots are shown. Energy states in rectangular dots are more 's-type' and 'p-type'. However, in a triangular dot the wave functions are mixed due to confinement symmetry.

The three dimensional case is analogous to the one dimensional case; the wave function is a complex function $\psi(x, y, z)$ defined over three dimensional space, and the square of its absolute value is interpreted as a three dimensional probability density function:

$$\mathbf{P}_R = \int_R |\psi(x, y, z)|^2 \, dV$$

The normalization condition is likewise

$$\int_R |\psi(x, y, z)|^2 \, dV = 1$$

where the preceding integral is taken over all space.

Two distinguishable particles in three spatial dimensions

In this case, the wave function is a complex function of *six* spatial variables, $\psi(x_1, y_1, z_1, x_2, y_2, z_2)$, and $|\psi|^2$ is the joint probability density associated with the positions of both particles. Thus the probability that a measurement of the positions of *both particles* indicates particle one is in region R and particle two is in region S is

$$\mathbf{P}_{R,S} = \int_R \int_S |\psi|^2 \, dV_2 \, dV_1$$

where $dV_1 = dx_1 dy_1 dz_1$, and similarly for dV_2.

The normalization condition is then:

$$\int \int |\psi(x_1, y_1, z_1, x_2, y_2, z_2)|^2 \, dV_2 \, dV_1 = 1$$

in which the preceding integral is taken over the full range of all six variables.

Given a wave function ψ of a system consisting of two (or more) particles, it is in general not possible to assign a definite wave function to a single-particle subsystem. In other words, the particles in the system can be entangled.

One particle in one dimensional momentum space

The wave function for a one dimensional particle in momentum space is a complex function $\psi(p)$ defined over the real line. The quantity $|\psi|^2$ is interpreted as a probability density function in momentum space:

$$\mathbf{P}_{ab} = \int_a^b |\psi(p)|^2 \, dp$$

As in the position space case, this leads to the normalization condition:

$$\int_{-\infty}^{\infty} |\psi(p)|^2 \, dp = 1.$$

Spin 1/2

The wave function for a spin-½ particle (ignoring its spatial degrees of freedom) is a column vector

$$\vec{\psi} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

The meaning of the vector's components depends on the basis, but typically c_1 and c_2 are respectively the coefficients of spin up and spin down in the z direction. In Dirac notation this is:

$$|\psi\rangle = c_1|\uparrow_z\rangle + c_2|\downarrow_z\rangle$$

The values $|c_1|^2$ and $|c_2|^2$ are then respectively interpreted as the probability of obtaining spin up or spin down in the z direction when a measurement of the particle's spin is performed. This leads to the normalization condition

$$|c_1|^2 + |c_2|^2 = 1.$$

Interpretation

A wave function describes the state of a physical system, $|\psi\rangle$, by expanding it in terms of other possible states of the same system, $|\phi_i\rangle$. Collectively the latter are referred to as a *basis* or *representation*. In what follows, all wave functions are assumed to be normalized.

Finite dimensional basis vectors

A wave function which is a vector $\vec{\psi}$ with *n* components describes how to express the state of the physical system $|\psi\rangle$ as the linear combination of finitely many basis elements $|\phi_i\rangle$, where *i* runs from 1 to *n*. In particular the equation

$$\vec{\psi} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix},$$

which is a relation between column vectors, is equivalent to

$$|\psi\rangle = \sum_{i=1}^{n} c_i|\phi_i\rangle,$$

which is a relation between the states of a physical system. Note that to pass between these expressions one must know the basis in use, and hence, two column vectors with the same components can represent two different states of a system if their associated basis states are different. An example of a wave function which is a finite vector is furnished by the spin state of a spin-1/2 particle, as described above.

The physical meaning of the components of $\vec{\psi}$ is given by the wave function collapse postulate:

If the states $|\phi_i\rangle$ have distinct, definite values, λ_i, of some dynamical variable (e.g. momentum, position, etc) and a measurement of that variable is performed on a system in the state

$$|\psi\rangle = \sum_i c_i|\phi_i\rangle$$

then the probability of measuring λ_i is $|c_i|^2$, and if the measurement yields λ_i, the system is left in the state $|\phi_i\rangle$.

Infinite dimensional basis vectors

The case of an infinite vector with a discrete index is treated in the same manner a finite vector, except the sum is extended over all the basis elements. Hence

$$\vec{\psi} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \\ \vdots \end{bmatrix}$$

is equivalent to

$$|\psi\rangle = \sum_i c_i|\psi_i\rangle,$$

where it is understood that the above sum includes all the components of $\vec{\psi}$. The interpretation of the components is the same as the finite case (apply the collapse postulate).

Continuously indexed vectors (functions)

In the case of a continuous index, the sum is replaced by an integral; an example of this is the spatial wave function of a particle in one dimension, which expands the physical state of the particle, $|\psi\rangle$, in terms of states with definite position, $|x\rangle$. Thus

$$|\psi\rangle = \int_{-\infty}^{\infty} \psi(x)|x\rangle\, dx.$$

Note that $|\psi\rangle$ is *not* the same as $\psi(x)$. The former is the actual state of the particle, whereas the latter is simply a wave function describing how to express the former as a superposition of states with definite position. In this case the base states themselves can be expressed as

$$|x_0\rangle = \int_{-\infty}^{\infty} \delta(x - x_0)|x\rangle \, dx$$

and hence the spatial wave function associated with $|x_0\rangle$ is $\delta(x - x_0)$ (where $\delta(x - x_0)$ is the Dirac delta function).

Formalism

Given an isolated physical system, the allowed states of this system (i.e. the states the system could occupy without violating the laws of physics) are part of a Hilbert space *H*. Some properties of such a space are

1. If $|\psi\rangle$ and $|\phi\rangle$ are two allowed states, then
$a|\psi\rangle + b|\phi\rangle$
is also an allowed state, provided $|a|^2 + |b|^2 = 1$. (This condition is due to normalisation.)
2. There is always an orthonormal basis of allowed states of the vector space *H*.

The wave function associated with a particular state may be seen as an expansion of the state in a basis of *H*. For example,

$$\{|\uparrow_z\rangle, |\downarrow_z\rangle\}$$

is a basis for the space associated with the spin of a spin-1/2 particle and consequently the spin state of any such particle can be written uniquely as

$$a|\uparrow_z\rangle + b|\downarrow_z\rangle.$$

Sometimes it is useful to expand the state of a physical system in terms of states which are *not* allowed, and hence, not in *H*. An example of this is the spacial wave function associated with a particle in one dimension which expands the state of the particle in terms of states with definite position.

Every Hilbert space *H* is equipped with an inner product. Physically, the nature of the inner product is contingent upon the kind of basis in use. When the basis is a countable set $\{|\phi_i\rangle\}$, and orthonormal, i.e.

$$\langle\phi_i|\phi_j\rangle = \delta_{ij}.$$

Then an arbitrary vector $|\psi\rangle$ can be expressed as

$$|\psi\rangle = \sum_i c_i|\phi_i\rangle$$

where $c_i = \langle \phi_i | \psi \rangle$.

If one chooses a "continuous" basis as, for example, the *position* or *coordinate* basis consisting of all states of definite position $\{|x\rangle\}$, the orthonormality condition holds similarly:

$$\langle x | x' \rangle = \delta(x - x').$$

We have the analogous identity

$$\langle x | \int \psi(x') | x' \rangle \, dx' = \int \psi(x') \delta(x - x') \, dx' = \psi(x).$$

Ontology

Whether the wave function is real, and what it represents, are major questions in the interpretation of quantum mechanics. Many famous physicists of a previous generation once puzzled over this problem, such as Erwin Schrödinger, Albert Einstein and Niels Bohr. Some approaches regard it as merely representing information in the mind of the observer. Some, ranging from Schrödinger, Einstein, David Bohm and Hugh Everett III and others, argued that the wavefunction must have an objective existence.

de Broglie waves

Louis de Broglie postulated that all particles with momentum have a wavelength

$$\lambda = \frac{h}{p},$$

where h is Planck's constant, and p is the magnitude of the momentum of the particle. This hypothesis was at the basis of quantum mechanics. Nowadays, this wavelength is called the de Broglie wavelength. For example, the electrons in a CRT display have a de Broglie wavelength of about 10^{-13} m.

A wave representing such a particle traveling in the k-direction is expressed by the wave function:

$$\psi(\mathbf{r},\ t = 0) = A\ e^{i\mathbf{k}\cdot\mathbf{r}}\ ,$$

where the wavelength is determined by the wave vector **k** as:

$$\lambda = \frac{2\pi}{k}\ ,$$

and the momentum by:

$$\mathbf{p} = \hbar \mathbf{k} \ .$$

However, a wave like this with definite wavelength is not localized in space, and so cannot represent a particle localized in space. To localize a particle, de Broglie proposed a superposition of different wavelengths ranging around a central value in a wave packet, a waveform often used in quantum mechanics to describe the wave function of a particle. In a wave packet, the wavelength of the particle is not precise, and the local wavelength deviates on either side of the main wavelength value.

In representing the wave function of a localized particle, the wave packet is often taken to have a Gaussian shape and is called a *Gaussian wave packet*. Gaussian wave packets also are used to analyze water waves.

For example, a Gaussian wavefunction ψ might take the form:

$$\psi(x,\ t=0) = A\ \exp\left(-\frac{x^2}{2\sigma^2} + ik_0 x\right)\ ,$$

at some initial time $t = 0$, where the central wavelength is related to the central wave vector k_0 as $\lambda_0 = 2\pi / k_0$. It is well known from the theory of Fourier analysis, or from the Heisenberg uncertainty principle (in the case of quantum mechanics) that a narrow range of wavelengths is necessary to produce a localized wave packet, and the more localized the envelope, the larger the spread in required wavelengths. The Fourier transform of a Gaussian is itself a Gaussian. Given the Gaussian:

$$f(x) = e^{-x^2/(2\sigma^2)}\ ,$$

the Fourier transform is:

$$\tilde{f}(k) = \sigma e^{-\sigma^2 k^2/2}\ .$$

The Gaussian in space therefore is made up of waves:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty}\ \tilde{f}(k) e^{ikx}\ dk\ ;$$

that is, a number of waves of wavelengths λ such that $k\lambda = 2\ \pi$.

The parameter σ decides the spatial spread of the Gaussian along the x-axis, while the Fourier transform shows a spread in wave vector k determined by $1/\sigma$. That is, the smaller the extent in space, the larger the extent in k, and hence in $\lambda = 2\pi/k$.

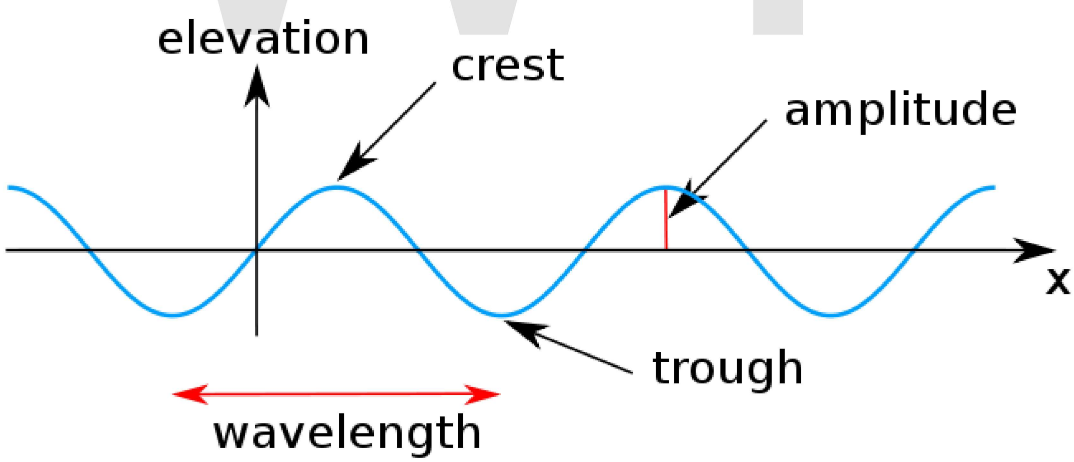
Chapter 8

Airy Wave Theory

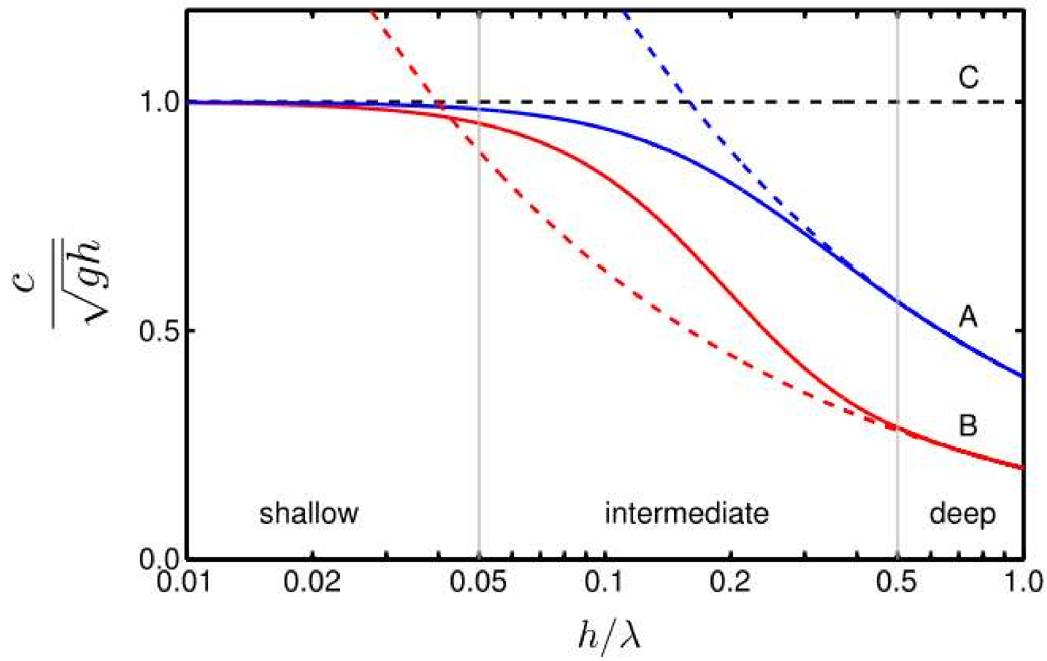
In fluid dynamics, **Airy wave theory** (often referred to as **linear wave theory**) gives a linearised description of the propagation of gravity waves on the surface of a homogeneous fluid layer. The theory assumes that the fluid layer has a uniform mean depth, and that the fluid flow is inviscid, incompressible and irrotational. This theory was first published, in correct form, by George Biddell Airy in the 19th century.

Airy wave theory is often applied in ocean engineering and coastal engineering for the modelling of random sea states — giving a description of the wave kinematics and dynamics of high-enough accuracy for many purposes. Further, several second-order nonlinear properties of surface gravity waves, and their propagation, can be estimated from its results. This linear theory is often used to get a quick and rough estimate of wave characteristics and their effects.

Description



Wave characteristics.



Dispersion of gravity waves on a fluid surface. Phase and group velocity divided by √*(gh)* as a function of *h/λ*. **A**: phase velocity, **B**: group velocity, **C**: phase and group velocity √*(gh)* valid in shallow water. *Drawn lines*: based on dispersion relation valid in arbitrary depth. *Dashed lines*: based on dispersion relation valid in deep water.

Airy wave theory uses a potential flow approach to describe the motion of gravity waves on a fluid surface. The use of — inviscid and irrotational — potential flow in water waves is remarkably successful, given its failure to describe many other fluid flows where it is often essential to take viscosity, vorticity, turbulence and/or flow separation into account. This is due to the fact that for the oscillatory part of the fluid motion, wave-induced vorticity is restricted to some thin oscillatory Stokes boundary layers at the boundaries of the fluid domain.

Airy wave theory is often used in ocean engineering and coastal engineering. Especially for random waves, sometimes called wave turbulence, the evolution of the wave statistics — including the wave spectrum — is predicted well over not too long distances (in terms of wavelengths) and in not too shallow water. Diffraction is one of the wave effects which can be described with Airy wave theory. Further, by using the WKBJ approximation, wave shoaling and refraction can be predicted.

Earlier attempts to describe surface gravity waves using potential flow were made by, among others, Laplace, Poisson, Cauchy and Kelland. But Airy was the first to publish the correct derivation and formulation in 1841. Soon after, in 1847, the linear theory of Airy was extended by Stokes for non-linear wave motion, correct up to third order in the wave steepness. Even before Airy's linear theory, Gerstner derived a nonlinear trochoidal wave theory in 1804, which however is not irrotational.

Airy wave theory is a linear theory for the propagation of waves on the surface of a potential flow and above a horizontal bottom. The free surface elevation $\eta(x,t)$ of one wave component is sinusoidal, as a function of horizontal position x and time t:

$$\eta(x,t) = a \cos(kx - \omega t)$$

where

- a is the wave amplitude in metre,
- cos is the cosine function,
- k is the angular wavenumber in radian per metre, related to the wavelength λ as

$$k = \frac{2\pi}{\lambda},$$

- ω is the angular frequency in radian per second, related to the period T and frequency f by

$$\omega = \frac{2\pi}{T} = 2\pi f.$$

The waves propagate along the water surface with the phase speed c_p:

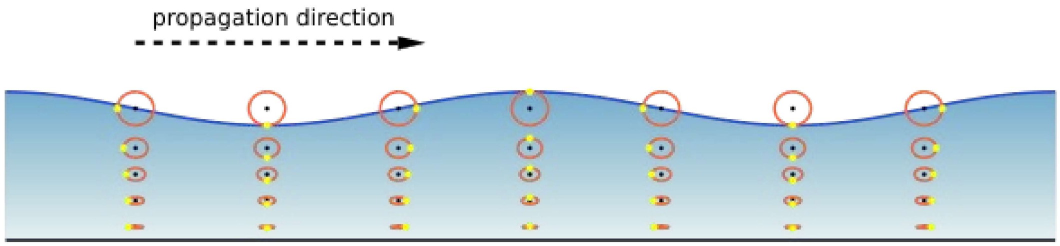
$$c_p = \frac{\omega}{k} = \frac{\lambda}{T}.$$

The angular wavenumber k and frequency ω are not independent parameters (and thus also wavelength λ and period T are not independent), but are coupled. Surface gravity waves on a fluid are dispersive waves — exhibiting frequency dispersion — meaning that each wavenumber has its own frequency and phase speed.

Note that in engineering the wave height H — the difference in elevation between crest and trough — is often used:

$$H = 2a \quad \text{and} \quad a = \frac{1}{2}H,$$

valid in the present case of linear periodic waves.



Orbital motion under linear waves. The yellow dots indicate the momentary position of fluid particles on their (orange) orbits. The black dots are the centres of the orbits.

Underneath the surface, there is a fluid motion associated with the free surface motion. While the surface elevation shows a propagating wave, the fluid particles are in an orbital motion. Within the framework of Airy wave theory, the orbits are closed curves: circles in deep water, and ellipses in finite depth—with the ellipses becoming flatter near the bottom of the fluid layer. So while the wave propagates, the fluid particles just orbit (oscillate) around their average position. With the propagating wave motion, the fluid particles transfer energy in the wave propagation direction, without having a mean velocity. The diameter of the orbits reduces with depth below the free surface. In deep water, the orbit's diameter is reduced to 4% of its free-surface value at a depth of half a wavelength.

In a similar fashion, there is also a pressure oscillation underneath the free surface, with wave-induced pressure oscillations reducing with depth below the free surface — in the same way as for the orbital motion of fluid parcels.

Mathematical formulation of the wave motion

Flow problem formulation

The waves propagate in the horizontal direction, with coordinate x, and a fluid domain bound above by a free surface at $z = \eta(x,t)$, with z the vertical coordinate (positive in the upward direction) and t being time. The level $z = 0$ corresponds with the mean surface elevation. The impermeable bed underneath the fluid layer is at $z = -h$. Further, the flow is assumed to be incompressible and irrotational — a good approximation of the flow in the fluid interior for waves on a liquid surface — and potential theory can be used to describe the flow. The velocity potential $\Phi(x,z,t)$ is related to the flow velocity components u_x and u_z in the horizontal (x) and vertical (z) directions by:

$$u_x = \frac{\partial \Phi}{\partial x} \quad \text{and} \quad u_z = \frac{\partial \Phi}{\partial z}.$$

Then, due to the continuity equation for an incompressible flow, the potential Φ has to satisfy the Laplace equation:

$$(1) \qquad \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial z^2} = 0.$$

Boundary conditions are needed at the bed and the free surface in order to close the system of equations. For their formulation within the framework of linear theory, it is necessary to specify what the base state (or zeroth-order solution) of the flow is. Here, we assume the base state is rest, implying the mean flow velocities are zero.

The bed being impermeable, leads to the kinematic bed boundary-condition:

$$(2) \qquad \frac{\partial \Phi}{\partial z} = 0 \quad \text{at } z = -h.$$

In case of deep water — by which is meant infinite water depth, from a mathematical point of view — the flow velocities have to go to zero in the limit as the vertical coordinate goes to minus infinity: $z \to -\infty$.

At the free surface, for infinitesimal waves, the vertical motion of the flow has to be equal to the vertical velocity of the free surface. This leads to the kinematic free-surface boundary-condition:

$$(3) \qquad \frac{\partial \eta}{\partial t} = \frac{\partial \Phi}{\partial z} \quad \text{at } z = \eta(x,t).$$

If the free surface elevation $\eta(x,t)$ was a known function, this would be enough to solve the flow problem. However, the surface elevation is an extra unknown, for which an additional boundary condition is needed. This is provided by Bernoulli's equation for an unsteady potential flow. The pressure above the free surface is assumed to be constant. This constant pressure is taken equal to zero, without loss of generality, since the level of such a constant pressure does not alter the flow. After linearisation, this gives the dynamic free-surface boundary condition:

$$(4) \qquad \frac{\partial \Phi}{\partial t} + g\,\eta = 0 \quad \text{at } z = \eta(x,t).$$

Because this is a linear theory, in both free-surface boundary conditions — the kinematic and the dynamic one, equations (3) and (4) — the value of Φ and $\partial\Phi/\partial z$ at the fixed mean level $z = 0$ is used.

Solution for a progressive monochromatic wave

For a propagating wave of a single frequency — a monochromatic wave — the surface elevation is of the form:

$$\eta = a \cos(kx - \omega t).$$

The associated velocity potential, satisfying the Laplace equation (1) in the fluid interior, as well as the kinematic boundary conditions at the free surface (2), and bed (3), is:

$$\Phi = \frac{\omega}{k} a \frac{\cosh\left(k\,(z+h)\right)}{\sinh\left(k\,h\right)} \sin\left(kx - \omega t\right),$$

with sinh and cosh the hyperbolic sine and hyperbolic cosine function, respectively. But η and Φ also have to satisfy the dynamic boundary condition, which results in non-trivial (non-zero) values for the wave amplitude *a* only if the linear dispersion relation is satisfied:

$$\omega^2 = g\,k\,\tanh\left(kh\right),$$

with tanh the hyperbolic tangent. So angular frequency ω and wavenumber k — or equivalently period T and wavelength λ — cannot be chosen independently, but are related. This means that wave propagation at a fluid surface is an eigenproblem. When ω and k satisfy the dispersion relation, the wave amplitude *a* can be chosen freely (but small enough for Airy wave theory to be a valid approximation).

Table of wave quantities

In the table below, several flow quantities and parameters according to Airy wave theory are given. The given quantities are for a bit more general situation as for the solution given above. Firstly, the waves may propagate in an arbitrary horizontal direction in the $\boldsymbol{x} = (x,y)$ plane. The wavenumber vector is $\boldsymbol{k}$, and is perpendicular to the cams of the wave crests. Secondly, allowance is made for a mean flow velocity $\boldsymbol{U}$, in the horizontal direction and uniform over (independent of) depth z. This introduces a Doppler shift in the dispersion relations. At an Earth-fixed location, the *observed angular frequency* (or *absolute angular frequency*) is ω. On the other hand, in a frame of reference moving with the mean velocity $\boldsymbol{U}$ (so the mean velocity as observed from this reference frame is zero), the angular frequency is different. It is called the *intrinsic angular frequency* (or *relative angular frequency*), denoted as σ. So in pure wave motion, with $\boldsymbol{U}=\boldsymbol{0}$, both frequencies ω and σ are equal. The wave number k (and wave length λ) are independent of the frame of reference, and have no Doppler shift (for monochromatic waves).

The table only gives the oscillatory parts of flow quantities — velocities, particle excursions and pressure — and not their mean value or drift. The oscillatory particle excursions $\boldsymbol{\xi}_x$ and ξ_z are the time integrals of the oscillatory flow velocities $\boldsymbol{u}_x$ and u_z respectively.

Water depth is classified into three regimes:

- **deep water** — for a water depth larger than half the wavelength, $h > \frac{1}{2}\lambda$, the phase speed of the waves is hardly influenced by depth (this is the case for most wind waves on the sea and ocean surface),

- **shallow water** — for a water depth smaller than the wavelength divided by 20, $h < \frac{1}{20}\lambda$, the phase speed of the waves is only dependent on water depth, and no longer a function of period or wavelength; and
- **intermediate depth** — all other cases, $\frac{1}{20}\lambda < h < \frac{1}{2}\lambda$, where both water depth and period (or wavelength) have a significant influence on the solution of Airy wave theory.

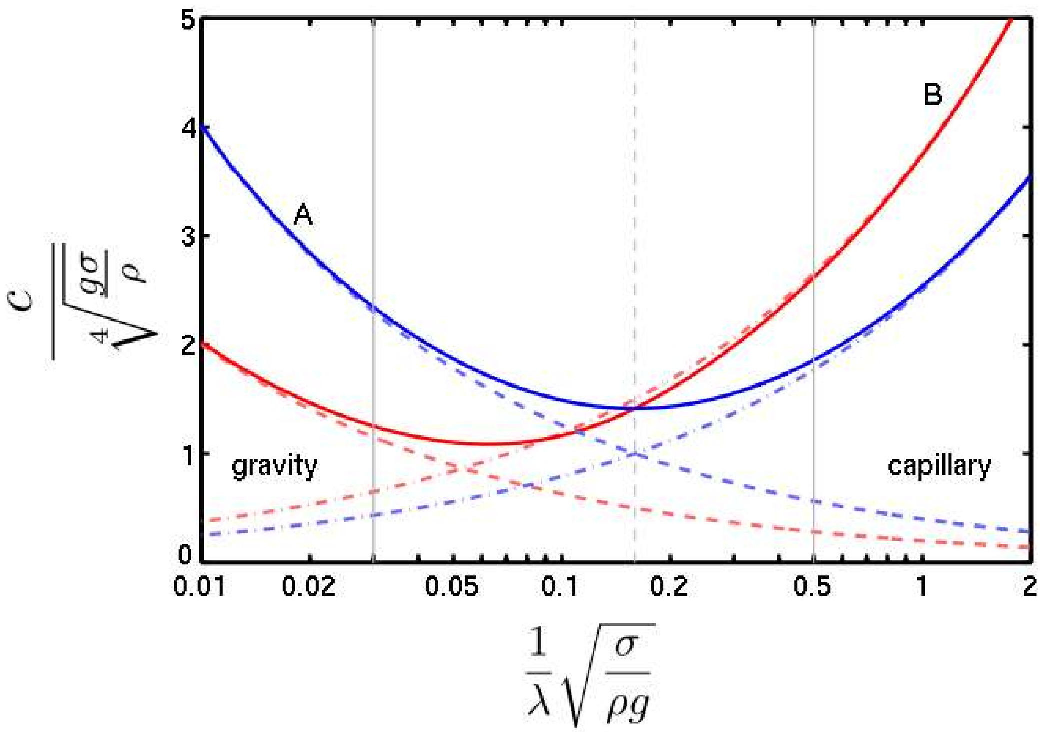
In the limiting cases of deep and shallow water, simplifying approximations to the solution can be made. While for intermediate depth, the full formulations have to be used.

Properties of gravity waves on the surface of deep water, shallow water and at intermediate depth, according to Airy wave theory

quantity	symbol	units	deep water ($h > \frac{1}{2}\lambda$)	shallow water ($h < 0.05\,\lambda$)	intermediate depth (all λ and h)
surface elevation	$\eta(\boldsymbol{x},t)$	m	$a\,\cos\,\theta(\boldsymbol{x},t)$		
wave phase	$\theta(\boldsymbol{x},t)$	rad	$\boldsymbol{k}\cdot\boldsymbol{x} - \omega t$		
observed angular frequency	ω	rad / s	$(\omega - \boldsymbol{k}\cdot\boldsymbol{U})^2 = \left(\Omega(k)\right)^2 \quad \text{with} \quad k = \lvert\boldsymbol{k}\rvert$		
intrinsic angular frequency	σ	rad / s	$\sigma^2 = \left(\Omega(k)\right)^2 \quad \text{with} \quad \sigma = \omega - \boldsymbol{k}\cdot\boldsymbol{U}$		
unit vector in the wave propagation direction	$\boldsymbol{e}_k$	–	$\dfrac{\boldsymbol{k}}{k}$		
dispersion relation	$\Omega(k)$	rad / s	$\Omega(k) = \sqrt{g\,k}$	$\Omega(k) = k\sqrt{g\,h}$	$\Omega(k) = \sqrt{g\,k\,\tanh(k\,h)}$
phase speed	$c_p = \dfrac{\Omega(k)}{k}$	m / s	$\sqrt{\dfrac{g}{k}} = \dfrac{g}{\sigma}$	$\sqrt{gh}$	$\sqrt{\dfrac{g}{k}\,\tanh\,(k\,h)}$
group speed	$c_g = \dfrac{\partial\Omega}{\partial k}$	m / s	$\dfrac{1}{2}\sqrt{\dfrac{g}{k}} = \dfrac{1}{2}\dfrac{g}{\sigma}$	$\sqrt{gh}$	$\dfrac{1}{2}c_p\left(1 + kh\dfrac{1-\tanh^2(kh)}{\tanh(kh)}\right)$

ratio	$\frac{c_g}{c_p}$	–	$\frac{1}{2}$	1	$\frac{1}{2}\left(1 + kh\frac{1 - \tanh^2(kh)}{\tanh(kh)}\right)$
horizontal velocity	$\boldsymbol{u}_x(\boldsymbol{x}, z, t)$	m / s	$\boldsymbol{e}_k \, \sigma \, a \, \mathrm{e}^{kz} \cos\theta$	$\boldsymbol{e}_k \sqrt{\frac{g}{h}} \, a \cos\theta$	$\boldsymbol{e}_k \, \sigma \, a \frac{\cosh(k(z+h))}{\sinh(kh)} \cos\theta$
vertical velocity	$u_z(\boldsymbol{x}, z, t)$	m / s	$\sigma \, a \, \mathrm{e}^{kz} \sin\theta$	$\sigma \, a \frac{z + h}{h} \sin\theta$	$\sigma \, a \frac{\sinh(k(z+h))}{\sinh(kh)} \sin\theta$
horizontal particle excursion	$\boldsymbol{\xi}_x(\boldsymbol{x}, z, t)$	m	$-\boldsymbol{e}_k \, a \, \mathrm{e}^{kz} \sin\theta$	$-\boldsymbol{e}_k \frac{1}{kh} \, a \sin\theta$	$-\boldsymbol{e}_k \, a \frac{\cosh(k(z+h))}{\sinh(kh)} \sin\theta$
vertical particle excursion	$\xi_z(\boldsymbol{x}, z, t)$	m	$a \, \mathrm{e}^{kz} \cos\theta$	$a \frac{z + h}{h} \cos\theta$	$a \frac{\sinh(k(z+h))}{\sinh(kh)} \cos\theta$
pressure oscillation	$p(\boldsymbol{x}, z, t)$	N / m²	$\rho \, g \, a \, \mathrm{e}^{kz} \cos\theta$	$\rho \, g \, a \cos\theta$	$\rho \, g \, a \frac{\cosh(k(z+h))}{\cosh(kh)} \cos\theta$

Surface tension effects



Dispersion of gravity–capillary waves on the surface of deep water. Phase and group velocity divided by $\sqrt[4]{g\sigma/\rho}$ as a function of inverse relative wavelength $\frac{1}{\lambda}\sqrt{\sigma/(\rho g)}$.
Blue lines (A): phase velocity c_p, Red lines (B): group velocity c_g.
Drawn lines: gravity–capillary waves.
Dashed lines: deep-water gravity waves.
Dash-dot lines: deep–water pure capillary waves.

Due to surface tension, the dispersion relation changes to:

$$\Omega^2(k) = \left(g + \frac{\gamma}{\rho}k^2\right) k \tanh(k\,h),$$

with γ the surface tension, with SI units in N/m. All above equations for linear waves remain the same, if the gravitational acceleration g is replaced by

$$\tilde{g} = g + \frac{\gamma}{\rho}k^2.$$

As a result of surface tension, the waves propagate faster. Surface tension only has influence for short waves, with wavelengths less than a few decimeters in case of a water–air interface. For very short wavelengths — two millimeter in case of the interface between air and water – gravity effects are negligible.

Interfacial waves

Surface gravity waves are a special case of interfacial waves, on the interface between two fluids of different density. Consider two fluids separated by an interface, and without further boundaries. Then their dispersion relation becomes:

$$\Omega^2(k) \;=\; |k| \left(\frac{\rho-\rho'}{\rho+\rho'} g \;+\; \frac{\sigma}{\rho+\rho'}\, k^2 \right),$$

where ρ and ρ' are the densities of the two fluids, below (ρ) and above (ρ') the interface, respectively. For interfacial waves to exist, the lower layer has to be heavier than the upper one, $\rho > \rho'$. Otherwise, the interface is unstable and a Rayleigh–Taylor instability develops.

Second-order wave properties

Several second-order wave properties, *i.e.* quadratic in the wave amplitude a, can be derived directly from Airy wave theory. They are of importance in many practical applications, *e.g.* forecasts of wave conditions. Using a WKBJ approximation, second-order wave properties also find their applications in describing waves in case of slowly-varying bathymetry, and mean-flow variations of currents and surface elevation. As well as in the description of the wave and mean-flow interactions due to time and space-variations in amplitude, frequency, wavelength and direction of the wave field itself.

Table of second-order wave properties

In the table below, several second-order wave properties — as well as the dynamical equations they satisfy in case of slowly-varying conditions in space and time — are given. More details on these can be found below. The table gives results for wave propagation in one horizontal spatial dimension.

Second-order quantities and their dynamics, using results of Airy wave theory

quantity	symbol	units	formula
mean wave-energy density per unit horizontal area	E	$\mathrm{J/m^2}$	$E = \frac{1}{2}\rho g a^2$

radiation stress or excess horizontal momentum flux due to the wave motion	S_{xx}	N / m	$S_{xx} = \left(2\,\frac{c_g}{c_p} - \frac{1}{2}\right) E$
wave action	$\mathcal{A}$	J·s / m²	$\mathcal{A} = \frac{E}{\sigma} = \frac{E}{\omega - k\,U}$
mean mass-flux due to the wave motion or the wave pseudo-momentum	M	kg / (m·s)	$M = \frac{E}{c_p} = k\,\frac{E}{\sigma}$
mean horizontal mass-transport velocity	$\tilde{U}$	m / s	$\tilde{U} = U + \frac{M}{\rho\,h} = U + \frac{E}{\rho\,h\,c_p}$
Stokes drift	$\bar{u}_S$	m / s	$\bar{u}_S = \frac{1}{2}\,\sigma\,k\,a^2\,\frac{\cosh 2k\,(z+h)}{\sinh^2(k\,h)}$
wave-energy propagation		J / (m²·s)	$\frac{\partial E}{\partial t} + \frac{\partial}{\partial x}\Big((U + c_g)\,E\Big) + S_{xx}\,\frac{\partial U}{\partial x} = 0$
wave action conservation		J / m²	$\frac{\partial \mathcal{A}}{\partial t} + \frac{\partial}{\partial x}\Big((U + c_g)\,\mathcal{A}\Big) = 0$
wave-crest conservation		rad / (m·s)	$\frac{\partial k}{\partial t} + \frac{\partial \omega}{\partial x} = 0$ with $\omega = \Omega(k) + k\,U$
mean mass conservation		kg / (m²·s)	$\frac{\partial}{\partial t}\Big(\rho\,h\Big) + \frac{\partial}{\partial x}\Big(\rho\,h\,\tilde{U}\Big) = 0$

mean horizontal-momentum evolution	N / m^2	$\frac{\partial}{\partial t}\left(\rho h \tilde{U}\right) + \frac{\partial}{\partial x}\left(\rho h \tilde{U}^2 + \frac{1}{2}\rho g h^2 + S_{xx}\right) = \rho g h \frac{\partial d}{\partial x}$

The last four equations describe the evolution of slowly-varying wave trains over bathymetry in interaction with the mean flow, and can be derived from a variational principle: Whitham's average Lagrangian method. In the mean horizontal-momentum equation, $d(x)$ is the still water depth, *i.e.* the bed underneath the fluid layer is located at $z = -d$. Note that the mean-flow velocity in the mass and momentum equations is the *mass transport velocity* $\tilde{U}$, including the splash-zone effects of the waves on horizontal mass transport, and not the mean Eulerian velocity (e.g. as measured with a fixed flow meter).

Wave energy density

Wave energy is a quantity of primary interest, since it is a primary quantity that is transported with the wave trains. As can be seen above, many wave quantities like surface elevation and orbital velocity are oscillatory in nature with zero mean (within the framework of linear theory). In water waves, the most used energy measure is the mean wave energy density per unit horizontal area. It is the sum of the kinetic and potential energy density, integrated over the depth of the fluid layer and averaged over the wave phase. Simplest to derive is the mean potential energy density per unit horizontal area E_{pot} of the surface gravity waves, which is the deviation of the potential energy due to the presence of the waves:

$$E_{\text{pot}} = \overline{\int_{-h}^{\eta} \rho\, g\, z\, \mathrm{d}z} - \int_{-h}^{0} \rho\, g\, z\, \mathrm{d}z = \overline{\frac{1}{2}\rho\, g\, \eta^2} = \frac{1}{4}\rho\, g\, a^2,$$

with an overbar denoting the mean value (which in the present case of periodic waves can be taken either as a time average or an average over one wavelength in space).

The mean kinetic energy density per unit horizontal area E_{kin} of the wave motion is similarly found to be:

$$E_{\text{kin}} = \overline{\int_{-h}^{0} \frac{1}{2}\rho\left[\left|\boldsymbol{U} + \boldsymbol{u}_x\right|^2 + u_z^2\right] \mathrm{d}z} - \int_{-h}^{0} \frac{1}{2}\rho\left|\boldsymbol{U}\right|^2 \mathrm{d}z = \frac{1}{4}\rho \frac{\sigma^2}{k \tanh(k\,h)} a^2,$$

with σ the intrinsic frequency, see the table of wave quantities. Using the dispersion relation, the result for surface gravity waves is:

$$E_{\text{kin}} = \frac{1}{4}\rho\, g\, a^2.$$

As can be seen, the mean kinetic and potential energy densities are equal. This is a general property of energy densities of progressive linear waves in a conservative system. Adding potential and kinetic contributions, E_{pot} and E_{kin}, the mean energy density per unit horizontal area E of the wave motion is:

$$E = E_{\text{pot}} + E_{\text{kin}} = \frac{1}{2}\rho g a^2.$$

In case of surface tension effects not being negligible, their contribution also adds to the potential and kinetic energy densities, giving

$$E_{\text{pot}} = E_{\text{kin}} = \frac{1}{4}\left(\rho g + \gamma k^2\right) a^2, \qquad \text{so} \qquad E = E_{\text{pot}} + E_{\text{kin}} = \frac{1}{2}\left(\rho g + \gamma k^2\right) a^2,$$

with γ the surface tension.

Wave action, wave energy flux and radiation stress

In general, there can be an energy transfer between the wave motion and the mean fluid motion. This means, that the wave energy density is not in all cases a conserved quantity (neglecting dissipative effects), but the total energy density — the sum of the energy density per unit area of the wave motion and the mean flow motion — is. However, there is for slowly-varying wave trains, propagating in slowly-varying bathymetry and mean-flow fields, a similar and conserved wave quantity, the wave action $\mathcal{A} = E/\sigma$:

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot \left[\left(\boldsymbol{U} + \boldsymbol{c}_g\right) \mathcal{A}\right] = 0,$$

with $\left(\boldsymbol{U} + \boldsymbol{c}_g\right) \mathcal{A}$ the action flux and $\boldsymbol{c}_g = c_g\, \boldsymbol{e}_k$ the group velocity vector. Action conservation forms the basis for many wind wave models and wave turbulence models. It is also the basis of coastal engineering models for the computation of wave shoaling. Expanding the above wave action conservation equation leads to the following evolution equation for the wave energy density:

$$\frac{\partial E}{\partial t} + \nabla \cdot \left[\left(\boldsymbol{U} + \boldsymbol{c}_g\right) E\right] + \mathbb{S} : \left(\nabla \boldsymbol{U}\right) = 0,$$

with:

- $\left(\boldsymbol{U} + \boldsymbol{c}_g\right) E$ is the mean wave energy density flux,
- $\mathbb{S}$ is the radiation stress tensor and
- $\nabla \boldsymbol{U}$ is the mean-velocity shear-rate tensor.

In this equation in non-conservation form, the Frobenius inner product $\mathbb{S} : (\nabla \boldsymbol{U})$ is the source term describing the energy exchange of the wave motion with the mean flow. Only in case the mean shear-rate is zero, $\nabla \boldsymbol{U} = 0$, the mean wave energy density E is conserved. The two tensors $\mathbb{S}$ and $\nabla \boldsymbol{U}$ are in a Cartesian coordinate system of the form:

$$\mathbb{S} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix} = \mathbb{I} \left(\frac{c_g}{c_p} - \frac{1}{2} \right) E + \frac{1}{k^2} \begin{pmatrix} k_x\, k_x & k_x\, k_y \\ k_y\, k_x & k_y\, k_y \end{pmatrix} \frac{c_g}{c_p} E,$$

$$\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and}$$

$$\nabla \boldsymbol{U} = \begin{pmatrix} \dfrac{\partial U_x}{\partial x} & \dfrac{\partial U_y}{\partial x} \\ \dfrac{\partial U_x}{\partial y} & \dfrac{\partial U_y}{\partial y} \end{pmatrix},$$

with k_x and k_y the components of the wavenumber vector $\boldsymbol{k}$ and similarly U_x and U_y the components in of the mean velocity vector $\boldsymbol{U}$.

Wave mass flux and wave momentum

The mean horizontal momentum per unit area $\boldsymbol{M}$ induced by the wave motion — and also the wave-induced mass flux or mass transport — is:

$$\boldsymbol{M} = \overline{\int_{-h}^{\eta} \rho \left(\boldsymbol{U} + \boldsymbol{u}_x \right) \, \mathrm{d}z} - \int_{-h}^{0} \rho \boldsymbol{U} \, \mathrm{d}z = \frac{E}{c_p} \boldsymbol{e}_k,$$

which is an exact result for periodic progressive water waves, also valid for nonlinear waves. However, its validity strongly depends on the way how wave momentum and mass flux are defined. Stokes already identified two possible definitions of phase velocity for periodic nonlinear waves:

- *Stokes first definition of wave celerity* (S1) — with the mean Eulerian flow velocity equal to zero for all elevations z below the wave troughs, and
- *Stokes second definition of wave celerity* (S2) — with the mean mass transport equal to zero.

The above relation between wave momentum $\boldsymbol{M}$ and wave energy density E is valid within the framework of Stokes' first definition.

However, for waves perpendicular to a coast line or in closed laboratory wave channel, the second definition (S2) is more appropriate. These wave systems have zero mass flux and momentum when using the second definition. In contrast, according to Stokes' first definition (S1), there is a wave-induced mass flux in the wave propagation direction, which has to be balanced by a mean flow $\boldsymbol{U}$ in the opposite direction — called the undertow.

So in general, there are quite some subtleties involved. Therefore also the term pseudo-momentum of the waves is used instead of wave momentum.

Mass and momentum evolution equations

For slowly-varying bathymetry, wave and mean-flow fields, the evolution of the mean flow can de described in terms of the mean mass-transport velocity $\tilde{\boldsymbol{U}}$ defined as:

$$\tilde{\boldsymbol{U}} = \boldsymbol{U} + \frac{\boldsymbol{M}}{\rho h}.$$

Note that for deep water, when the mean depth h goes to infinity, the mean Eulerian velocity $\boldsymbol{U}$ and mean transport velocity $\tilde{\boldsymbol{U}}$ become equal.

The equation for mass conservation is:

$$\frac{\partial}{\partial t}(\rho h) + \nabla \cdot \left(\rho h \tilde{\boldsymbol{U}}\right) = 0,$$

where $h(\boldsymbol{x},t)$ is the mean water-depth, slowly varying in space and time. Similarly, the mean horizontal momentum evolves as:

$$\frac{\partial}{\partial t}\left(\rho h \tilde{\boldsymbol{U}}\right) + \nabla \cdot \left(\rho h \tilde{\boldsymbol{U}} \otimes \tilde{\boldsymbol{U}} + \frac{1}{2}\rho g h^2 \mathbb{I} + \mathbb{S}\right) = \rho g h \nabla d,$$

with d the still-water depth (the sea bed is at $z=-d$), $\mathbb{S}$ is the wave radiation-stress tensor, $\mathbb{I}$ is the identity matrix and $\otimes$ is the dyadic product:

$$\tilde{\boldsymbol{U}} \otimes \tilde{\boldsymbol{U}} = \begin{pmatrix} \tilde{U}_x \tilde{U}_x & \tilde{U}_x \tilde{U}_y \\ \tilde{U}_y \tilde{U}_x & \tilde{U}_y \tilde{U}_y \end{pmatrix}.$$

Note that mean horizontal momentum is only conserved if the sea bed is horizontal (*i.e.* the still-water depth d is a constant), in agreement with Noether's theorem.

The system of equations is closed through the description of the waves. Wave energy propagation is described through the wave-action conservation equation (without dissipation and nonlinear wave interactions):

$$\frac{\partial}{\partial t}\left(\frac{E}{\sigma}\right) + \nabla \cdot \left[\left(\boldsymbol{U} + \boldsymbol{c}_g\right) \frac{E}{\sigma}\right] = 0.$$

The wave kinematics are described through the wave-crest conservation equation:

$$\frac{\partial \boldsymbol{k}}{\partial t} + \nabla \omega = \boldsymbol{0},$$

with the angular frequency ω a function of the (angular) wavenumber $\boldsymbol{k}$, related through the dispersion relation. For this to be possible, the wave field must be coherent. By taking the curl of the wave-crest conservation, it can be seen that an initially irrotational wavenumber field stays irrotational.

Stokes drift

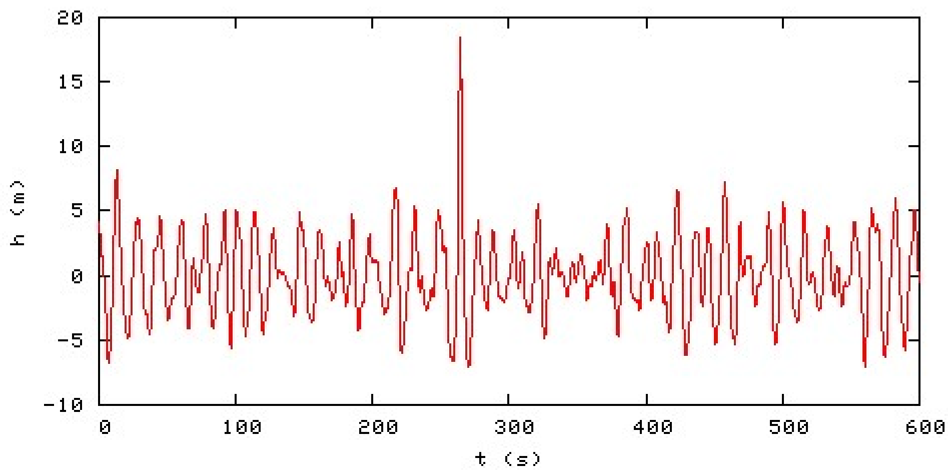
When following a single particle in pure wave motion $(\boldsymbol{U} = \boldsymbol{0})$, according to linear Airy wave theory the particles are in closed elliptical orbit. However, in nonlinear waves this is no longer the case and the particles exhibit a Stokes drift. The Stokes drift velocity $\bar{\boldsymbol{u}}_S$, which is the Stokes drift after one wave cycle divided by the period, can be estimated using the results of linear theory:

$$\bar{\boldsymbol{u}}_S = \frac{1}{2} \sigma\, k\, a^2 \frac{\cosh\, 2\, k\, (z + h)}{\sinh^2 (k\, h)}\, \boldsymbol{e}_k,$$

so it varies as a function of elevation. The given formula is for Stokes first definition of wave celerity. When $\rho\, \bar{\boldsymbol{u}}_S$ is integrated over depth, the expression for the mean wave momentum $\boldsymbol{M}$ is recovered.

Chapter 9

Rogue Wave



The Draupner wave, a single giant wave measured on New Year's Day 1995, finally confirmed the existence of freak waves, which had previously been considered near-mythical

Rogue waves (also known as **freak waves**, **monster waves**, **killer waves**, **extreme waves**, and **abnormal waves**) are relatively large and spontaneous ocean surface waves that occur far out in sea, and are a threat even to large ships and ocean liners. In oceanography, they are more precisely defined as waves whose height is more than twice the significant wave height (SWH), which is itself defined as the mean of the largest third of waves in a wave record. Therefore rogue waves are not necessarily the biggest waves found at sea; they are, rather, surprisingly large waves for a given sea state. "Rogue waves are not tsunamis, which are set in motion by earthquakes [and] travel at high speed, building up as they approach the shore. Rogue waves seem to occur in deep water or where a number of physical factors such as strong winds and fast currents converge. This may have a focusing effect, which can cause a number of waves to join together."

Background

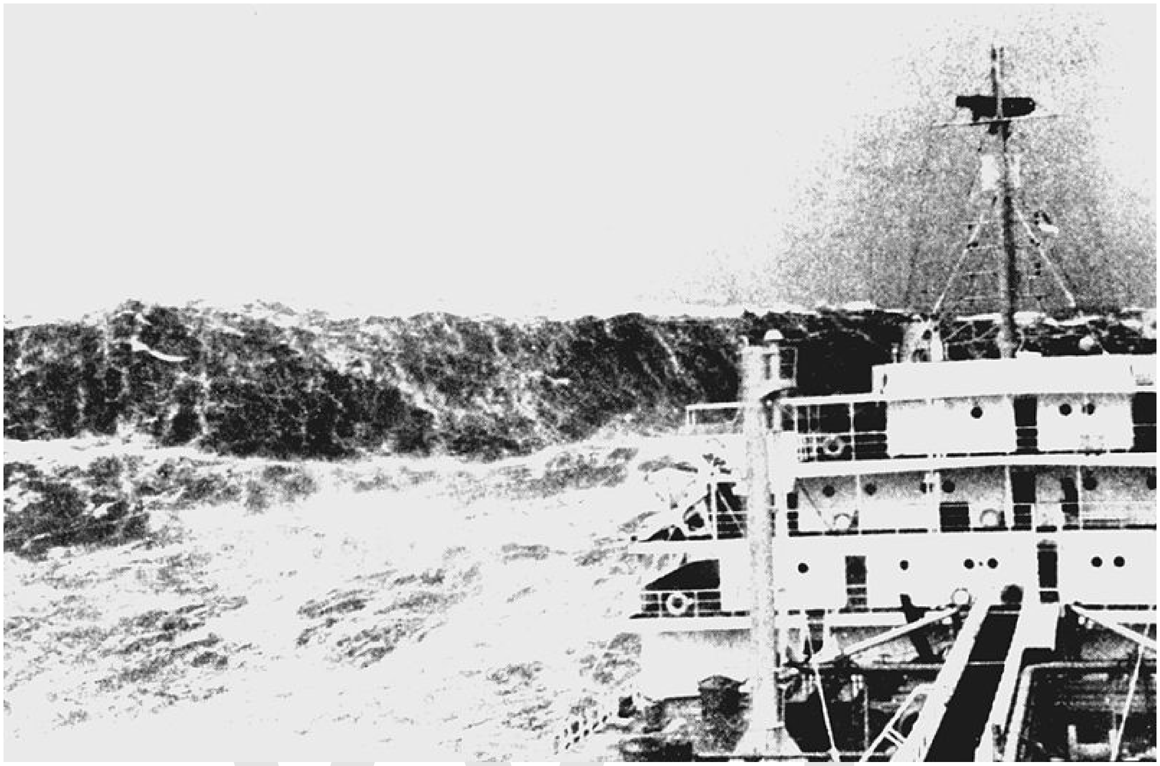
Once thought by scientists to exist only in legends, rogue waves are now known to be a natural ocean phenomenon. Eyewitness accounts from mariners and damages inflicted on ships have long suggested they occurred; however, their scientific measurement was only positively confirmed following measurements of the "Draupner wave", a rogue wave at the Draupner platform, in the North Sea on January 1, 1995. During this event, minor damage was inflicted on the platform, confirming that the reading was valid. Satellite images have also confirmed their existence.

Freak waves have been cited in the media as a likely source of the sudden, inexplicable disappearance of many ocean-going vessels. One of the very few cases in which evidence exists that *may* indicate a freak wave incident is the 1978 loss of the freighter MS *München*, detailed below. In February 2000, a British oceanographic research vessel sailing in the Rockall Trough west of Scotland encountered the largest waves ever recorded by scientific instruments in the open ocean, with a SWH of 18.5 meters (61 ft) and individual waves up to 29.1 meters (95 ft). "In 2004 scientists using three weeks of radar images from European Space Agency satellites found ten rogue waves, each 25 meters or higher."

A rogue wave is not the same as a tsunami. Tsunamis are caused by mass displacement, such as sudden movement of the ocean floor. They propagate at high speed over a wide area and are more or less unnoticeable in deep water, only becoming dangerous as they approach the shoreline and the ocean floor becomes shallower. They do not present a threat to shipping at sea (the only ships lost in the 2004 Asian tsunami were in port). A rogue wave, on the other hand, is a highly localized phenomenon both in space and duration, most frequently occurring far out at sea.

Rogue waves may sometimes be referred to as "hundred-year waves," due to the supposed likelihood of their occurrence. They should not be confused, however, with the hundred-year wave, which is a statistical prediction of the highest wave likely to occur in a hundred-year period in a particular body of water. These predictions are typically based on wave models which do not take rogue waves into account.

History



Merchant ship labouring in heavy seas as a huge wave looms astern. Huge waves are common near the 100-fathom line in the Bay of Biscay. Published in Fall 1993 issue of Mariner's Weather Log. Credits: NOAA Photo Library

It is common for mid-ocean storm waves to reach 7 meters (23 ft) in height, and in extreme conditions such waves can reach heights of 15 meters (49 ft). However, for centuries maritime folklore told of the existence of vastly more massive waves — veritable monsters up to 30 meters (98 ft) in height (approximately the height of a 10-story building) — that could appear without warning in mid-ocean, against the prevailing current and wave direction, and often in perfectly clear weather. Such waves were said to consist of an almost vertical wall of water preceded by a trough so deep that it was referred to as a "hole in the sea"; a ship encountering a wave of such magnitude would be unlikely to survive the tremendous pressures of up to 980 kPa (142 psi) exerted by the weight of the breaking water, and would almost certainly be sunk in a matter of minutes.

Many years of research have confirmed that waves of up to 35 meters (115 ft) in height are much more common than mathematical probability theory would predict using a Rayleigh distribution of wave heights. In addition, pressure readings from buoys moored in the Gulf of Mexico at the time of Hurricane Katrina also indicated the presence of such large waves at the time of the storm. In fact, they seem to occur in all of the world's oceans many times every year. This has caused a re-examination of the reasons for their existence, as well as reconsideration of the implications for ocean-going ship design.

Rogue waves also occur on the Great Lakes. A phenomenon known as the "Three Sisters" is said to occur in Lake Superior when a series of three waves form that are one-third larger than the regular waves. The second wave hits the ship's deck before the first wave clears. The third incoming wave adds to the two accumulated backwashes and suddenly overloads the ship deck with tons of water. The House Subcommittee on Coast Guard and Navigation took testimony from commercial fisherman Lyle A. McDonald of Laurium, Michigan that the "Three Sister" phenomenon may have contributed to the infamous sinking of the SS *Edmund Fitzgerald* on Lake Superior in November 1975.

Occurrence

In the course of Project MaxWave, researchers from the GKSS Research Centre, using data collected by ESA satellites, identified a large number of radar signatures that have been portrayed as evidence for rogue waves. Further research is under way to develop better methods of translating the radar echoes into sea surface elevation, but at present this technique is not proven.

Causes

Because the phenomenon of rogue waves is still a matter of active research, it is premature to state clearly what the most common causes are or whether they vary from place to place. The areas of highest predictable risk appear to be where a strong current runs counter to the primary direction of travel of the waves; the area near Cape Agulhas off the southern tip of Africa is one such area; the warm Agulhas current runs to the southwest, while the dominant winds are westerlies. However, since this thesis does not explain the existence of all waves that have been detected, several different mechanisms are likely, with localised variation. Suggested mechanisms for freak waves include the following:

- Diffractive focusing — According to this hypothesis, coast shape or seabed shape directs several small waves to meet in phase. Their crest heights combine to create a freak wave.
- Focusing by currents — Waves from one current are driven into an opposing current. This results in shortening of wavelength, causing shoaling (i.e., increase in wave height), and oncoming wave trains to compress together into a rogue wave. This happens off the South African coast, where the Agulhas current is countered by westerlies.
- Nonlinear effects — It seems possible to have a rogue wave occur by natural, nonlinear processes from a random background of smaller waves. In such a case, it is hypothesised, an unusual, unstable wave type may form which 'sucks' energy from other waves, growing to a near-vertical monster itself, before becoming too unstable and collapsing shortly after. One simple model for this is a wave equation known as the nonlinear Schrödinger equation (NLS), in which a normal and perfectly accountable (by the standard linear model) wave begins to 'soak' energy from the waves immediately fore and aft, reducing them to minor ripples

compared to other waves. The NLS can be used in deep water conditions. In shallow water, waves are described by the Korteweg–de Vries equation or the Boussinesq equation. These equations also have non-linear contributions and show solitary-wave solutions.

- Normal part of the wave spectrum — Rogue waves are not freaks at all but are part of normal wave generation process, albeit a rare extremity.
- Wind waves — While it is unlikely that wind alone can generate a rogue wave, its effect combined with other mechanisms may provide a fuller explanation of freak wave phenomena. As wind blows over the ocean, energy is transferred to the sea surface. When strong winds from a storm happen to blow in the opposing direction of the ocean current the forces might be strong enough to randomly generate rogue waves. Theories of instability mechanisms for the generation and growth of wind waves—although not on the causes of rogue waves—are provided by Phillips and Miles.

The spatio-temporal focusing seen in the NLS equation can also occur when the nonlinearity is removed. In this case, focusing is primarily due to different waves coming into phase, rather than any energy transfer processes. Further analysis of rogue waves using a fully nonlinear model by R.H. Gibbs (2005) brings this mode into question, as it is shown that a typical wavegroup focuses in such a way as to produce a significant wall of water, at the cost of a reduced height.

A rogue wave, and the deep trough commonly seen before and after it, may last only for some minutes before either breaking, or reducing in size again. Apart from one single rogue wave, the rogue wave may be part of a wave packet consisting of a few rogue waves. Such rogue wave groups have been observed in nature.

There are three categories of freak waves:

- "Walls of water" travelling up to 10 km (6.2 mi) through the ocean
- "Three Sisters", groups of three waves
- Single, giant storm waves, building up to fourfold the storm's waves height and collapsing after some seconds

A research group at the Umeå University, Sweden in August 2006 showed that normal stochastic wind driven waves can suddenly give rise to monster waves. The nonlinear evolution of the instabilities was investigated by means of direct simulations of the time-dependent system of nonlinear equations.

Applications

The possibility of the artificial stimulation of rogue wave phenomena has attracted research funding from DARPA, an agency of the United States Department of Defense. Bahram Jalali and other researchers at UCLA studied microstructured optical fibers near the threshold of soliton supercontinuum generation and observed rogue wave phenomena.

After modeling the effect, the researchers announced that they had successfully characterized the proper initial conditions for generating rogue waves in any medium.

Reported encounters

It should be noted that many of these encounters are only reported in the media, and are not examples of open ocean rogue waves. Often, in popular culture, an endangering huge wave is loosely denoted as a *rogue wave*, while it has not been (and most often cannot be) established that the reported event is a rogue wave in the scientific sense — *i.e.* of a very different nature in characteristics as the surrounding waves in that sea state and with very low probability of occurrence (according to a Gaussian process description as valid for linear wave theory).

Nineteenth century

- The Eagle Island lighthouse (1861) – water broke the glass of the structure's east tower and flooded it, implying a wave that surmounted the 40 m (130 ft) cliff and overwhelmed the 26 m (85 ft) tower.
- Flannan Isles (1900) – three lighthouse keepers vanished after a storm that resulted in wave-damaged equipment being found 34 meters (112 ft) above sea level.

Twentieth century

- SS *Waratah* - In 1909, it left Durban, South Africa with 211 passengers and crew but never made it to Cape Town, South Africa.
- Voyage of the James Caird - In 1916 Sir Ernest Shackleton encountered a wave he termed "gigantic" while piloting a lifeboat/whaler from Elephant Island to South Georgia Island.
- USS *Ramapo* (1933) – triangulated at 112 feet (34 m).
- RMS *Queen Mary* (1942) – broadsided by a 92-foot (28 m) wave and listed briefly about 52 degrees before slowly righting.
- *SS Michelangelo* (1966) – hole torn in superstructure, heavy glass smashed 80 feet (24 m) above the waterline, and 3 deaths.
- SS *Edmund Fitzgerald* (1975) – lost on Lake Superior. A Coast Guard report blamed water entry to the hatches, which gradually filled the hold, or alternatively errors in navigation or charting causing damage from running onto shoals. However, another nearby ship, the SS *Arthur M. Anderson*, was hit at a similar time by two rogue waves and possibly a third, and this appeared to coincide with the sinking around ten minutes later.
- MS *München* (1978) – lost at sea leaving only "a few bits of wreckage" and signs of sudden damage including extreme forces 66 feet (20 m) above the water line. Although more than one wave was probably involved, this remains the most likely sinking due to a freak wave.

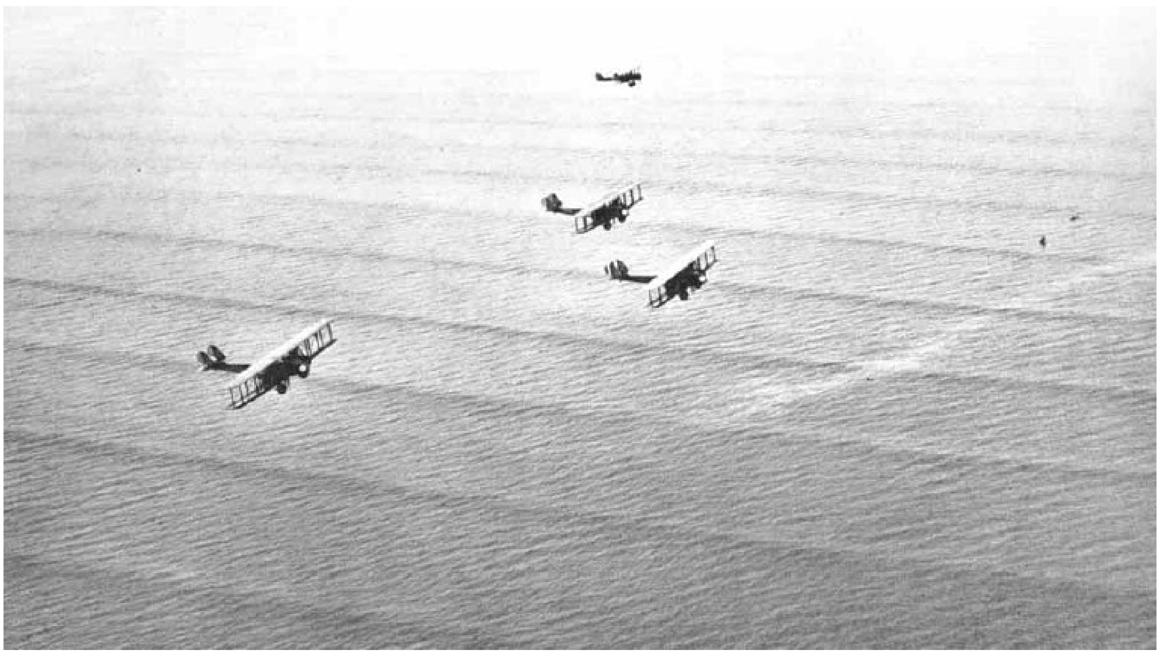
- Esso Languedoc A 25-30m wave washed across the deck from the stern of the French supertanker, and was photographed by the first mate, Philippe Lijour, in 1980.
- Fastnet Lighthouse Struck by 48m wave in 1985
- Draupner wave (North Sea, 1995) – First rogue wave confirmed with scientific evidence, it had a maximum height of 25.6 meters (84 ft).
- RMS *Queen Elizabeth 2* – North Atlantic, September 1995, 29 meters (95 ft), during Hurricane Luis:
  The Master said it "came out of the darkness" and "looked like the White Cliffs of Dover." Newspaper reports at the time described the cruise liner as attempting to "surf" the near-vertical wave in order not to be sunk.

Twenty-first century

- MS *Bremen* and *Caledonian Star* (South Atlantic, 2001) encountered 30-meter (98 ft) freak waves. Bridge windows on both ships were smashed, and all power and instrumentation lost.
- U.S. Naval Research Laboratory ocean-floor pressure sensors detected a freak wave caused by Hurricane Ivan in the Gulf of Mexico, 2004. The wave was around 27.7 meters (91 ft) high from peak to trough, and around 200 meters (660 ft) long.
- *Aleutian Ballad*, (Bering Sea, 2005)
  Footage of what is identified as a rogue wave appears in an episode of *Deadliest Catch*. The wave cripples the vessel, causing the boat to tip for a short period onto its side. One of the few video recordings of what might be a rogue wave.
- It has been suggested that these types of waves may be responsible for the loss of several low-flying aircraft, namely U.S. Coast Guard helicopters on Search and Rescue missions.
- *MS Louis Majesty*, (Gulf of Marseille, Mediterranean Sea, March 2010). An unexpected pack of three waves of 26 feet (7.9 m) hit the ship while on a cruise between Carthagena and Marseille. Two passengers were killed in a lounge by flying glass from a shattered window. Damage to the ship was done by the second and third waves. While this wave was much lower than freak waves appearing in open oceans, published evidence indicates that its behaviour was similar to that of freak waves.

Chapter 10

Cnoidal Wave



US Army bombers flying over near-periodic swell in shallow water, close to the Panama coast (1933). The sharp crests and very flat troughs are characteristic for cnoidal waves.

In fluid dynamics, a **cnoidal wave** is a nonlinear and exact periodic wave solution of the Korteweg–de Vries equation. These solutions are in terms of the Jacobi elliptic function *cn*, which is why they are coined *cn*oidal waves. They are used to describe surface gravity waves of fairly long wavelength, as compared to the water depth.

The cnoidal wave solutions were derived by Korteweg and de Vries, in their 1895 paper in which they also propose their dispersive long-wave equation, now known as the Korteweg–de Vries equation. In the limit of infinite wavelength, the cnoidal wave becomes a solitary wave.

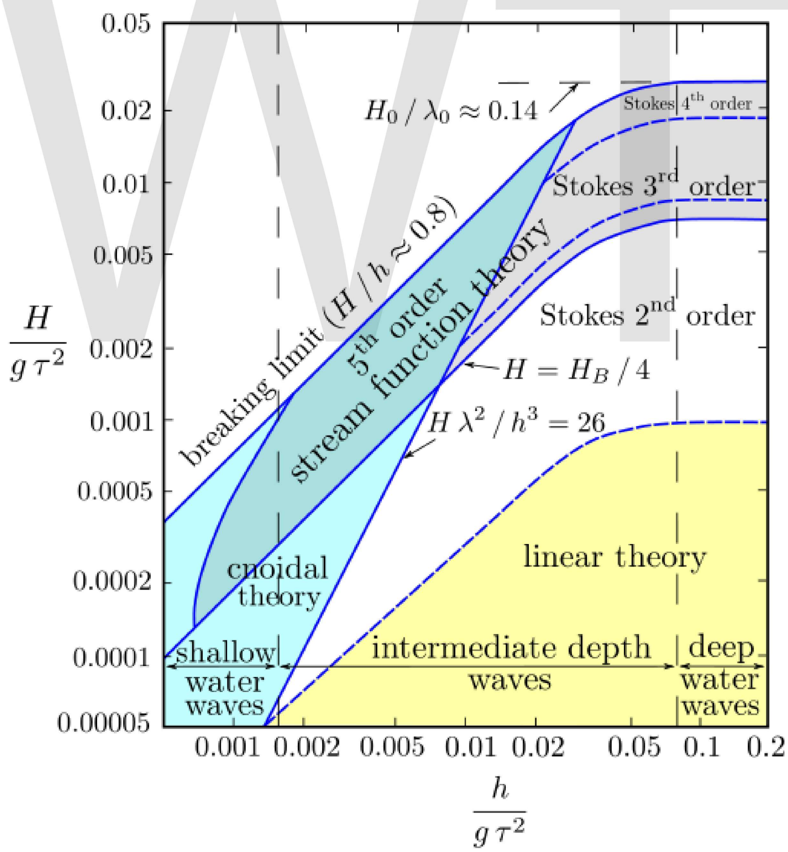
The Benjamin–Bona–Mahony equation has improved short-wavelength behaviour, as compared to the Korteweg–de Vries equation, and is another uni-directional wave equation with cnoidal wave solutions. Further, since the Korteweg–de Vries equation is an approximation to the Boussinesq equations for the case of one-way wave propagation, cnoidal waves are approximate solutions to the Boussinesq equations.

Cnoidal wave solutions can appear in other applications than surface gravity waves as well, for instance to describe ion acoustic waves in plasma physics.

A cnoidal wave, characterised by sharper crests and flatter troughs than in a sine wave. For the shown case, the elliptic parameter is $m = 0.9$.

Background

Korteweg–de Vries, and Benjamin–Bona–Mahony equations



Validity of several theories for periodic water waves, according to Le Méhauté (1976). The light-blue area gives the range of validity of cnoidal wave theory; light-yellow for Airy wave theory; and the dashed blue lines demarcate between the required order in

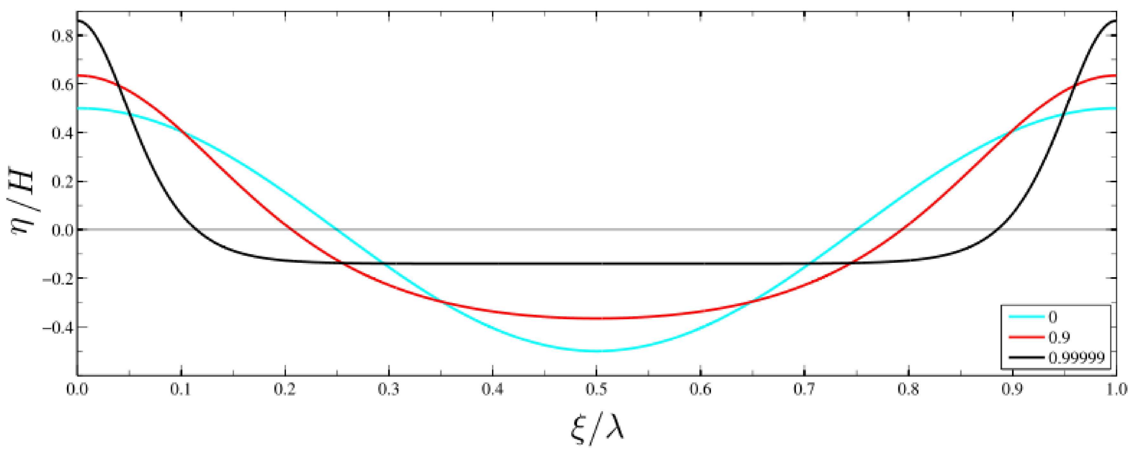
Stokes' wave theory. The light-gray shading gives the range extension by numerical approximations using fifth-order stream-function theory, for high waves ($H > ¼ H_{breaking}$).

The Korteweg–de Vries equation (KdV equation) can be used to describe the uni-directional propagation of weakly nonlinear and long waves—where long wave means: having long wavelengths as compared with the mean water depth—of surface gravity waves on a fluid layer. The KdV equation is a dispersive wave equation, including both frequency dispersion and amplitude dispersion effects. In its classical use, the KdV equation is applicable for wavelengths λ in excess of about five times the average water depth h, so for $\lambda > 5\ h$; and for the period τ greater than $7\sqrt{h/g}$ with g the strength of the gravitational acceleration. To envisage the position of the KdV equation within the scope of classical wave approximations, it distinguishes itself in the following ways:

- **Korteweg–de Vries equation** — describes the forward propagation of weakly nonlinear and dispersive waves, for long waves with $\lambda > 7\ h$.
- **Shallow water equations** — are also nonlinear and do have amplitude dispersion, but no frequency dispersion; they are valid for very long waves, $\lambda > 20\ h$.
- **Boussinesq equations** — have the same range of validity as the KdV equation (in their classical form), but allow for wave propagation in arbitrary directions, so not only forward-propagating waves. The drawback is that the Boussinesq equations are often more difficult to solve than the KdV equation; and in many applications wave reflections are small and may be neglected.
- **Airy wave theory** — has full frequency dispersion, so valid for arbitrary depth and wavelength, but is a linear theory without amplitude dispersion, limited to low-amplitude waves.
- **Stokes' wave theory** — a perturbation-series approach to the description of weakly nonlinear and dispersive waves, especially successful in deeper water for relative short wavelengths, as compared to the water depth. However, for long waves the Boussinesq approach—as also applied in the KdV equation—is often preferred. This is because in shallow water the Stokes' perturbation series needs many terms before convergence towards the solution, due to the peaked crests and long flat troughs of the nonlinear waves. While the KdV or Boussinesq models give good approximations for these long nonlinear waves.

The KdV equation can be derived from the Boussinesq equations, but additional assumptions are needed to be able to split-off of the forward wave propagation. For practical applications, the Benjamin–Bona–Mahony equation (BBM equation) is preferable over the KdV equation, a forward-propagating model similar to KdV but with much better frequency-dispersion behaviour at shorter wavelengths. Further improvements in short-wave performance can be obtained by starting to derive a one-way wave equation from a modern improved Boussinesq model, valid for even shorter wavelengths.

Cnoidal waves



Cnoidal wave profiles for three values of the elliptic parameter *m*.

blue : *m* = 0,
red : *m* = 0.9 and
black : *m* = 0.99999.

The cnoidal wave solutions of the KdV equation were presented by Korteweg and de Vries in their 1895 paper, which article is based on the PhD thesis by de Vries in 1894. Solitary wave solutions for nonlinear and dispersive long waves had been found earlier by Boussinesq in 1872, and Rayleigh in 1876. The search for these solutions was triggered by the observations of this solitary wave (or "wave of translation") by Russell, both in nature and laboratory experiments. Cnoidal wave solutions of the KdV equation are stable with respect to small perturbations.

The surface elevation $\eta(x,t)$, as a function of horizontal position x and time t, for a cnoidal wave is given by:

$$\eta(x,t) = \eta_2 + H\,\mathrm{cn}^2\left(2\,K(m)\,\frac{x-ct}{\lambda} \,\middle|\, m \right),$$

where H is the wave height, λ is the wavelength, c is the phase speed and η_2 is the trough elevation. Further cn is one of the Jacobi elliptic functions and $K(m)$ is the complete elliptic integral of the first kind; both are dependent on the elliptic parameter m. The latter, m, determines the shape of the cnoidal wave. For m equal to zero the cnoidal wave becomes a cosine function, while for values close to one the cnoidal wave gets peaked crests and (very) flat troughs.

An important dimensionless parameter for nonlinear long waves ($\lambda \gg h$) is the Ursell parameter:

$$U = \frac{H\,\lambda^2}{h^3} \qquad = \frac{H}{h}\,\left(\frac{\lambda}{h}\right)^2.$$

For small values of U, say $U < 5$, a linear theory can be used, and at higher values nonlinear theories have to be used, like cnoidal wave theory. The demarcation zone between—third or fifth order—Stokes' and cnoidal wave theories is in the range 10–25 of the Ursell parameter. As can be seen from the formula for the Ursell parameter, for a given relative wave height H/h the Ursell parameter—and thus also the nonlinearity—grows quickly with increasing relative wavelength λ/h.

Based on the analysis of the full nonlinear problem of surface gravity waves within potential flow theory, the above cnoidal waves can be considered the lowest-order term in a perturbation series. Higher-order cnoidal wave theories remain valid for shorter and more nonlinear waves. A fifth-order cnoidal wave theory was developed by Fenton in 1979. A detailed description and comparison of fifth-order Stokes' and fifth-order cnoidal wave theories is given in the review article by Fenton.

Cnoidal wave descriptions, through a renormalisation, are also well suited to waves on deep water, even infinite water depth; as found by Clamond. A description of the interactions of cnoidal waves in shallow water, as found in real seas, has been provided by Osborne in 1994.

Periodic wave solutions

Korteweg–de Vries equation

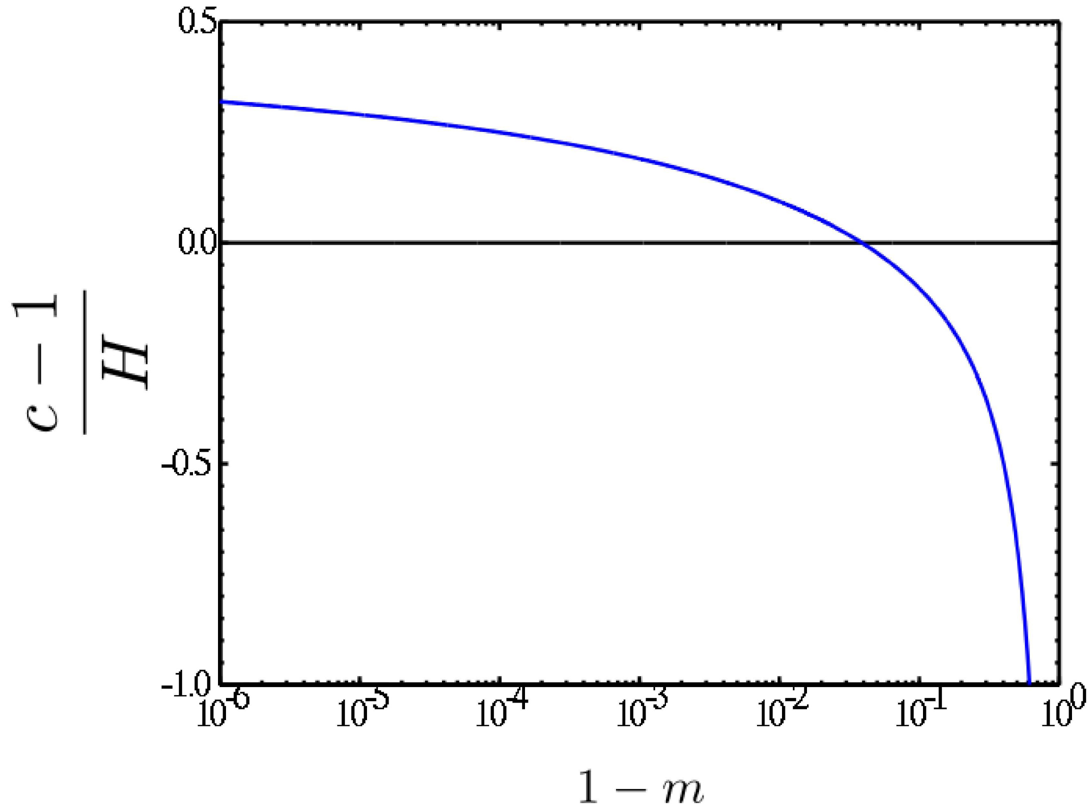
The Korteweg–de Vries equation (KdV equation), as used for water waves and in dimensional form, is:

$$\partial_t\eta + \sqrt{gh}\,\partial_x\eta + \tfrac{3}{2}\sqrt{\frac{g}{h}}\,\eta\,\partial_x\eta + \tfrac{1}{6}\,h^2\,\sqrt{gh}\,\partial_x^3\eta = 0,$$

where

η	: surface elevation, a function of x and t, with the positive direction upwards (opposing gravity),
x	: horizontal coordinate,
t	: time,
g	: the value of Earth's gravity,
h	: the mean water depth, and
∂_x and ∂_t	: partial derivative operators with respect to x and t.

Details of the derivation



Relative phase speed increase of cnoidal wave solutions for the Korteweg–de Vries equation as a function of 1−*m*, with *m* the elliptic parameter.
The horizontal axis is on a logarithmic scale, from 10^{-6} to $10^{0}=1$.
The figure is for non-dimensional quantities, *i.e.* the phase speed *c* is made dimensionless with the shallow-water phase speed $\sqrt{gh}$, and the wave height *H* is made dimensionless with the mean water depth *h*.

The cnoidal-wave solution of the KdV equation is:

$$\eta(x,t) = \eta_2 + H \operatorname{cn}^2 \left(\frac{x-ct}{\Delta} \middle| m \right),$$

with *H* the wave height—the difference between crest and trough elevation, η_2 the trough elevation, *m* the elliptic parameter, *c* the phase speed and cn one of the Jacobi elliptic functions. The trough level η_2 and width parameter Δ can be expressed in terms of *H*, *h* and *m*:

$$\eta_2 = \frac{H}{m}\left(1 - m - \frac{E(m)}{K(m)}\right), \quad \text{and} \quad \Delta = \frac{\lambda}{2\,K(m)} = h\sqrt{\frac{4}{3}\frac{m\,h}{H}},$$

with $K(m)$ the complete elliptic integral of the first kind and $E(m)$ the complete elliptic integral of the second kind. Note that $K(m)$ and $E(m)$ are denoted here as a function of the elliptic parameter m and not as a function of the elliptic modulus k, with $m = k^2$.

The wavelength λ, phase speed c and wave period τ are related to H, h and m by:

$$\lambda = h\sqrt{\frac{16}{3}\frac{m\,h}{H}}\;K(m),$$
$$c = \sqrt{gh}\left[1 + \frac{H}{m\,h}\left(1 - \frac{1}{2}m - \frac{3}{2}\frac{E(m)}{K(m)}\right)\right] \quad \text{and} \quad \tau = \frac{\lambda}{c},$$

with g the Earth's gravity.

Most often, the known wave parameters are the wave height H, mean water depth h, gravitational acceleration g, and either the wavelength λ or else the period τ. Then the above relations for λ, c and τ are used to find the elliptic parameter m. This requires numerical solution by some iterative method.

Benjamin–Bona–Mahony equation

The Benjamin–Bona–Mahony equation (BBM equation), or regularised long wave (RLW) equation, is in dimensional form given by:

$$\partial_t\eta + \sqrt{g\,h}\,\partial_x\eta + \tfrac{3}{2}\sqrt{\frac{g}{h}}\,\eta\,\partial_x\eta - \tfrac{1}{6}\,h^2\,\partial_t\,\partial_x^2\eta = 0.$$

All quantities have the same meaning as for the KdV equation. The BBM equation is often preferred over the KdV equation because it has a better short-wave behaviour.

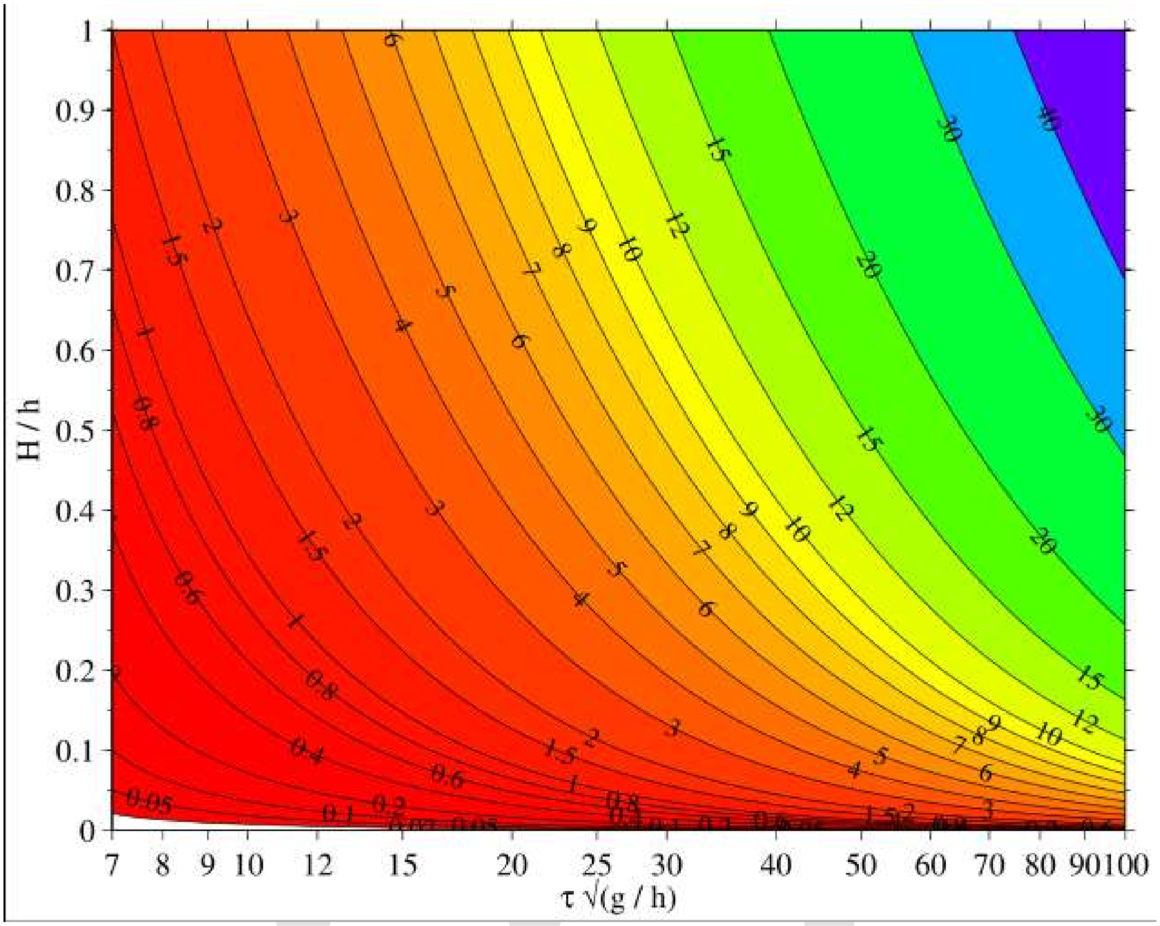
Details of the derivation

The cnoidal wave solution of the BBM equation, together with the associated relationships for the parameters is:

$$
\begin{aligned}
\eta(x,t) &= \eta_2 + H\,\mathrm{cn}^2\left(\left. \frac{x-c\,t}{\Delta} \right| m \right), \\
\eta_2 &= \frac{H}{m}\left(1 - m - \frac{E(m)}{K(m)}\right), \\
\Delta &= h\sqrt{\frac{4}{3}\,\frac{m\,h}{H}\,\frac{c}{\sqrt{g\,h}}} \qquad\qquad = \frac{\lambda}{2\,K(m)}, \\
\lambda &= h\sqrt{\frac{16}{3}\,\frac{m\,h}{H}\,\frac{c}{\sqrt{gh}}}\;K(m), \\
c &= \sqrt{gh}\,\left[1 + \frac{H}{m\,h}\left(1 - \frac{1}{2}\,m - \frac{3}{2}\,\frac{E(m)}{K(m)}\right)\right] \quad \text{and} \\
\tau &= \frac{\lambda}{c}.
\end{aligned}
$$

The only difference with the cnoidal wave solution of the KdV equation is in the equation for the wavelength λ. For practical applications, usually the water depth h, wave height H, gravitational acceleration g, and either the wavelength λ, or—most often—the period (physics) τ are provided. Then the elliptic parameter m has to be determined from the above relations for λ, c and τ through some iterative method.

Example



Parameter relations for cnoidal wave solutions of the Korteweg–de Vries equation. Shown is $-\log_{10}(1-m)$, with m the elliptic parameter of the complete elliptic integrals, as a function of dimensionless period $\tau\sqrt{(g/h)}$ and relative wave height H / h. The values along the contour lines are $-\log_{10}(1-m)$, so a value 1 corresponds with $m = 1 - 10^{-1} = 0.9$ and a value 40 with $m = 1 - 10^{-40}$.

In this example, a cnoidal wave according to the Korteweg–de Vries (KdV) equation is considered. The following parameters of the wave are given:

- mean water depth $h = 5$ m (16 ft),
- wave height $H = 3$ m (9.8 ft),
- wave period $\tau = 7$ s, and
- gravitational acceleration $g = 9.81$ m/s^2 (32 ft/s^2).

Instead of the period τ, in other cases the wavelength λ may occur as a quantity known beforehand.

First, the dimensionless period is computed:

$$\tau \sqrt{\frac{g}{h}} = 9.80,$$

which is larger than seven, so long enough for cnoidal theory to be valid. The main unknown is the elliptic parameter *m*. This has to be determined in such a way that the wave period τ, as computed from cnoidal wave theory for the KdV equation:

$$\lambda = h \sqrt{\frac{16}{3} \frac{m\,h}{H}}\, K(m),$$
$$c = \sqrt{gh}\, \left[1 + \frac{H}{m\,h} \left(1 - \frac{1}{2}\, m - \frac{3}{2}\, \frac{E(m)}{K(m)} \right) \right] \text{ and } \tau = \frac{\lambda}{c},$$

is consistent with the given value of τ; here λ is the wavelength and *c* is the phase speed of the wave. Further, *K*(*m*) and *E*(*m*) are complete elliptic integrals of the first and second kind, respectively. Searching for the elliptic parameter *m* can be done by trial and error, or by use of a numerical root-finding algorithm. In this case, starting from an initial guess $m_{\text{init}} = 0.99$, by trial and error the answer

$$m = 0.9832$$

is found. Within the process, the wavelength λ and phase speed *c* have been computed:

- wavelength λ = 50.8 m (167 ft), and
- phase speed *c* = 7.26 m/s (23.8 ft/s).

The phase speed *c* can be compared with its value $\sqrt{gh}$ according to the shallow water equations:

$$\frac{c}{\sqrt{g\,h}} = 1.0376,$$

showing a 3.8% increase due to the effect of nonlinear amplitude dispersion, which wins in this case from the reduction of phase speed by frequency dispersion.

Now the wavelength is known, the Ursell number can be computed as well:

$$U = \frac{H\,\lambda^2}{h^3} = 62,$$

which is not small, so linear wave theory is not applicable, but cnoidal wave theory is. Finally, the ratio of wavelength to depth is $\lambda / h = 10.2 > 7$, again indicating this wave is long enough to being considered as a cnoidal wave.

Solitary-wave limit

For very long nonlinear waves, with the parameter m close to one, $m \rightarrow 1$, the Jacobi elliptic function cn can be approximated by

$$\operatorname{cn}(z|m) \approx \operatorname{sech}(z) - \tfrac{1}{4}(1-m)\Big[\sinh(z)\ \cosh(z) - z\Big]\ \tanh(z)\operatorname{sech}(z),$$

with $\operatorname{sech}(z) = \dfrac{1}{\cosh(z)}.$

Here sinh, cosh, tanh and sech are hyperbolic functions. In the limit $m = 1$:

$$\operatorname{cn}(z|m) \longrightarrow \operatorname{sech}(z),$$

with $\operatorname{sech}(z) = 1 / \cosh(z)$.

Further, for the same limit of $m \rightarrow 1$, the complete elliptic integral of the first kind $K(m)$ goes to infinity, while the complete elliptic integral of the second kind $E(m)$ goes to one. This implies that the limiting values of the phase speed c and minimum elevation η_2 become:

$$c = \sqrt{g\,h}\,\left(1 + \frac{1}{2}\,\frac{H}{h}\right) \quad \text{and} \quad \eta_2 = 0.$$

Consequently, in terms of the width parameter Δ, the solitary wave solution to both the KdV and BBM equation is:

$$\eta(x,t) = H\operatorname{sech}^2\left(\frac{x - c\,t}{\Delta}\right).$$

The width parameter, as found for the cnoidal waves and now in the limit $m \rightarrow 1$, is different for the KdV and the BBM equation:

$$\Delta = h\sqrt{\frac{4\,h}{3\,H}} \qquad \text{: KdV equation, and}$$

$$\Delta = h\sqrt{\frac{4\,h}{3\,H}\,\frac{c}{\sqrt{g\,h}}} \qquad \text{: BBM equation.}$$

But the phase speed of the solitary wave in both equations is the same, for a certain combination of height H and depth h.

Limit of infinitesimal wave height

For infinitesimal wave height the results of cnoidal wave theory are expected to converge towards those of Airy wave theory for the limit of long waves $\lambda \gg h$. First the surface elevation, and thereafter the phase speed, of the cnoidal waves for infinitesimal wave height will be examined.

Surface elevation

Details of the derivation

$$n = 0 : \frac{\pi}{\sqrt{m}\, K(m)} \operatorname{sech}\left(\frac{\pi\, K'(m)}{2\, K(m)}\right) = 1 - \tfrac{1}{16}\, m - \tfrac{9}{16}\, m^2 + \cdots ,$$

$$n = 1 : \frac{\pi}{\sqrt{m}\, K(m)} \operatorname{sech}\left(\frac{3\,\pi\, K'(m)}{2\, K(m)}\right) = \tfrac{1}{16}\, m + \tfrac{1}{32}\, m^2 + \cdots ,$$

$$n = 2 : \frac{\pi}{\sqrt{m}\, K(m)} \operatorname{sech}\left(\frac{5\,\pi\, K'(m)}{2\, K(m)}\right) = \tfrac{1}{256}\, m^2 + \cdots .$$

Note: The limiting behaviour for zero m—at infinitesimal wave height—can also be seen from:

$$\operatorname{cn}(z|m) \approx \cos(z) + \tfrac{1}{4}\, m\, \big(z - \sin(z)\, \cos(z)\big)\, \sin(z) + \cdots ,$$

but the higher-order term proportional to m in this approximation contains a secular term, due to the mismatch between the period of cn($z|m$), which is 4 $K(m)$, and the period 2π for the cosine cos(z). The above Fourier series for small m does not have this drawback, and is consistent with forms as found using the Lindstedt–Poincaré method in perturbation theory.

For infinitesimal wave height, in the limit $m \to 0$, the free-surface elevation becomes:

$$\eta(x,t) = \tfrac{1}{2}\, H\, \cos\, \theta, \text{ with } \theta = 2\,\pi\, \frac{\xi}{\lambda} = 2\,\pi\, \frac{x - c\,t}{\lambda}.$$

So the wave amplitude is ½H, half the wave height. This is of the same form as studied in Airy wave theory, but note that cnoidal wave theory is only valid for long waves with their wavelength much longer than the average water depth.

Phase speed

Details of the derivation

$$\text{KdV}: \quad \lambda = h\sqrt{\frac{16}{3}\,\frac{mh}{H}}\,K(m),$$
$$\text{BBM}: \quad \lambda = h\sqrt{\frac{16}{3}\,\frac{mh}{H}\,\frac{c}{\sqrt{g\,h}}}\,K(m),$$

$$\text{KdV}: \quad c = \sqrt{gh}\,\left[1 + (\kappa\,h)^2\,\frac{4}{3\,\pi^2}\,K^2(m)\,\left(1 - \frac{1}{2}\,m - \frac{3}{2}\,\frac{E(m)}{K(m)}\right)\right],$$
$$\text{BBM}: \quad c = \sqrt{gh}\,\frac{1}{1 - (\kappa\,h)^2\,\frac{4}{3\,\pi^2}\,K^2(m)\,\left(1 - \frac{1}{2}\,m - \frac{3}{2}\,\frac{E(m)}{K(m)}\right)}.$$

The phase speeds for infinitesimal wave height, according to the cnoidal wave theories for the KdV equation and BBM equation, are

$$\text{KdV} \quad : \quad c = \left[1 - \tfrac{1}{6}\,(\kappa h)^2\right]\sqrt{g\,h},$$
$$\text{BBM} \quad : \quad c = \frac{1}{1 + \tfrac{1}{6}\,(\kappa h)^2}\,\sqrt{g\,h},$$

with $\kappa = 2\pi / \lambda$ the wavenumber and κh the relative wavenumber. These phase speeds are in full agreement with the result obtained by directly searching for sine-wave solutions of the linearised KdV and BBM equations. As is evident from these equations, the linearised BBM equation has a positive phase speed for all κh. On the other hand, the phase speed of the linearised KdV equation changes sign for short waves with $\kappa h > \sqrt{6}$. This is in conflict with the derivation of the KdV equation as a one-way wave equation.

Potential energy

The potential energy density

$$E_{\text{pot}} = \frac{1}{\lambda}\int_0^\lambda \tfrac{1}{2}\,\rho\,g\,\eta^2(x,t)\,\mathrm{d}x$$

with ρ the fluid density, is one of the infinite number of invariants of the KdV equation. This can be seen by multiplying the KdV equation with the surface elevation $\eta(x,t)$; after repeated use of the chain rule the result is:

$$\partial_t\left(\tfrac{1}{2}\,\eta^2\right)+\partial_x\left\{\tfrac{1}{2}\,\sqrt{g\,h}\,\eta^2+\tfrac{1}{2}\,\sqrt{\frac{g}{h}}\,\eta^3+\tfrac{1}{12}\,h^2\sqrt{g\,h}\,\left[\partial_x^2\left(\eta^2\right)-3\left(\partial_x\eta\right)^2\right]\right\}=0,$$

which is in conservation form, and is an invariant after integration over the interval of periodicity—the wavelength for a cnoidal wave. The potential energy is not an invariant of the BBM equation, but $\tfrac{1}{2}\rho g\,[\eta^2 + \tfrac{1}{6}\,h^2\,(\partial_x\,\eta)^2]$ is.

First the variance of the surface elevation in a cnoidal wave is computed. Note that $\eta_2 = -(1/\lambda)\,{}_0\!\int^{\lambda} H\,\mathrm{cn}^2(\xi/\Delta|\mathrm{m})\,\mathrm{d}x$, $\mathrm{cn}(\xi/\Delta|\mathrm{m}) = \cos\psi(\xi)$ and $\lambda = 2\,\Delta\,K(m)$, so

$$\begin{aligned}\frac{1}{\lambda}\int_0^{\lambda}\eta^2\,\mathrm{d}x &= \frac{1}{\lambda}\int_0^{\lambda}\left\{\eta_2+H\,\mathrm{cn}^2\left(\left.\frac{\xi}{\Delta}\right|m\right)\right\}^2\,\mathrm{d}\xi=\frac{H^2}{\lambda}\int_0^{\lambda}\mathrm{cn}^4\left(\left.\frac{\xi}{\Delta}\right|m\right)\,\mathrm{d}\xi-\eta_2^2\\ &=\frac{\Delta\,H^2}{\lambda}\int_0^{\pi}\cos^4\psi\,\frac{\mathrm{d}\xi}{\mathrm{d}\psi}\,\mathrm{d}\psi-\eta_2^2=\frac{H^2}{2\,K(m)}\int_0^{\pi}\frac{\cos^4\psi}{\sqrt{1-m\,\sin^2\psi}}\,\mathrm{d}\psi-\eta_2^2\\ &=\frac{1}{3}\,\frac{H^2}{m^2}\,\left[\left(2-5\,m+3\,m^2\right)+\left(4\,m-2\right)\,\frac{E(m)}{K(m)}\right]-\frac{H^2}{m^2}\,\left(1-m-\frac{E(m)}{K(m)}\right)^2\end{aligned}$$

The potential energy, both for the KdV and the BBM equation, is subsequently found to be

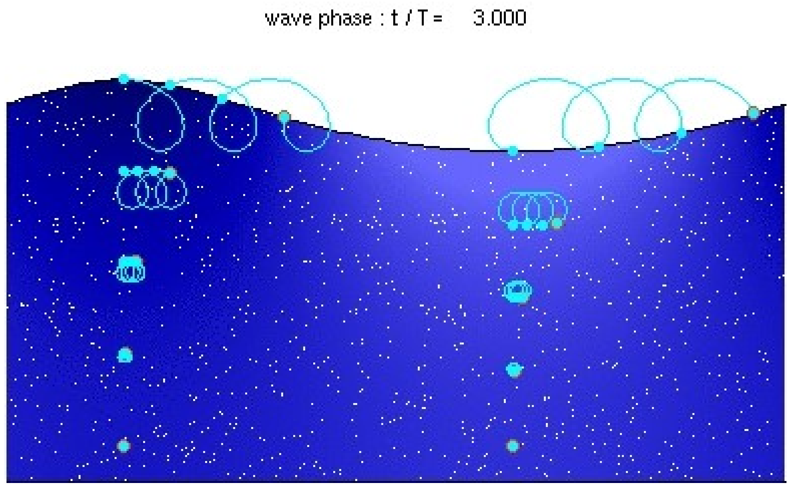
$$E_{\text{pot}}=\tfrac{1}{2}\,\rho\,g\,H^2\,\left[-\frac{1}{3\,m}+\frac{2}{3\,m}\,\left(1+\frac{1}{m}\right)\left(1-\frac{E(m)}{K(m)}\right)-\frac{1}{m^2}\,\left(1-\frac{E(m)}{K(m)}\right)^2\right].$$

The infinitesimal wave-height limit of the potential energy is $E_{pot} = \tfrac{1}{16}\,\rho\,g\,H^2$, which is in agreement with Airy wave theory. The wave height is twice the amplitude, $H = 2a$, in the infinitesimal wave limit.

Chapter 11

Stokes Drift

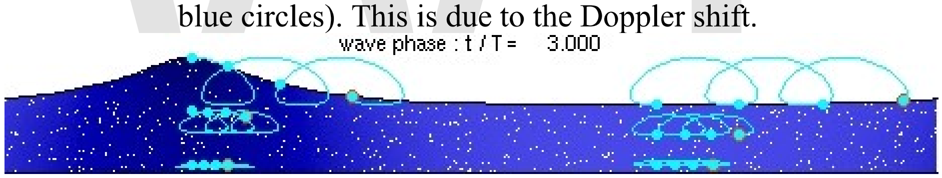
An expanse of driftwood along the northern coast of Washington state.



Stokes drift in deep water waves, with a wave length of about twice the water depth.

The red circles are the present positions of massless particles, moving with the flow velocity. The light-blue line gives the path of these particles, and the light-blue circles the particle position after each wave period. The white dots are fluid particles, also followed in time. In the case shown here, the mean Eulerian horizontal velocity below the wave trough is zero.

Observe that the wave period, experienced by a fluid particle near the free surface, is different from the wave period at a fixed horizontal position (as indicated by the light-blue circles). This is due to the Doppler shift.



Stokes drift in shallow water waves, with a wave length much longer than the water depth.

The red circles are the present positions of massless particles, moving with the flow velocity. The light-blue line gives the path of these particles, and the light-blue circles the particle position after each wave period. The white dots are fluid particles, also followed in time. In the case shown here, the mean Eulerian horizontal velocity below the wave trough is zero.

Observe that the wave period, experienced by a fluid particle near the free surface, is different from the wave period at a fixed horizontal position (as indicated by the light-blue circles). This is due to the Doppler shift.

For a pure wave motion in fluid dynamics, the **Stokes drift velocity** is the average velocity when following a specific fluid parcel as it travels with the fluid flow. For instance, a particle floating at the free surface of water waves, experiences a net Stokes drift velocity in the direction of wave propagation.

More generally, the Stokes drift velocity is the difference between the average Lagrangian flow velocity of a fluid parcel, and the average Eulerian flow velocity of the fluid at a fixed position. This nonlinear phenomenon is named after George Gabriel Stokes, who derived expressions for this drift in his 1847 study of water waves.

The **Stokes drift** is the difference in end positions, after a predefined amount of time (usually one wave period), as derived from a description in the Lagrangian and Eulerian coordinates. The end position in the Lagrangian description is obtained by following a specific fluid parcel during the time interval. The corresponding end position in the Eulerian description is obtained by integrating the flow velocity at a fixed position—equal to the initial position in the Lagrangian description—during the same time interval. The Stokes drift velocity equals the Stokes drift divided by the considered time interval. Often, the Stokes drift velocity is loosely referred to as Stokes drift. Stokes drift may occur in all instances of oscillatory flow which are inhomogeneous in space. For instance in water waves, tides and atmospheric waves.

In the Lagrangian description, fluid parcels may drift far from their initial positions. As a result, the unambiguous definition of an average Lagrangian velocity and Stokes drift velocity, which can be attributed to a certain fixed position, is by no means a trivial task. However, such an unambiguous description is provided by the *Generalized Lagrangian Mean* (GLM) theory of Andrews and McIntyre in 1978.

The Stokes drift is important for the mass transfer of all kind of materials and organisms by oscillatory flows. Further the Stokes drift is important for the generation of Langmuir circulations. For nonlinear and periodic water waves, accurate results on the Stokes drift have been computed and tabulated.

Mathematical description

The Lagrangian motion of a fluid parcel with position vector $x = \boldsymbol{\xi}(\boldsymbol{\alpha},t)$ in the Eulerian coordinates is given by:

$$\dot{\boldsymbol{\xi}} = \frac{\partial \boldsymbol{\xi}}{\partial t} = \boldsymbol{u}(\boldsymbol{\xi}, t),$$

where $\partial\xi / \partial t$ is the partial derivative of $\boldsymbol{\xi}(\boldsymbol{\alpha},t)$ with respect to t, and

$\boldsymbol{\xi}(\boldsymbol{\alpha},t)$ is the Lagrangian position vector of a fluid parcel, in meters,
$\boldsymbol{u}(\boldsymbol{x},t)$ is the Eulerian velocity, in meters per second,
$\boldsymbol{x}$ is the position vector in the Eulerian coordinate system, in meters,
$\boldsymbol{\alpha}$ is the position vector in the Lagrangian coordinate system, in meters,
t is the time, in seconds.

Often, the Lagrangian coordinates $\boldsymbol{\alpha}$ are chosen to coincide with the Eulerian coordinates $\boldsymbol{x}$ at the initial time $t = t_0$:

$$\boldsymbol{\xi}(\boldsymbol{\alpha}, t_0) = \boldsymbol{\alpha}.$$

But also other ways of labeling the fluid parcels are possible.

If the average value of a quantity is denoted by an overbar, then the average Eulerian velocity vector $\bar{\boldsymbol{u}}_E$ and average Lagrangian velocity vector $\bar{\boldsymbol{u}}_L$ are:

$$\overline{\boldsymbol{u}}_E = \overline{\boldsymbol{u}(\boldsymbol{x}, t)},$$

$$\overline{\boldsymbol{u}}_L = \overline{\dot{\boldsymbol{\xi}}(\boldsymbol{\alpha}, t)} = \overline{\left(\frac{\partial \boldsymbol{\xi}(\boldsymbol{\alpha}, t)}{\partial t}\right)} = \overline{\boldsymbol{u}(\boldsymbol{\xi}(\boldsymbol{\alpha}, t), t)}.$$

Different definitions of the average may be used, depending on the subject of study:

- time average,
- space average,
- ensemble average and
- phase average.

Now, the Stokes drift velocity $\bar{\boldsymbol{u}}_S$ equals

$$\overline{\boldsymbol{u}}_S = \overline{\boldsymbol{u}}_L - \overline{\boldsymbol{u}}_E.$$

In many situations, the mapping of average quantities from some Eulerian position $\boldsymbol{x}$ to a corresponding Lagrangian position $\boldsymbol{\alpha}$ forms a problem. Since a fluid parcel with label $\boldsymbol{\alpha}$ traverses along a path of many different Eulerian positions $\boldsymbol{x}$, it is not possible to assign $\boldsymbol{\alpha}$ to a unique $\boldsymbol{x}$. A mathematical sound basis for an unambiguous mapping between average Lagrangian and Eulerian quantities is provided by the theory of the *Generalized Lagrangian Mean* (GLM) by Andrews and McIntyre (1978).

Example: Deep water waves

The Stokes drift was formulated for water waves by George Gabriel Stokes in 1847. For simplicity, the case of infinite-deep water is considered, with linear wave propagation of a sinusoidal wave on the free surface of a fluid layer:

$$\eta = a \cos(kx - \omega t),$$

where

η is the elevation of the free surface in the z-direction (meters),
a is the wave amplitude (meters),
k is the wave number: $k = 2\pi / \lambda$ (radians per meter),
ω is the angular frequency: $\omega = 2\pi / T$ (radians per second),
x is the horizontal coordinate and the wave propagation direction (meters),

z is the vertical coordinate, with the positive z direction pointing out of the fluid layer (meters),
λ is the wave length (meters), and
T is the wave period (seconds).

It is assumed that the waves are of infinitesimal amplitude and the free surface oscillates around the mean level $z = 0$. The waves propagate under the action of gravity, with a constant acceleration vector by gravity (pointing downward in the negative z-direction). Further the fluid is assumed to be inviscid and incompressible, with a constant mass density. The fluid flow is irrotational. At infinite depth, the fluid is taken to be at rest.

Now the flow may be represented by a velocity potential φ, satisfying the Laplace equation and

$$\varphi = \frac{\omega}{k} a \, \mathrm{e}^{kz} \sin(kx - \omega t).$$

In order to have non-trivial solutions for this eigenvalue problem, the wave length and wave period may not be chosen arbitrarily, but must satisfy the deep-water dispersion relation:

$$\omega^2 = g k.$$

with g the acceleration by gravity in (m / s^2). Within the framework of linear theory, the horizontal and vertical components, ξ_x and ξ_z respectively, of the Lagrangian position $\boldsymbol{\xi}$ are:

$$\xi_x = x + \int \frac{\partial \varphi}{\partial x} \, \mathrm{d}t = x - a \, \mathrm{e}^{kz} \sin(kx - \omega t),$$
$$\xi_z = z + \int \frac{\partial \varphi}{\partial z} \, \mathrm{d}t = z + a \, \mathrm{e}^{kz} \cos(kx - \omega t).$$

The horizontal component $\bar{u}_S$ of the Stokes drift velocity is estimated by using a Taylor expansion around $\boldsymbol{x}$ of the Eulerian horizontal-velocity component $u_x = \partial \xi_x / \partial t$ at the position $\boldsymbol{\xi}$:

$$
\begin{aligned}
\overline{u}_S &= \overline{u_x(\boldsymbol{\xi},t)} - \overline{u_x(\boldsymbol{x},t)} \\
&= \overline{\left[u_x(\boldsymbol{x},t) + (\xi_x - x)\,\frac{\partial u_x(\boldsymbol{x},t)}{\partial x} + (\xi_z - z)\,\frac{\partial u_x(\boldsymbol{x},t)}{\partial z} + \cdots \right]} - \overline{u_x(\boldsymbol{x},t)} \\
&\approx \overline{(\xi_x - x)\,\frac{\partial^2 \xi_x}{\partial x\,\partial t}} + \overline{(\xi_z - z)\,\frac{\partial^2 \xi_x}{\partial z\,\partial t}} \\
&= \overline{\left[-a\,\mathrm{e}^{kz}\,\sin\,(kx - \omega t) \right] \left[-\omega\,k\,a\,\mathrm{e}^{kz}\,\sin\,(kx - \omega t) \right]} \\
&\quad + \overline{\left[a\,\mathrm{e}^{kz}\,\cos\,(kx - \omega t) \right] \left[\omega\,k\,a\,\mathrm{e}^{kz}\,\cos\,(kx - \omega t) \right]} \\
&= \overline{\omega\,k\,a^2\,\mathrm{e}^{2kz} \left[\sin^2\,(kx - \omega t) + \cos^2\,(kx - \omega t) \right]}.
\end{aligned}
$$

Performing the averaging, the horizontal component of the Stokes drift velocity for deep-water waves is approximately:

$$
\overline{u}_S \approx \omega\,k\,a^2\,\mathrm{e}^{2kz} = \frac{4\pi^2\,a^2}{\lambda T}\,\mathrm{e}^{4\pi\,z/\lambda}.
$$

As can be seen, the Stokes drift velocity $\bar{u}_S$ is a nonlinear quantity in terms of the wave amplitude *a*. Further, the Stokes drift velocity decays exponentially with depth: at a depth of a quart wavelength, $z = -¼\,\lambda$, it is about 4% of its value at the mean free surface, $z = 0$.

Chapter 12

Breaking Wave

A large wave breaking.

In physics, a **breaking wave** is a wave whose amplitude reaches a critical level at which some process can suddenly start to occur that causes large amounts of wave energy to be transformed into turbulent kinetic energy. At this point, simple physical models that describe wave dynamics often become invalid, particularly those that assume linear behavior.

The most generally familiar sort of breaking wave is the breaking of water surface waves on a coastline. Because of the horizontal component of the fluid velocity associated with the wave motion, wave crests steepen as the amplitude increases; wave breaking generally occurs where the amplitude reaches the point that the crest of the wave actually overturns — though the types of breaking water surface waves are discussed in more detail below. Certain other effects in fluid dynamics have also been termed "breaking waves," partly by analogy with water surface waves. In meteorology, gravity waves are said to break when the wave produces regions where the potential temperature decreases with height, leading to energy dissipation through convective instability; likewise Rossby waves are said to break when the potential vorticity gradient is overturned. Wave breaking also occurs in plasmas, when the particle velocities exceed the wave's phase speed.

An analogous effect in electromagnetic waves, although not commonly termed *wave breaking*, is when the electric field amplitude associated with the wave is sufficient to cause dielectric breakdown and sparking (such as when sharp metal objects are placed in a microwave oven).

Breaking water surface waves

Breaking waves at the village of Porto Covo, west coast of Portugal.

Breaking wave on a slope in a laboratory wave channel (movie)

Breaking of water surface waves may occur anywhere that the amplitude is sufficient, including in mid-ocean. However, it is particularly common on beaches because waves are amplified in the region of shallower water (because the group velocity is lower there).

There are four basic types of breaking water waves. They are spilling, plunging, collapsing, and surging.

Spilling breakers

When the ocean floor has a gradual slope, the wave will steepen until the crest becomes unstable, resulting in turbulent whitewater spilling down the face of the wave. This continues as the wave approaches the shore, and the wave's energy is slowly dissipated in the whitewater. Because of this, spilling waves break for a longer time than other waves, and create a relatively gentle wave. Onshore wind conditions make spillers more likely.

Plunging breakers

A plunging wave occurs when the ocean floor is steep or has sudden depth changes, such as from a reef or sandbar. The crest of the wave becomes much steeper than a spilling wave, becomes vertical, then curls over and drops onto the trough of the wave, releasing most of its energy at once in a relatively violent impact. A plunging wave breaks with more energy than a significantly larger spilling wave. The wave can trap and compress the air under the lip, which creates the "crashing" sound associated with waves. With large waves, this crash can be felt by beachgoers on land. Offshore wind conditions can make plungers more likely.

If a plunging wave is not parallel to the beach (or the ocean floor), the section of the wave which reaches shallow water will break first, and the breaking section (or curl) will move laterally across the face of the wave as the wave continues. This is the "tube" that is so highly sought after by surfers (also called a "barrel", a "pit", and "the greenhouse", among other terms). The surfer tries to stay near or under the curling lip, often trying to stay as "deep" in the tube as possible while still being able to shoot forward and exit the tube before it closes. A plunging wave that is parallel to the beach can break along its whole length at once, rendering it unrideable and dangerous. Surfers refer to these waves as "closed out".

Surging

On steeper beaches, the energy of the wave can be reflected by the bottom back into the ocean, causing standing waves. This can result in a very narrow surf zone, or no breaking waves at all.

Collapsing

Collapsing waves are a cross between plunging and surging, in which the crest never fully breaks, yet the bottom face of the wave gets steeper and collapses, resulting in foam.

Physics

During breaking, a deformation (usually a bulge) forms at the wave crest, either leading side of which is known as the "toe." Parasitic capillary waves are formed, with short wavelengths. Those above the "toe" tend to have much longer wavelengths. This theory is anything but perfect, however, as it's linear. There have been a couple non-linear theories of motion (regarding waves). One put forth uses a perturbation method to expand the description all the way to the third order, and better solutions have been found since then. As for wave deformation, methods much like the boundary integral method and the Boussinesq model have been created.

It has been accounted for, that the high-frequencies detail present in a breaking wave play a part in crest deformation and destabilzation. The same theory expands on this, stating that the valleys of the capillary waves create a source for vorticity. It is said that surface tension (and viscosity) are significant for waves up to 2m in wavelength.

These models are flawed, however, as they can't take into account what happens to the water after the wave breaks. Post-break eddy forms and the turbulence created via the breaking is mostly unresearched. Understandably, it might be difficult to glean predictable results from the ocean.

After the tip of the wave overturns and the jet collapses, it creates a very coherent and defined horizontal vertex. The plunging breakers create secondary eddies down the face of the wave. Small horizontal random eddides that form on the sides of the wave suggest that, perhaps, prior to breaking, the water's velocity is more or less two dimensional. This becomes three dimensional upon breaking.

The main vortex along the front of the wave diffuses rapidly into the interior of the wave after breaking, as the eddies on the surface become more viscous. Advection and molecular diffusion play a part in stretching the vortex and redistributing the vorticity, as well as the formation turbulence cascades. The energy of the large vortices are, by this method, is transferred to much smaller isotropic vortices.

Experiments have been conducted to deduce the evolution of turbulence after break, both in deep water and on a beach.

Chapter 13

Radiation Stress

Breaking waves on beaches induce variations in radiation stress, driving longshore currents. The resulting longshore sediment transport shapes the beaches, and may result in beach erosion or accretion.

In fluid dynamics, the **radiation stress** is the depth-integrated – and thereafter phase-averaged – excess momentum flux caused by the presence of the surface gravity waves exerted on the mean flow. The radiation stresses behave as a second-order tensor.

The radiation stress tensor describes the additional forcing of the mean depth-integrated horizontal momentum in the fluid layer, due to the presence of the waves. As a result, for varying radiation stresses, there may be changes in the mean surface elevation (wave setup) and the mean flow (wave-induced currents).

For the mean energy density in the oscillatory part of the fluid motion, the radiation stress tensor is important for the dynamics in case of an inhomogeneous mean-flow field.

The radiation stress tensor, as well as several of its implications on the physics of surface gravity waves and mean flows, were formalized in a series of papers by Longuet-Higgins and Stewart in 1960–1964.

Radiation stress derives its name from the analogous effect of radiation pressure for electromagnetic radiation.

Physical significance

The radiation stress – mean excess momentum-flux due to the presence of the waves – plays an important role in the explanation and modeling of various coastal processes:

- *Wave setup* and *setdown* – the radiation stress consists in part of a radiation pressure, exerted at the free surface elevation of the mean flow. If the radiation stress varies spatially, as it does in the surf zone where the wave height reduces by wave breaking, this results in changes of the mean surface elevation called wave setup (in case of an increased level) and setdown (for a decreased water level);
- *Wave-driven current*, especially a *longshore current* in the surf zone – for oblique incidence of waves on a beach, the reduction in wave height inside the surf zone (by breaking) introduces a variation of the shear-stress component S_{xy} of the radiation stress over the width of the surf zone. This provides the forcing of a wave-driven longshore current, which is of importance for sediment transport (longshore drift) and the resulting coastal morphology;
- *Bound long waves* or *forced long waves* – for wave groups the radiation stress varies along the group. As a result, a non-linear long wave propagates together with the group, at the group velocity of the modulated short waves within the group. While, according to the dispersion relation, a long wave of this length should propagate at its own – higher – phase velocity. The amplitude of this bound long wave varies with the square of the wave height, and is only significant in shallow water;
- *Wave–current interaction* – in varying mean-flow fields, the energy exchanges between the waves and the mean flow, as well as the mean-flow forcing, can be modeled by means of the radiation stress.

Definitions and values derived from linear wave theory

One-dimensional wave propagation

For uni-directional wave propagation – say in the x-coordinate direction – the component of the radiation stress tensor of dynamical importance is S_{xx}. It is defined as:

$$S_{xx} = \overline{\int_{-h}^{\eta} \left(p + \rho u^2\right)\, \mathrm{d}z} - \frac{1}{2}\rho g \left(h + \overline{\eta}\right)^2,$$

where $p(x,z,t)$ is the fluid pressure, $u(x,z,t)$ is the horizontal x-component of the oscillatory part of the flow velocity vector, z is the vertical coordinate, t is time, $z = -h(x)$ is the bed elevation of the fluid layer, and $z = \eta(x,t)$ is the surface elevation. Further ρ is the fluid density and g is the acceleration by gravity, while an overbar denotes phase averaging. The last term on the right-hand side, $\frac{1}{2}\rho g(h+\eta)^2$, is the integral of the hydrostatic pressure over the still-water depth.

To lowest (second) order, the radiation stress S_{xx} for traveling periodic waves can be determined from the properties of surface gravity waves according to Airy wave theory:

$$S_{xx} = \left(2\frac{C_g}{C_p} - \frac{1}{2}\right) E,$$

where C_p is the phase speed and C_g is the group speed of the waves. Further E is the mean depth-integrated wave energy density (the sum of the kinetic and potential energy) per unit of horizontal area. From the results of Airy wave theory, to second order, the mean energy density E equals:

$$E = \frac{1}{2}\rho g a^2 = \frac{1}{8}\rho g H^2,$$

with a the wave amplitude and $H = 2a$ the wave height. Note this equation is for periodic waves: in random waves the root-mean-square wave height H_{rms} should be used with $H_{\mathrm{rms}} = H_{\mathrm{m0}} / \sqrt{2}$, where H_{m0} is the significant wave height. Then $E = \frac{1}{16}\rho g H_{\mathrm{m0}}^2$.

Two-dimensional wave propagation

For wave propagation in two horizontal dimensions the radiation stress $\mathbb{S}$ is a second-order tensor with components:

$$\mathbb{S} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}.$$

With, in a cartesian coordinate system (x,y,z):

$$S_{xx} = \overline{\int_{-h}^{\eta} \left(p + \rho u^2\right)\, \mathrm{d}z} - \frac{1}{2}\rho g \left(h + \overline{\eta}\right)^2,$$
$$S_{xy} = \overline{\int_{-h}^{\eta} \left(\rho u v\right)\, \mathrm{d}z} = S_{yx},$$
$$S_{yy} = \overline{\int_{-h}^{\eta} \left(p + \rho v^2\right)\, \mathrm{d}z} - \frac{1}{2}\rho g \left(h + \overline{\eta}\right)^2,$$

where u and v are the horizontal x- and y-components of the oscillatory part $\mathbf{u}(x,y,z,t)$ of the flow velocity vector.

To second order – in wave amplitude a – the components of the radiation stress tensor for progressive periodic waves are:

$$S_{xx} = \left[\frac{k_x^2}{k^2}\frac{C_g}{C_p} + \left(\frac{C_g}{C_p} - 1\right)\right] E,$$
$$S_{xy} = \left(\frac{k_x k_y}{k^2}\frac{C_g}{C_p}\right) E = S_{yx}, \quad \text{and}$$
$$S_{yy} = \left[\frac{k_y^2}{k^2}\frac{C_g}{C_p} + \left(\frac{C_g}{C_p} - 1\right)\right] E,$$

where k_x and k_y are the x- and y-components of the wavenumber vector $\mathbf{k}$, with length $k = |\mathbf{k}| = \sqrt{k_x^2+k_y^2}$ and the vector $\mathbf{k}$ perpendicular to the wave crests. The phase and group speeds, C_p and C_g respectively, are the lengths of the phase and group velocity vectors: $C_p = |\mathbf{C}_p|$ and $C_g = |\mathbf{C}_g|$.

Dynamical significance

The radiation stress tensor is an important quantity in the description of the phase-averaged dynamical interaction between waves and mean flows. Here, the depth-integrated dynamical conservation equations are given, but – in order to model three-dimensional mean flows forced, or interacting with, surface waves – a three-dimensional description of the radiation stress over the fluid layer is needed.

Mass transport velocity

Propagating waves induce a – relatively small – mean mass transport in the wave propagation direction, also called the wave (pseudo) momentum. To lowest order, the wave momentum $\mathbf{M}_w$ is, per unit of horizontal area:

$$\boldsymbol{M}_w = \frac{\boldsymbol{k}}{k} \frac{E}{\rho C_p},$$

which is exact for progressive waves of permanent form in irrotational flow. Above, C_p is the phase speed relative to the mean flow:

$$C_p = \frac{\sigma}{k} \qquad \text{with} \qquad \sigma = \omega - \boldsymbol{k} \cdot \boldsymbol{V},$$

with σ the *intrinsic angular frequency*, as seen by an observer moving with the mean flow **V**. While ω is the *apparent angular frequency* of an observer at rest (with respect to 'Earth'). The difference $\mathbf{k} \cdot \mathbf{V}$ is the Doppler shift.

The mean momentum **M**, also per unit of horizontal area, is the mean value of the integral of momentum over depth:

$$\boldsymbol{M} = \overline{\int_{-h}^{\eta} \rho \boldsymbol{v} \, \mathrm{d}z},$$

with $\mathbf{v}(x,y,z,t)$ the total flow velocity at any point below the free surface $z = \eta(x,y,t)$.

Now the mass transport velocity **U** is defined as:

$$\boldsymbol{U} = \frac{\boldsymbol{M}}{\rho \left(h + \overline{\eta}\right)}.$$

Observe that first the depth-integrated horizontal momentum is averaged, before the division by the mean water depth $(h+\eta)$ is made.

Mass and momentum conservation

Vector notation

The equation of mean mass conservation is, in vector notation:

$$\frac{\partial}{\partial t} \left[\rho \left(h + \overline{\eta}\right)\right] + \nabla \cdot \left[\rho \left(h + \overline{\eta}\right) \boldsymbol{U}\right] = 0,$$

with **U** including the contribution of the wave momentum $\mathbf{M}_w$.

The equation for the conservation of horizontal mean momentum is:

$$\frac{\partial}{\partial t} \left[\rho \left(h + \overline{\eta}\right) \boldsymbol{U}\right] + \nabla \cdot \left[\rho \left(h + \overline{\eta}\right) \boldsymbol{U} \otimes \boldsymbol{U} + \mathbb{S}\right] = -\rho g \left(h + \overline{\eta}\right) \nabla h + \boldsymbol{\tau}_w - \boldsymbol{\tau}_b,$$

where $\mathbf{U} \otimes \mathbf{U}$ denotes the tensor product of $\mathbf{U}$ with itself, and $\boldsymbol{\tau}_w$ is the mean wind shear stress at the free surface, while $\boldsymbol{\tau}_b$ is the bed shear stress. Note that the right hand side of the momentum equation provides the non-conservative contributions of the bed slope ∇h, as well the forcing by the wind and the bed friction.

In terms of the horizontal momentum $\mathbf{M}$ the above equations become:

$$\frac{\partial}{\partial t}\left[\rho\left(h+\bar{\eta}\right)\right]+\nabla\cdot\boldsymbol{M}=0,$$
$$\frac{\partial \boldsymbol{M}}{\partial t}+\nabla\cdot\left[\boldsymbol{U}\otimes\boldsymbol{M}+\mathbb{S}\right]=-\rho g\left(h+\bar{\eta}\right)\nabla h+\boldsymbol{\tau}_w-\boldsymbol{\tau}_b.$$

Component form in cartesian coordinates

In a cartesian coordinate system, the mass conservation equation becomes:

$$\frac{\partial}{\partial t}\left[\rho\left(h+\bar{\eta}\right)\right]+\frac{\partial}{\partial x}\left[\rho\left(h+\bar{\eta}\right)U_x\right]+\frac{\partial}{\partial y}\left[\rho\left(h+\bar{\eta}\right)U_y\right]=0,$$

with U_x and U_y respectively the x and y components of the mass transport velocity $\mathbf{U}$.

The horizontal momentum equations are:

$$\begin{aligned}\frac{\partial}{\partial t}\left[\rho\left(h+\bar{\eta}\right)U_x\right]+\frac{\partial}{\partial x}\left[\rho\left(h+\bar{\eta}\right)U_xU_x+S_{xx}\right]+\frac{\partial}{\partial y}\left[\rho\left(h+\bar{\eta}\right)U_xU_y+S_{xy}\right]\\ =-\rho g\left(h+\bar{\eta}\right)\frac{\partial}{\partial x}h+\tau_{w,x}-\tau_{b,x},\\ \frac{\partial}{\partial t}\left[\rho\left(h+\bar{\eta}\right)U_y\right]+\frac{\partial}{\partial x}\left[\rho\left(h+\bar{\eta}\right)U_yU_x+S_{yx}\right]+\frac{\partial}{\partial y}\left[\rho\left(h+\bar{\eta}\right)U_yU_y+S_{yy}\right]\\ =-\rho g\left(h+\bar{\eta}\right)\frac{\partial}{\partial y}h+\tau_{w,y}-\tau_{b,y}.\end{aligned}$$

Energy conservation

For a inviscid flow the mean mechanical energy of the total flow – that is the sum of the energy of the mean flow and the fluctuating motion – is conserved. However, the mean energy of the fluctuating motion itself is not conserved, nor is the energy of the mean flow. The mean energy E of the fluctuating motion (the sum of the kinetic and potential energies satisfies:

$$\frac{\partial E}{\partial t}+\nabla\cdot\left[\left(\boldsymbol{U}+\boldsymbol{C}_g\right)E\right]+\mathbb{S}:\left(\nabla\otimes\boldsymbol{U}\right)=\boldsymbol{\tau}_w\cdot\boldsymbol{U}-\boldsymbol{\tau}_b\cdot\boldsymbol{U}-\varepsilon,$$

where ":" denotes the double-dot product, and ε denotes the dissipation of mean mechanical energy (for instance by wave breaking). The term $\mathbb{S} : (\nabla \otimes \boldsymbol{U})$ is the exchange of energy with the mean motion, due to wave–current interaction. The mean horizontal wave-energy transport $(\mathbf{U} + \mathbf{C}_g)\, E$ consists of two contributions:

- $\mathbf{U}\, E$: the transport of wave energy by the mean flow, and
- $\mathbf{C}_g\, E$: the mean energy transport by the waves themselves, with the group velocity $\mathbf{C}_g$ as the wave-energy transport velocity.

In a cartesian coordinate system, the above equation for the mean energy E of the flow fluctuations becomes:

$$\begin{aligned} &\frac{\partial E}{\partial t} + \frac{\partial}{\partial x}\left[\left(U_x + C_{g,x}\right) E\right] + \frac{\partial}{\partial y}\left[\left(U_y + C_{g,y}\right) E\right] \\ &\quad + S_{xx}\frac{\partial U_x}{\partial x} + S_{xy}\left(\frac{\partial U_y}{\partial x} + \frac{\partial U_x}{\partial y}\right) + S_{yy}\frac{\partial U_y}{\partial y} \\ &\quad = \left(\tau_{w,x} - \tau_{b,x}\right) U_x + \left(\tau_{w,y} - \tau_{b,y}\right) U_y - \varepsilon. \end{aligned}$$

Chapter 14

Internal Wave & Kelvin Wave

Internal Wave

Internal waves (marked with arrows) caused by the Strait of Gibraltar

Internal waves are gravity waves that oscillate within, rather than on the surface of, a fluid medium. They are one of many types of wave motion in stratified fluids (another example being Lee waves). A simple example is a wave propagating on the interface between two fluids of different densities, such as oil and water. Internal wave motions are ubiquitous in both the ocean and atmosphere, where they create wave clouds. Nonlinear solitary internal waves are called solitons.

Gravity versus buoyancy

Suppose a water column is in hydrostatic equilibrium and displace a small packet of fluid with density ρ_0 vertically by a distance Δz. The balance between the force of gravity and the buoyant restoring force is disturbed, and the motion of the packet obeys the equation

$$\rho_0 \frac{d^2 \Delta z}{dt^2} = g \frac{d\rho_0}{dz} \Delta z,$$

where g is the gravitational acceleration. The packet oscillates at the Brunt–Väisälä frequency:

$$N = \left(-\frac{g}{\rho_0} \frac{d\rho_0}{dz} \right)^{1/2}.$$

The Brunt–Väisälä frequency is an upper limit for the frequency in actual internal waves.

Internal waves in the ocean

Most people think of waves as a surface phenomenon, which acts between water (as in lakes or oceans) and the air. Where low density water overlies high density water in the ocean, internal waves propagate along the boundary. They are especially common over the continental shelf regions of the world oceans and where brackish water overlies salt water at the outlet of large rivers. There is typically little surface expression of the waves, aside from slick bands that can form over the trough of the waves.

Internal waves are the source of a curious phenomenon called **dead water**, first reported by the Norwegian oceanographer Fridtjof Nansen, in which a boat may experience strong resistance to forward motion in apparently calm conditions. This occurs when the ship is sailing on a layer of relatively fresh water whose depth is comparable to the ship's draft. This causes a wake of internal waves that dissipates a lot of energy.

Properties of internal waves

Internal waves typically have much lower frequencies and higher amplitudes than surface gravity waves because the density differences (and therefore the restoring forces) within a fluid are usually much smaller than the density of the fluid itself. Wavelengths vary from centimetres to kilometres with periods of seconds to hours respectively.

The atmosphere and ocean are continuously stratified: potential density generally increases steadily downward. Internal waves in a continuously stratified medium may propagate vertically as well as horizontally. The dispersion relation for such waves is curious: For a freely-propagating internal wave packet, the direction of propagation of energy (group velocity) is perpendicular to the direction of propagation of wave crests and troughs (phase velocity). An internal wave may also become confined to a finite region of altitude or depth, as a result of varying stratification or wind. Here, the wave is said to be *ducted* or *trapped*, and a vertically standing wave may form, where the vertical component of group velocity approaches zero. A ducted internal wave *mode* may propagate horizontally, with parallel group and phase velocity vectors, analogous to propagation within a waveguide.

At large scales, internal waves are influenced both by the rotation of the Earth as well as by the stratification of the medium. The frequencies of these geophysical wave motions vary from a lower limit of the Coriolis frequency (inertial motions) up to the Brunt–Väisälä, or buoyancy frequency (buoyancy oscillations). Above the Brunt–Väisälä frequency may exist evanescent internal wave motions, for example those resulting from partial reflection. Internal waves at tidal frequencies are produced by tidal flow over topography/bathymetry, and are known as internal tides. Similarly, Atmospheric tides arise from, for example, non-uniform solar heating associated with diurnal motion.

Kelvin Wave

A **Kelvin wave** is a wave in the ocean or atmosphere that balances the Earth's Coriolis force against a topographic boundary such as a coastline, or a waveguide such as the equator. A feature of a Kelvin wave is that it is non-dispersive, i.e., the phase speed of the wave crests is equal to the group speed of the wave energy for all frequencies. This means that it retains its shape in the alongshore direction over time.

A Kelvin wave (fluid dynamics) is also a long scale perturbation mode of a vortex in superfluid dynamics; in terms of the meteorological or oceanographical derivation, one may assume that the meridional velocity component vanishes (i.e. there is no flow in the north–south direction, thus making the momentum and continuity equations much simpler).

Coastal Kelvin wave

In a stratified ocean of mean depth H, free waves propagate along coastal boundaries (and hence become trapped in the vicinity of the coast itself) in the form of internal Kelvin waves on a scale of about 30 km. These waves are called coastal Kelvin waves, and have propagation speeds of approximately 2 m/s in the ocean. Utilizing the assumption that the cross-shore velocity *v* is zero at the coast, *v* = 0, one may solve a frequency relation for the phase speed of coastal Kelvin waves, which are among the class of waves called boundary waves, edge waves, trapped waves, or surface waves (similar to the Lamb waves). The (linearised) primitive equations then become the following:

- the continuity equation (accounting for the effects of horizontal convergence and divergence):

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = \frac{-1}{H}\frac{\partial \eta}{\partial t}$$

- the *u*-momentum equation (zonal wind component):

$$\frac{\partial u}{\partial t} = -g\frac{\partial \eta}{\partial x} + fv$$

- the *v*-momentum equation (meridional wind component):

$$\frac{\partial v}{\partial t} = -g\frac{\partial \eta}{\partial y} - fu.$$

If one assumes that the Coriolis coefficient *f* is constant along the right boundary conditions and the zonal wind speed is set equal to zero, then the primitive equations become the following:

- the continuity equation:

$$\frac{\partial v}{\partial y} = \frac{-1}{H}\frac{\partial \eta}{\partial t}$$

- the *u*-momentum equation:

$$g\frac{\partial \eta}{\partial x} = fv$$

- the *v*-momentum equation:

$$\frac{\partial v}{\partial t} = -g\frac{\partial \eta}{\partial y}.$$

The solution to these equations yields the following phase speed: $c^2 = gH$, which is the same speed as for shallow-water gravity waves without the effect of Earth's rotation. It is important to note that for an observer traveling with the wave, the coastal boundary (maximum amplitude) is always to the right in the northern hemisphere and to the left in the southern hemisphere (i.e. these waves move equatorward/southward – negative phase speed – on a western boundary and poleward/northward – positive phase speed – on an eastern boundary; the waves move cyclonically around an ocean basin).

Equatorial Kelvin wave

The equatorial zone essentially acts as a waveguide, causing disturbances to be trapped in the vicinity of the equator, and the equatorial Kelvin wave illustrates this fact because the equator acts analogously to a topographic boundary for both the Northern and Southern Hemispheres, making this wave very similar to the coastally-trapped Kelvin wave. The primitive equations are identical to those used to develop the coastal Kelvin wave phase speed solution (U-momentum, V-momentum, and continuity equations) and the motion is unidirectional and parallel to the equator. Because these waves are equatorial, the Coriolis parameter vanishes at 0 degrees; therefore, it is necessary to use the equatorial beta plane approximation that states:

$$f = \beta y,$$

where β is the variation of the Coriolis parameter with latitude. This equatorial Beta plane assumption requires a geostrophic balance between the eastward velocity and the north-south pressure gradient. The phase speed is identical to that of coastal Kelvin waves, indicating that the equatorial Kelvin waves propagate toward the east without dispersion (as if the earth were a non-rotating planet). For the first baroclinic mode in the ocean, a typical phase speed would be about 2.8 m/s, causing an equatorial Kelvin wave to take 2 months to cross the Pacific Ocean between New Guinea and South America; for higher ocean and atmospheric modes, the phase speeds are comparable to fluid flow speeds.

When the motion at the equator is to the east, any deviation toward the north is brought back toward the equator because the Coriolis force acts to the right of the direction of motion in the Northern Hemisphere, and any deviation to the south is brought back toward the equator because the Coriolis force acts to the left of the direction of motion in the Southern Hemisphere. Note that for motion toward the west, the Coriolis force would not restore a northward or southward deviation back toward the equator; thus, equatorial Kelvin waves are only possible for eastward motion (as noted above). Both atmospheric and oceanic equatorial Kelvin waves play an important role in the dynamics of El Nino-Southern Oscillation, by transmitting changes in conditions in the Western Pacific to the Eastern Pacific.

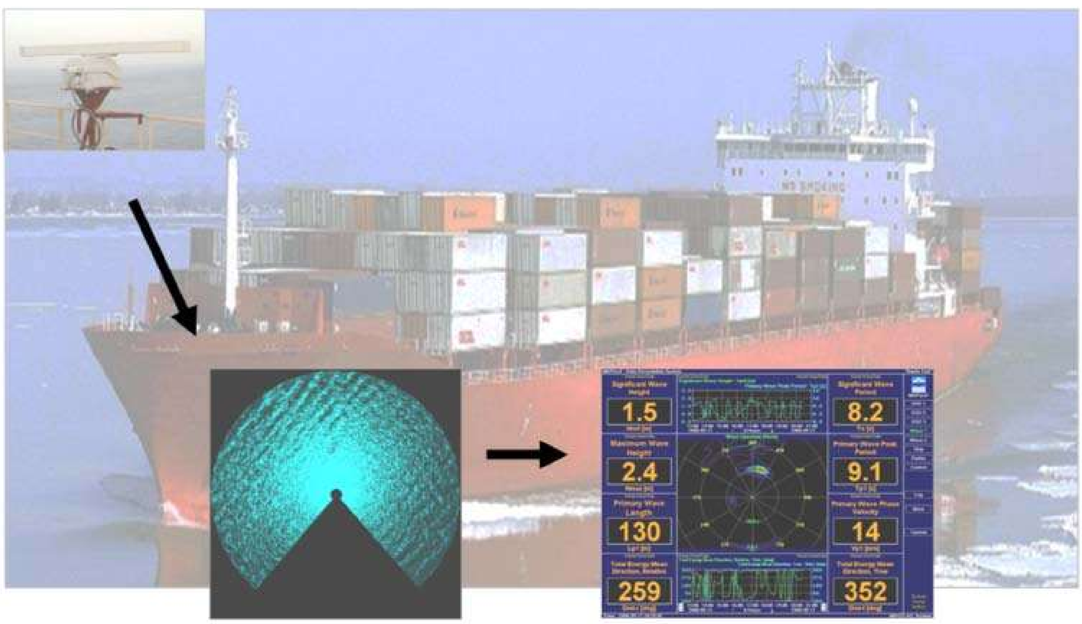
There have been studies that connect equatorial Kelvin waves to coastal Kelvin waves. Moore (1968) found that as an equatorial Kelvin wave strikes an "eastern boundary," part of the energy is reflected in the form of planetary and gravity waves; and the remainder of the energy is carried poleward along the eastern boundary as coastal Kelvin waves.

This process indicates that some energy may be lost from the equatorial region and transported to the poleward region.

Equatorial Kelvin waves are often associated with anomalies in surface wind stress. For example, positive (eastward) anomalies in wind stress in the central Pacific excite positive anomalies in 20°C isotherm depth which propagate to the east as equatorial Kelvin waves.

Chapter 15

Wave Radar



Measuring ocean waves by use of marine radars

Wind waves can be measured by several radar remote sensing techniques. Several instruments based on a variety of different concepts and techniques are available to the user and these are all often called **wave radars**.

Instruments based on radar remote sensing techniques have become of particular interest in applications where it is important to avoid direct contact with the water surface and avoid structural interference. A typical case is wave measurements from an offshore platform in deep waters with the presence of high currents making the mooring of a wave buoy enormously difficult. Another interesting case is a ship in transit where having instruments in the sea is highly impractical and interference from the ships hull must be avoided.

Radar remote sensing

Terms and definitions

Basically there are two different *classes* of radar remote sensors for ocean waves.

- **Direct sensor** measures directly some relevant parameter of the wave system (like surface elevation or water particle velocity).
- **Indirect sensors** observe the surface waves via the interaction with some other physical process as for example the radar cross section of the sea surface.

Microwave radars may be used in two different *modes*;

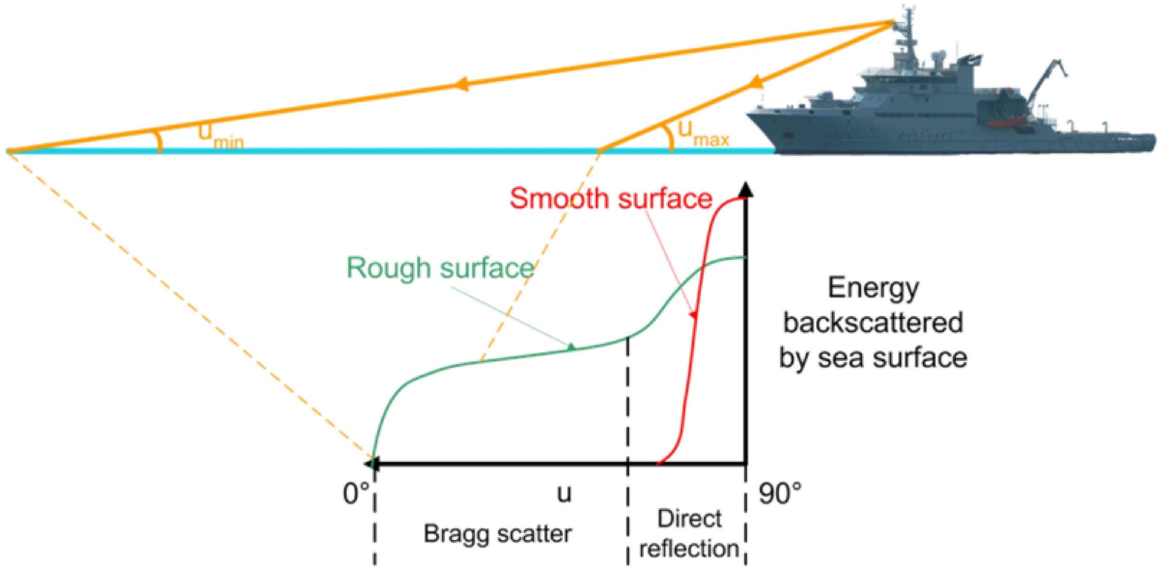
- The near **vertical mode**. The radar echo is generated by specular reflections from the sea surface.
- The **low grazing angle mode**. The radar echo is generated by Bragg scattering, hence wind generated surface ripple (capillary waves)must be present. The backscattered signal will be modulated by the large surface gravity waves and the gravity wave information is derived from the modulation of the backscattered signal. An excellent presentation of the theories of microwave remote sensing of the sea surface is given by Plant and Shuler (1980).

The radar footprint (the radial and azimuthal extent of the surface area to be illuminated by the radar) must be small in comparison with all ocean wavelength of interest. The radar spatial resolution is determined by the bandwidth of the radar signal and the beamwidth of the radar antenna.

The beam of a microwave antenna is dispersive, consequently the resolution becomes a function of range. The beam of an IR radar (laser) is non dispersive, the radar footprint is therefore independent of range.

HF radars utilize the Bragg scattering mechanism and do always operate at very low grazing angles. Due to the low frequency of operation the radar waves are backscattered directly from the gravity waves and surface ripple need not be present.

Radar transceivers may be coherent or non-coherent. Coherent radars measure Doppler-modulation as well as amplitude modulation, while non-coherent radars only measure amplitude modulation. Consequently, a non-coherent radar echo contains less information about the sea surface properties. Examples of non-coherent radars are conventional marine navigation radars.



Energy backscattered from sea surface as a function of angle.

The radar transmitter waveform may be either unmodulated continuous wave, modulated or pulsed. An unmodulated continuous wave radar has no range resolution, but can resolve targets on the basis of different velocity, while a modulated or pulsed radar can resolve echoes from different ranges. The radar waveform plays a very important role in radar theory (Plant and Shuler, 1980)).

The wave radar performance is highly dependent on

- Mode of operation or measurement geometry (vertical or grazing)
- Class of system (direct or indirect)
- Frequency of operation
- Radar waveform (unmodulated CW or modulated/pulsed)
- Type of transceiver (coherent or non-coherent)
- Radar antenna properties

Remote sensing techniques

An excellent survey of different radar techniques for remote sensing of waves is given by Tucker (1991).

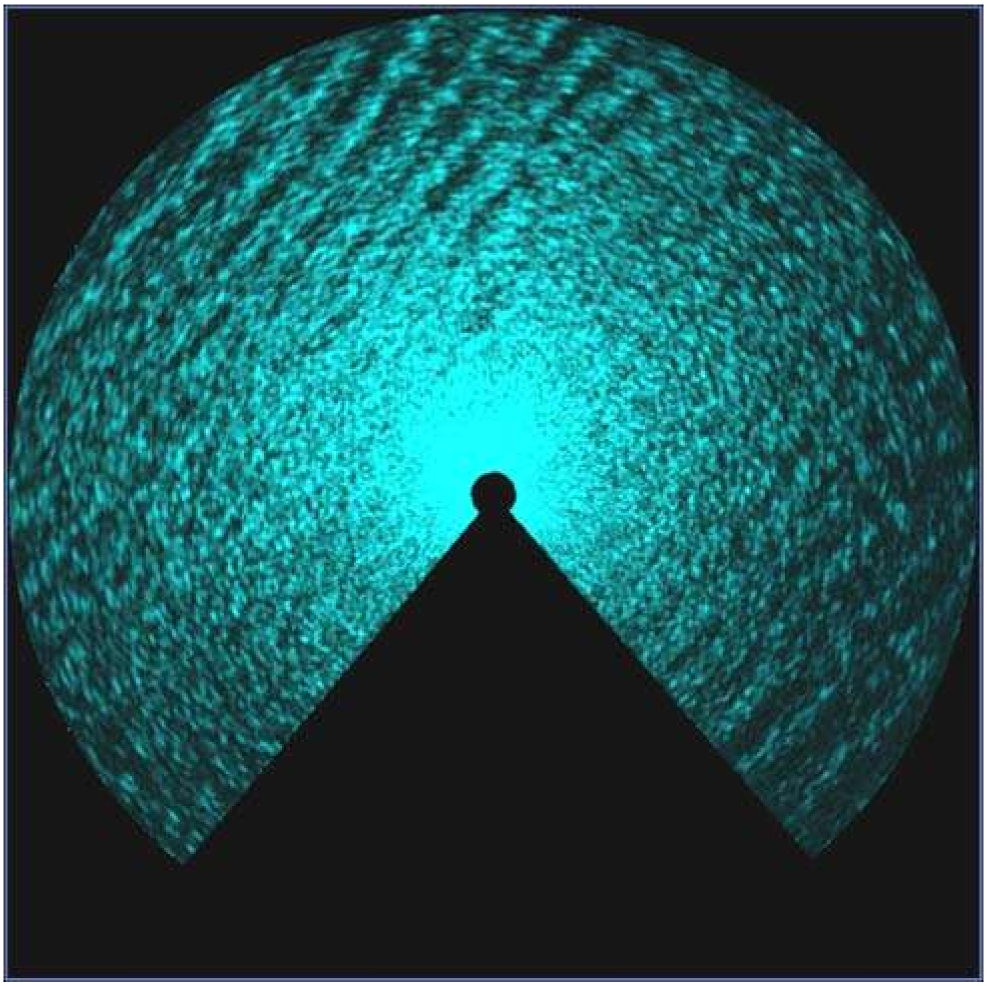
Laser altimeters

Laser altimeters are small and light weight and operate in the infra-red (IR) frequency band. They operate in vertical mode and normally use pulsed waveforms to perform direct measurements of sea surface elevation which easily can be converted to wave amplitude.

Microwave range finders

Microwave range finders also operate in vertical mode at GHz frequencies and is not as affected by fog and water spray as the laser altimeter. A continuous wave frequency modulated (CWFM) or pulsed radar waveform is normally used to provide range resolution. The beam is dispersive, hence the size of the footprint increases linearly with range.

One example of a microwave range finder is the Miros SM-094 which is designed for wave and water level (and tide) measurements. This sensor is applied as air gap (bridge clearance) sensor in NOAAs PORTS system. Another example is the Saab wave radar REX which is a derivative of a Saab tank radar.

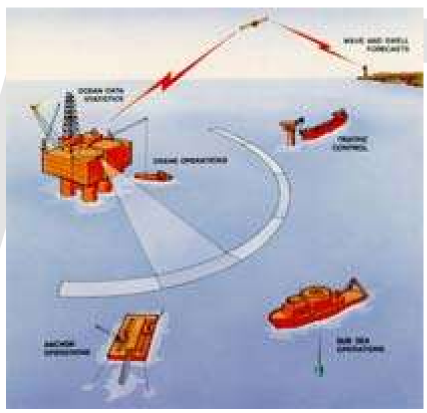


Digitized sea clutter image.

An array of three vertical radars in a triangular configuration can be used to measure a directional wave spectrum. Algorithms and signal processing software similar to what is used in the processing of heave, pitch, roll buoys. A commercial system called “Directional WaveGuide” is available from the Dutch companies Enraf and Radac.

Marine navigation radars

Marine navigation radars (X band) provide sea clutter images which contain a pattern resembling a sea wave pattern. By digitizing the radar video signal it can be processed by a digital computer. Sea surface parameters may be calculated on the basis of these digitized images. The marine navigation radar operates in low grazing angle mode and wind generated surface ripple must be present. The marine navigation radar is non-coherent and is a typical example of an indirect wave sensor, because there is no direct relation between wave height and radar back-scatter modulation amplitude. An empirical method of wave spectrum scaling is normally employed. Marine navigation radar based wave sensors are excellent tools for wave direction measurements. A marine navigation radar may also be a tool for surface current measurements. Point measurements of the current vector as well as current maps up to a distance of a few km can be provided (Gangeskar, 2002). Miros WAVEX has its main area of application as directional wave measurements from moving ships. Another example of a marine radar based system is OceanWaves WaMoS II.



Measurement geometry of pulsed Doppler wave and current radar.

The range gated pulsed Doppler microwave radar

The range gated pulsed Doppler microwave radar operates the in low grazing angle mode. By using several antennas it may be used as a directional wave sensor, basically measuring the directional spectrum of the horizontal water particle velocity. The velocity spectrum is directly related to the wave height spectrum by a mathematical model based on linear wave theory and accurate measurements of the wave spectrum can be provided under most conditions. As measurements are taken at a distance from the platform on which it is mounted, the wave field is to a small degree disturbed by interference from the platform structure.

Miros Wave and current radar is the only available wave sensor based on the range gated pulsed Doppler radar technique. This radar also uses the dual frequency technique to perform point measurements of the surface current vector

The dual frequency microwave radar

The dual frequency microwave radar transmits two microwave frequencies simultaneously. The frequency separation is chosen to give a “spatial beat” length which is in the range of the water waves of interest. The dual frequency radar may be considered a microwave equivalent of the high frequency (HF) radar. The dual frequency radar is suitable for the measurement of surface current. As far as wave measurements are concerned, the back-scatter processes are too complicated (and not well understood) to allow useful measurement accuracy to be attained.

The HF radar

The HF radar is well established as a powerful tool for sea current measurements up to a range of about 30 km. It operates in the MHz frequency band corresponding to a radar wavelength in the range of 10 to 300m. The Doppler shift of the first order Bragg lines of the radar echo is used to derive sea current estimates in very much the same way as for the dual frequency microwave radar. Two radar installations are normally required, looking at the same patch of the sea surface from different angles. CODAR Ocean Sensors (COS). The latest generation of shore-based ocean radar can reach more than 200 km for ocean current mapping and more than 100 km for wave measurements Helzel WERA. For all ocean radars, the accuracy in range is excellent. With shorter ranges, the range resolution gets finer. The angular resolution and accuracy depends on the used antenna array configuration and applied algorithms (direction finding or beam forming). The WERA system provides the option to use both techniques; the compact version with direction finding or the array type antenna system with beam forming methods.

Chapter 16

Wave Shoaling & Wind Wave Model

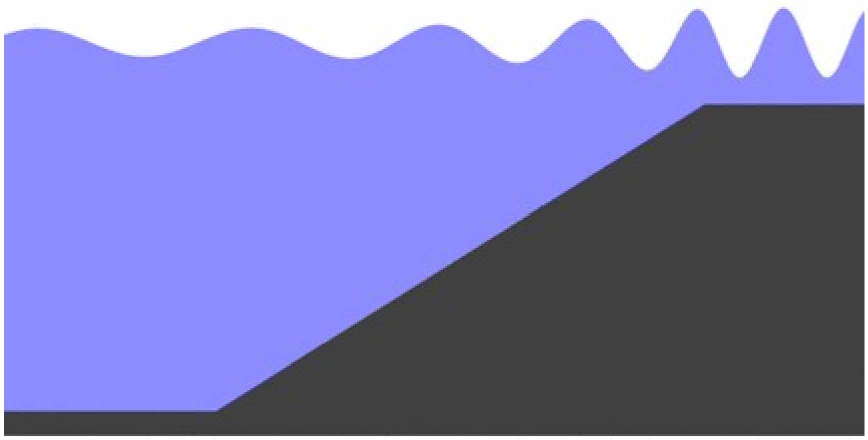
Wave Shoaling

Surfing on shoaling and breaking waves.

In fluid dynamics, **wave shoaling** is the effect by which surface waves entering shallower water increase in wave height (which is about twice the amplitude). It is caused by the fact that the group velocity, which is also the wave-energy transport velocity, decreases with the reduction of water depth. Under stationary conditions, this decrease in transport speed must be compensated by an increase in energy density in order to maintain a constant energy flux. Shoaling waves will also exhibit a reduction in wavelength while the frequency remains constant.

In shallow water and parallel depth contours, non-breaking waves will increase in wave height as the wave packet enters shallower water. This is particularly evident for tsunamis as they wax in height when approaching a coastline, with devastating results.

Mathematics



When waves enter shallow water they slow down. Under stationary conditions, the wave length is reduced. The energy flux must remain constant and the reduction in group (transport) speed is compensated by an increase in wave height (and thus wave energy density).

For non-breaking waves, the energy flux associated with the wave motion, which is the product of the wave energy density with the group velocity, between two wave rays is a conserved quantity (i.e. a constant when following the energy of a wave packet from one location to another). Under stationary conditions the total energy transport must be constant along the wave ray,

$$\frac{d}{ds}(c_g E) = 0,$$

where s is the co-ordinate along the wave ray and $c_g E$ is the energy flux per unit crest length. A decrease in group speed c_g must be compensated by an increase in energy density E. This can be formulated as a shoaling coefficient relative to the wave height in deep water.

Let us follow Phillips (1977) and Mei (1989) and denote the phase of a wave ray as

$$S = S(\mathbf{x}, t), 0 \leq S < 2\pi.$$

The local wave number vector is the gradient of the phase function,

$$\mathbf{k} = \nabla S,$$

and the angular frequency is proportional to its local rate of change,

$$\omega = -\partial S/\partial t.$$

Simplifying to one dimension and cross-differentiating it is now easily seen that the above definitions indicate simply that the rate of change of wavenumber is balanced by the convergence of the frequency along a ray;

$$\frac{\partial k}{\partial t} + \frac{\partial \omega}{\partial x} = 0.$$

Assuming stationary conditions ($\partial/\partial t = 0$), this implies that wave crests are conserved and the frequency must remain constant along a wave ray as $\partial \omega/\partial x = 0$. As waves enter shallower waters, the decrease in group velocity caused by the reduction in water depth leads to a reduction in wave length $\lambda = 2\pi / k$ because the nondispersive shallow water limit of the dispersion relation for the wave phase speed,

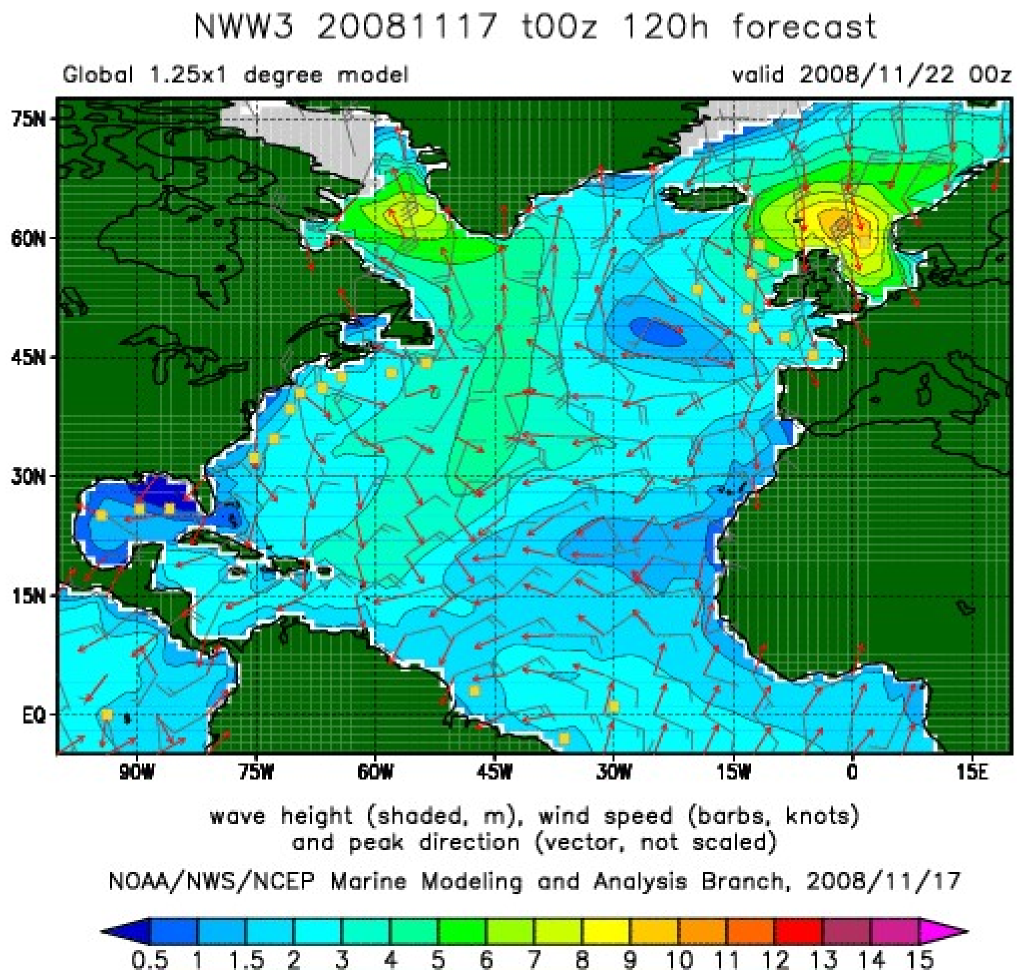
$$\omega/k \equiv c = \sqrt{gh}$$

dictates that

$$k = \omega/\sqrt{gh},$$

i.e, a steady increase in k (decrease in λ) as the phase speed decreases under constant ω.

Wind Wave Model



NOAA Wavewatch III 120-hour Forecast for the North Atlantic

In fluid dynamics, **wind wave modeling** describes the effort to depict the sea state and predict the evolution of the energy of wind waves using numerical techniques. These simulations consider atmospheric wind forcing, nonlinear wave interactions, and frictional dissipation, and they output statistics describing wave heights, periods, and propagation directions for regional seas or global oceans. Such **wave hindcasts** and **wave forecasts** are extremely important for commercial interests on the high seas. For example, the shipping industry requires guidance for operational planning and tactical seakeeping purposes.

For the specific case of predicting wind wave statistics on the ocean, the term **ocean surface wave model** is used.

Other applications, in particular coastal engineering, have led to the developments of wind wave models specifically designed for coastal applications.

Historical overview

Early forecasts of the sea state were created manually based upon empirical relationships between the present state of the sea, the expected wind conditions, the fetch/duration, and the direction of the wave propagation. Alternatively, the swell part of the state has been forecasted as early as 1920 using remote observations.

During the 1950s and 1960s, much of the theoretical groundwork necessary for numerical descriptions of wave evolution was laid. For forecasting purposes, it was realized that the random nature of the sea state was best described by a spectral decomposition in which the energy of the waves was attributed to as many wave trains as necessary, each with a specific direction and period. This approach allowed to make combined forecasts of wind seas and swells. The first numerical model based on the spectral decomposition of the sea state was operated in 1956 by the French Weather Service, and focused on the North Atlantic. The 1970s saw the first operational, hemispheric wave model: the spectral wave ocean model (SWOM) at the Fleet Numerical Oceanography Center.

First generation wave models did not consider nonlinear wave interactions. Second generation models, available by the early 1980s, parameterized these interactions. They included the “coupled hybrid” and “coupled discrete” formulations. Third generation models explicitly represent all the physics relevant for the development of the sea state in two dimensions. The wave modeling project (WAM), an international effort, led to the refinement of modern wave modeling techniques during the decade 1984-1994. Improvements included two-way coupling between wind and waves, assimilation of satellite wave data, and medium-range operational forecasting.

Wind wave models are used in the context of a forecasting or hindcasting system. Differences in model results arise, with decreasing order of importance, from differences in wind and sea ice forcing, differences in parameterizations of physical processes, the use of data assimilation and associated methods, the numerical techniques used to solve the wave energy evolution equation.

General strategy

Input

A wave model requires as initial conditions information describing the current state of the sea. An analysis of the sea or ocean can be created through data assimilation, where observations such as buoy or satellite altimeter measurements are combined with a background guess from a previous forecast or climatology to create the best estimate of the current conditions.

Representation

The sea state is described as a spectrum; the sea surface can be decomposed into waves of varying frequencies using the principle of superposition. The waves are also separated by their direction of propagation. The model domain size can range from regional to the global ocean. Smaller domains can be nested within a global domain to provide higher resolution in a region of interest. The sea state evolves according to physical equations, which include wave propagation / advection, and a source function which allows for wave energy to be augmented or diminished. The source function has at least three terms: wind forcing, nonlinear transfer, and dissipation by whitecapping. Wind data are typically provided from a separate atmospheric model from an operational weather forecasting center.

For intermediate water depths the effect of bottom friction should also be added . At ocean scales, the dissipation of swells - without breaking- is a very important term .

Output

The output of a wind wave model is a description of the wave spectra, with amplitudes associated with each frequency and propagation direction. Results are typically summarized by the significant wave height, which is the average height of the one-third largest waves, and the period and propagation direction of the dominant wave.

Coupled models

Wind waves also act to modify atmospheric properties through frictional drag of near-surface winds and heat fluxes. Two-way coupled models allow the wave activity to feed back upon the atmosphere.

Examples

WAVEWATCH

The operational wave forecasting systems at NOAA are based on the WAVEWATCH III model. This system has a global domain of approximately 100 km resolution, with nested regional domains for the northern hemisphere oceanic basins at approximately 25 km resolution. Physics includes wave field refraction, nonlinear resonant interactions, sub-grid representations of unresolved islands, and dynamically updated ice coverage. Wind data is provided from the GDAS data assimilation system for the GFS weather model. Up to 2008, the model was limited to regions outside the surf zone where the waves are not strongly impacted by shallow depths.

Other models

The European Centre for Medium-Range Weather Forecasts (ECMWF) has incorporated the WAM wave model as part of its ensemble forecasting system. This model includes 30 frequency bins, 24 propagation directions, and an average spatial resolution of 40 km.

Wind wave forecasts are issued regionally by Environment Canada.

Regional wave predictions are also produced by universities, such as Texas A&M University's use of the SWAN model to forecast waves in the Gulf of Mexico.

Other wind wave models include the U.S. Navy Standard Surf Model (NSSM).

The private sector is also active in producing wind wave simulations and surf forecasts. For example, Oceanweather Inc. provides global operational forecasts and hindcasts of the sea state.

Validation

Comparison of the wave model forecasts with observations is essential for characterizing model deficiencies and identifying areas for improvement. In-situ observations are obtained from buoys, ships and oil platforms. Altimetry data from satellites, such as GEOSAT and TOPEX, can also be used to infer the characteristics of wind waves.

Hindcasts of wave models during extreme conditions also serves as a useful test bed for the models.

Reanalyses

A retrospective analysis, or reanalysis, combines all available observations with a physical model to describe the state of a system over a time period of decades. Wind waves are a part of both the NCEP Reanalysis and the ERA-40 from the ECMWF. Such resources permit the creation of monthly wave climatologies, and can track the variation of wave activity on interannual and multi-decadal time scales. During the northern hemisphere winter, the most intense wave activity is located in the central North Pacific south of the Aleutians, and in the central North Atlantic south of Iceland. During the southern hemisphere winter, intense wave activity circumscribes the pole at around 50°S, with 5 m significant wave heights typical in the southern Indian Ocean.

Chapter 17

Mild-Slope Equation

In fluid dynamics, the **mild-slope equation** describes the combined effects of diffraction and refraction for water waves propagating over bathymetry and due to lateral boundaries — like breakwaters and coastlines. It is an approximate model, deriving its name from being originally developed for wave propagation over mild slopes of the sea floor. The mild-slope equation is often used in coastal engineering to compute the wave-field changes near harbours and coasts.

The mild-slope equation models the propagation and transformation of water waves, as they travel through waters of varying depth and interact with lateral boundaries such as cliffs, beaches, seawalls and breakwaters. As a result, it describes the variations in wave amplitude, or equivalently wave height. From the wave amplitude, the amplitude of the flow velocity oscillations underneath the water surface can also be computed. These quantities — wave amplitude and flow-velocity amplitude — may subsequently be used to determine the wave effects on coastal and offshore structures, ships and other floating objects, sediment transport and resulting geomorphology changes of the sea bed and coastline, mean flow fields and mass transfer of dissolved and floating materials. Most often, the mild-slope equation is solved by computer using methods from numerical analysis.

A first form of the mild-slope equation was developed by Eckart in 1952, and an improved version — the mild-slope equation in its classical formulation — has been derived independently by Juri Berkhoff in 1972. There after, many modified and extended forms have been proposed, to include the effects of, for instance: wave–current interaction, wave nonlinearity, steeper sea-bed slopes, bed friction and wave breaking. Also parabolic approximations to the mild-slope equation are often used, in order to reduce the computational cost.

In case of a constant depth, the mild-slope equation reduces to the Helmholtz equation for wave diffraction.

Formulation for monochromatic wave motion

For monochromatic waves according to linear theory — with the free surface elevation given as $\zeta(x,y,t) = \Re\left\{\eta(x,y)\,\mathrm{e}^{-i\omega t}\right\}$ and the waves propagating on a fluid layer of mean water depth *h*(*x*,*y*) — the mild-slope equation is:

$$\nabla\cdot\left(c_p\,c_g\,\nabla\eta\right)\,+\,k^2\,c_p\,c_g\,\eta\,=\,0,$$

where:

- η(*x*,*y*) is the complex-valued amplitude of the free-surface elevation ζ(*x*,*y*,*t*);
- (*x*,*y*) is the horizontal position;
- ω is the angular frequency of the monochromatic wave motion;
- *i* is the imaginary unit;
- $\Re\{\cdot\}$ means taking the real part of the quantity between braces;
- ∇ is the horizontal gradient operator;
- $\nabla\cdot$ is the divergence operator;
- *k* is the wavenumber;
- c_p is the phase speed of the waves and
- c_g is the group speed of the waves.

The phase and group speed depend on the dispersion relation, and are derived from Airy wave theory as:

$$\begin{aligned}\omega^2 &= g\,k\,\tanh\,(kh),\\ c_p &= \frac{\omega}{k}\quad\text{and}\\ c_g &= \frac{1}{2}\,c_p\,\left[1\,+\,kh\,\frac{1-\tanh^2(kh)}{\tanh\,(kh)}\right]\end{aligned}$$

where

- *g* is Earth's gravity and
- tanh is the hyperbolic tangent.

For a given angular frequency ω, the wavenumber *k* has to be solved from the dispersion equation, which relates these two quantities to the water depth *h*.

Transformation to an inhomogeneous Helmholtz equation

Through the transformation

$$\psi = \eta \sqrt{c_p\, c_g},$$

the mild slope equation can be cast in the form of an inhomogeneous Helmholtz equation:

$$\Delta\psi + k_c^2\, \psi = 0 \qquad \text{with} \qquad k_c^2 = k^2 - \frac{\Delta\left(\sqrt{c_p\, c_g}\right)}{\sqrt{c_p\, c_g}},$$

where Δ is the Laplace operator.

Propagating waves

In spatially coherent fields of propagating waves, it is useful to split the complex amplitude η(*x*,*y*) in its amplitude and phase, both real valued:

$$\eta(x,y) = a(x,y)\, \mathrm{e}^{i\,\theta(x,y)},$$

where

- $a = |\eta|$ is the amplitude or absolute value of η and
- $\theta = \arg\{\eta\}$ is the wave phase, which is the argument of η.

This transforms the mild-slope equation in the following set of equations (apart from locations for which $\nabla\theta$ is singular):

$$\frac{\partial \kappa_y}{\partial x} - \frac{\partial \kappa_x}{\partial y} = 0 \qquad \text{with } \kappa_x = \frac{\partial\theta}{\partial x} \text{ and } \kappa_y = \frac{\partial\theta}{\partial y},$$

$$\kappa^2 = k^2 + \frac{\nabla\cdot\left(c_p\, c_g\, \nabla a\right)}{c_p\, c_g\, a} \qquad \text{with } \kappa = \sqrt{\kappa_x^2 + \kappa_y^2} \quad \text{and}$$

$$\nabla\cdot\left(\boldsymbol{v}_g\, E\right) = 0 \qquad \text{with } E = \frac{1}{2}\rho\, g\, a^2 \quad \text{and} \quad \boldsymbol{v}_g = c_g\, \frac{\boldsymbol{\kappa}}{k},$$

where

- *E* is the average wave-energy density per unit horizontal area (the sum of the kinetic and potential energy densities),
- $\boldsymbol{\kappa}$ is the effective wavenumber vector, with components (κ_x, κ_y),
- $\boldsymbol{v}_g$ is the effective group velocity vector,
- ρ is the fluid density, and
- *g* is the acceleration by the Earth's gravity.

The last equation shows that wave energy is conserved in the mild-slope equation, and that the wave energy *E* is transported in the $\boldsymbol{\kappa}$-direction normal to the wave crests (in this case of pure wave motion without mean currents). The effective group speed $|\boldsymbol{v}_g|$ is different from the group speed c_g.

The first equation states that the effective wavenumber $\boldsymbol{\kappa}$ is irrotational, a direct consequence of the fact it is the derivative of the wave phase θ, a scalar field. The second equation is the eikonal equation. It shows the effects of diffraction on the effective wavenumber: only for more-or-less progressive waves, with $\nabla \cdot (c_p\, c_g\, \nabla a) \ll k^2\, c_p\, c_g\, a,$ the splitting into amplitude *a* and phase θ leads to consistent-varying and meaningful fields of *a* and $\boldsymbol{\kappa}$. Otherwise, κ^2 can even become negative. When the diffraction effects are totally neglected, the effective wavenumber κ is equal to *k*, and the geometric optics approximation for wave refraction can be used.

Details of the derivation of the above equations

Derivation of the mild-slope equation

The mild-slope equation can be derived by the use of several methods. Here, we will use a variational approach. The fluid is assumed to be inviscid and incompressible, and the flow is assumed to be irrotational. These assumptions are valid ones for surface gravity waves, since the effects of vorticity and viscosity are only significant in the Stokes boundary layers (for the oscillatory part of the flow). Because the flow is irrotational, the wave motion can be described using potential flow theory.

Details of the derivation of the mild-slope equation

The following time-dependent equations give the evolution of the free-surface elevation $\zeta(x,y,t)$ and free-surface potential $\varphi(x,y,t)$:

$$
\begin{aligned}
g\,\frac{\partial \zeta}{\partial t} + \nabla \cdot (c_p\, c_g\, \nabla \varphi) + \left(k^2\, c_p\, c_g\, - \omega_0^2\right)\, \varphi = 0, \\
\frac{\partial \varphi}{\partial t} + g\zeta = 0, \quad \text{with} \quad \omega_0^2\, =\, g\, k\, \tanh(kh).
\end{aligned}
$$

From the two evolution equations, one of the variables φ or ζ can be eliminated, to obtain the time-dependent form of the mild-slope equation:

$$
-\frac{\partial^2 \zeta}{\partial t^2} + \nabla \cdot (c_p\, c_g\, \nabla \zeta) + \left(k^2\, c_p\, c_g\, - \omega_0^2\right)\, \zeta = 0,
$$

and the corresponding equation for the free-surface potential is identical, with ζ replaced by φ. The time-dependent mild-slope equation can be used to model waves in a narrow band of frequencies around ω_0.

Monochromatic waves

Consider monochromatic waves with complex amplitude $\eta(x,y)$ and angular frequency ω:

$$\zeta(x,y,t) = \Re\left\{\eta(x,y)\,\mathrm{e}^{-i\omega t}\right\},$$

with ω and ω_0 chosen equal to each other, $\omega = \omega_0$. Using this in the time-dependent form of the mild-slope equation, recovers the classical mild-slope equation for time-harmonic wave motion:

$$\nabla \cdot (c_p\, c_g\, \nabla \eta) + k^2\, c_p\, c_g\, \eta = 0.$$

Applicability and validity of the mild-slope equation

The standard mild slope equation, without extra terms for bottom curvature, provides accurate results for the wave field over bottom slopes ranging from 0 to about 1/3. However, some subtle aspects, like the amplitude of reflected waves, can be completely wrong, even for slopes going to zero. This mathematical curiosity has little practical importance in general since this reflection becomes vanishingly small for bottom slopes.