Waves in Classical Physics

If I say the word "wave" in no particular context, the image that most probably springs to your mind is one of a roughly periodic undulation on the surface of a pond or, perhaps, the sea. But the concept of a wave in classical physics is actually much more general, and in particular the periodicity is not really an essential aspect. (Huygens, Fourier...)

At least in classical (i.e., pre-relativistic as well as pre-quantum) physics, a wave always occurs *in* (or *on*) some medium. Examples include surface waves on water, transverse waves on a plucked string, and sound waves (pressure, or density waves in air). We will assume that there exists a state where the medium is undisturbed (string stretched straight and taut, water flat, air still...), and until further notice, we will *choose an (inertial) frame of reference in which the undisturbed medium is at rest.* In general, we can describe a *disturbance* in the medium by specifying some quantity (call it generically q) that tells us how far the medium is displaced from its undisturbed (equilibrium) state *at a given point, at a given time.* (By contrast, in the case of a particle we ask: what is its position at a given time? I.e., we write $x = x(t)$.) Thus, for a stretched string, q would be the transverse displacement from the original straight configuration, and would be a function of one space variable, the distance x along the string, and of time:

$$q = q(x, t)$$

For the water surface, q would be the height above the equilibrium, and would be a function of the two variables x and y describing horizontal position and of t; for sound waves in air, q would be the excess pressure above equilibrium and would be a function of the three space coordinates x, y, z, and of time; and so on. Two points that are absolutely crucial to appreciate are (1) in all cases, the "disturbance" q *can be positive or negative* – crudely speaking, waves can have troughs as well as crests, and (2) q is always defined as a function of position as well as of time: in general it makes no sense to ask for "*the* position" of the disturbance (though it may make limited sense, see below). Any disturbance of this type is generically called a *wave*, and the quantity $q(x, y, z, t)$ is called its *amplitude*.

Suppose we take a "snapshot" of the disturbance q at some definite time t, as a function of position. If we take a second snapshot at a slightly later time and compare the two, $q(x)$ will in general look different; the disturbance *propagates*, i.e., changes its space-dependence as time evolves. Exactly how this happens depends on the details of the system studied; generally speaking, the system will obey some equation that relates the changes of q with time near some point (x, y, z) to the value of q and its space derivatives near (x, y, z). All the cases we shall be interested in have the property that the equations obeyed by $q(x, y, z, t)$ are *linear* – that is, crudely speaking, changes of q in time are directly proportional to changes of q in space (and not, for example to changes of q^2); as we shall see, this property has a very important consequence.

To see qualitatively how such an equation might arise, let us look briefly at a specific example, namely, transverse waves on a stretched string, such as on a violin. In the equilibrium (undisturbed) state the string is straight and taut. Consider now what happens if at a given time t the string is displaced so that its form is something like Fig. 1.

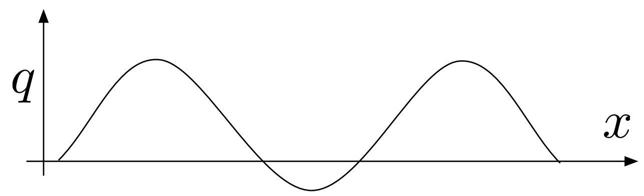


Figure 1

What determines how the string will behave in the next few instants? Actually, nothing but Newton's second law: acceleration = force/mass! Consider a small element of length* Δx and, therefore, mass $m = \rho\Delta x$ (ρ is the mass per unit length).

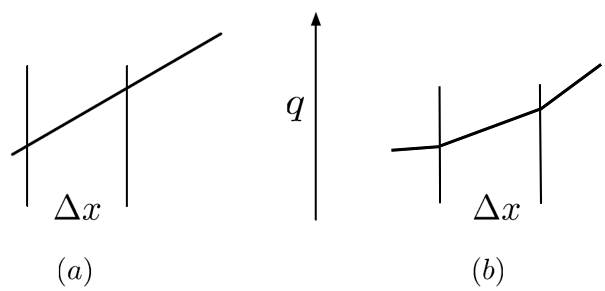


Figure 2

The forces acting on this element arise from the tension of the parts of the string immediately to its left and right; the actual magnitude of the tension is (approximately) that of the original unstretched string, but its direction may be different. In the special case (a), where the string happens to be straight at the point in question even though it is not parallel to the original unstretched one, it is clear that the forces exerted by the right and left "neighbors" are equal and opposite, so there is no net force and no acceleration. However, in case (b), there is a net resultant force, which is proportional to the difference in slope,† $\Delta(\Delta q/\Delta x)$; in fact, numerically, it is simply equal to $Y\Delta(\Delta q/\Delta x)$, where Y is the original tension in the unstretched string. But the quantity $\Delta(\Delta q/\Delta x)$ is approximately equal to Δx times the "rate of change of slope" (d^2q/dx^2 in the standard notation of differential calculus). Equating $Y\Delta(\Delta q/\Delta x)$ to the mass $\rho\Delta x$ times the (transverse) acceleration of this bit of the string, we find that the length Δx of the element in question cancels between the two sides and we get the result (for the transverse

*Note that the term Δx is being used in a slightly different sense than in lectures 6-8.

†If you have difficulty with the double deltas, it is recommended that you reread the discussion of the definition of "acceleration" in lecture 6.

acceleration):‡

$$\boxed{\text{acceleration} = (Y/\rho) \times \text{ rate of change of slope}} \qquad (\alpha)$$

Or, in the formal language of differential calculus:

$$\frac{d^2q(x,t)}{dt^2} = c^2\frac{d^2q(x,t)}{dx^2}, \quad c^2 \equiv Y/\rho \quad (c \text{ has dimensions of velocity})$$

This equation has the standard form of a *nondispersive* linear wave equation (we will see the reason for the name in a minute); a similar equation is satisfied, e.g., by sound waves in air or (in an appropriate limit only) waves on a water surface. However, not all waves satisfy a simple equation of this type; for example, the right-hand side may contain more complicated quantities related to the space variation. In this case, the wave equation is called "dispersive"; one example is water waves under more general conditions.

Whatever the detailed form of the equation obeyed by the wave (as long as it is linear), it always has, among its other solutions, the familiar "monochromatic wave"; that is, a disturbance that is wavelike in the intuitive sense, corresponding to a series of equally spaced crests and troughs that move to left or right with constant velocity.

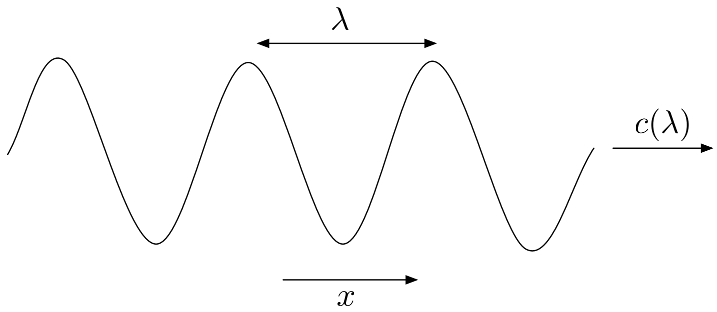


Figure 3

In such a case, a "snapshot" of the disturbance taken at a fixed time will produce the picture in Fig. 3; the (constant) separation between the wave crests is the *wavelength* and is conventionally denoted λ. On the other hand, if we stand at a fixed point in space and watch the surface of the water (or more generally the quantity $q(x,t)$) as a function of time, we find it oscillates regularly, with some *period* (interval between times of maximum height) conventionally denoted $T(\lambda)$. Now if I am standing at a given point and wait the period T between one crest and the next, it is clear that the last crest has moved on exactly a distance λ. Thus, the speed with which the pattern of crests travels – call it c – is related to the wavelength λ and the period $T(\lambda)$ for that λ by the simple relation

$$c(\lambda) = \lambda/T(\lambda)$$

‡It turns out (not obvious!) that, for small displacements, the longitudinal motion of the element Δx (i.e., motion in the x-direction) can be neglected.

In the general case, the velocity is a function of the wavelength λ, as the notation indicates. Note that all of this is completely independent of the overall "amplitude" of the wave (i.e., the height of the crests).[§]

In the special case of a nondispersive equation of the type of eqn. α, the situation is particularly simple. It is fairly clear that the "rate of change of slope" in space is proportional to the amplitude A and inversely proportional to the square of the wavelength λ. Similarly, the acceleration is proportional to A and inversely proportional to the square of T. So, in eqn. α, the factors of A cancel and we get

$$(\text{const.})T^{-2} \propto c^2(\text{const.})\lambda^{-2} \qquad (\beta)$$

where c is the *constant* quantity $\sqrt{Y/\rho}$. Thus, $T \propto \lambda$. It is straightforward to show that the other constants on the two sides of eqn. β are the same, so in fact we find the simple result:

$$T(\lambda) = \lambda/c \qquad (\text{or } \nu\lambda = c, \text{ where the "frequency" } \nu \text{ is just } 1/T)$$

In this case, the "wave velocity" $c(\lambda)$ is independent of λ and just given by the $c = \sqrt{Y/\rho}$ that appears in eqn. α. Thus, in this case, all simple "monochromatic" waves travel at the same speed, regardless of wavelength. This is no longer true for more complicated (dispersive) wave equations.

We now ask the following question: considering, for definitiveness, the case of a stretched string, suppose we pluck it at time zero in such a way that $q(x, 0)$ has a definite shape; say, for example, the one in Fig. 4 (displacement much exaggerated!) and let it go.

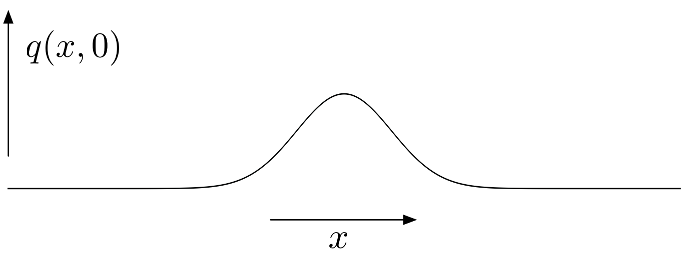


Figure 4

What does the disturbance (amplitude) $q(x, t)$ look like at a later time t? (In the following, I neglect for simplicity complications associated with the finite length of the string, which turns out to be a valid approximation if its length L is much greater than ct.)

The answer is given by a remarkable theorem of pure mathematics (Fourier's theorem): given any disturbance $q(x)$ (or more precisely, any disturbance of any type remotely likely to occur in practice), we can always regard it as "built up" of a set of (in

[§]In the case of a simple "monochromatic" wave of the type discussed, the term "amplitude" is usually used to refer to the height of the crests.

general, an infinite number of) monochromatic waves of different wavelengths λ. At first sight, this theorem may look implausible, because any *one* monochromatic wave will exist everywhere in space, whereas, e.g., the $q(x)$ in Fig. 4 corresponds to zero disturbance anywhere except the plucked region; nevertheless, it is true! What we do, therefore, is "analyze" (break up) the disturbance in question into a set of monochromatic waves, calculate the progress of the latter up to time t according to the prescription that a wave of wavelength λ moves with velocity $c(\lambda)$, and then reassemble at time t the different waves to give the total disturbance $q(x, t)$. In principle, this procedure gives a complete solution of the problem.

The case of nondispersive wave equation is particularly simple, because once one has isolated the set of waves that are moving (say) rightward, they all travel at exactly the same velocity c, and consequently the "disturbance" they form itself travels rightward at this same velocity and *without change of shape* (i.e., without "dispersion" – hence the name¶). Thus, for example, if we release the string as plucked in the above diagram from rest, what happens is that it splits into two disturbances of equal magnitude, which travel right and left, respectively, at constant velocity c:

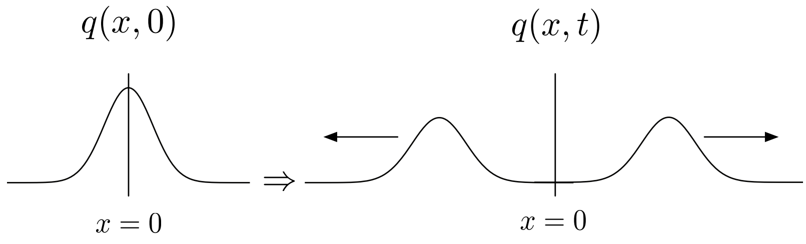


Figure 5

The two "compact" disturbances are known as "wave packets".

In the case of a dispersive wave, the different component waves (even the rightward ones) travel with different velocities, and hence, in general, the shape of the disturbance changes in time and, in particular, it (usually) tends to get wider ("dispersion"). However, the concept of a "wave packet" – that is, a reasonably compact or localized disturbance – can still remain valid under certain conditions (very crudely speaking, the initial width of the packet should be large compared to the characteristic wavelength of the waves that principally compose it, and the wait time should not be too great). Thus, although we emphasized above that, in general, it makes no sense to ask for the "position" of a wave-type disturbance, it makes some sort of approximate sense in the case of a wave packet – even in the dispersive case, provided the packet stays together fairly well. We will see that this feature is crucial in the quantum-mechanical picture of microscopic objects.

¶Actually, historically the term "dispersion" comes from optics: light of different wavelengths travels in glass with different velocities, and as a result, a prism "disperses" (separates) the different colors (wavelengths).

At this stage, we make a short digression to ask about the energy associated with a wave. Actually, this is quite a complicated question, and the exact expression depends in detail on the nature of the wave-type phenomenon considered. However, one can make two general statements.

(1) In the absence of dissipation, the *total* energy associated with a disturbance is conserved, though it may be "relocated" in time between different spatial regions.

(2) Crudely speaking, under most conditions, the energy associated, at any given time t, with a particular region of space in the neighborhood of x, will be proportional to the average of the *square* of the amplitude $q(x,t)$ in this region:

$$E(x,t) \propto \overline{q^2(x,t)}$$

where the notation $\overline{A(x,t)}$ indicates that the quantity A is averaged over a small region of space surrounding x (note that there is no average over time t). In particular, in the motion of a wave packet, as discussed above, the energy is relatively well-localized in the neighborhood of the packet (as we should no doubt expect intuitively!): regions where $q(x,t)$ is small have little or no energy associated with them.

In the above argument concerning the development of the disturbance $q(x,t)$, we made implicit use of a very fundamental principle that holds for any system obeying a *linear* wave equation (whether or not it is dispersive), namely the *principle of superposition*: given any two possible solutions, their sum is also a possible solution. (It was this principle that we used, implicitly, when we assumed that the propagation of the whole disturbance could be obtained by computing the propagation of the component monochromatic waves and reassembling them.) The principle of superposition has a number of fundamental and remarkable consequences, in particular the phenomenon of *interference*. To illustrate this, let us consider the following (symmetrical) arrangement of stretched strings.

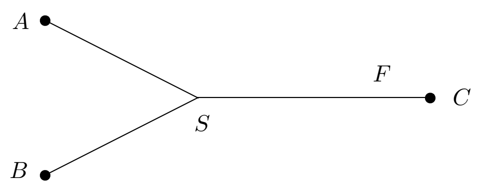


Figure 6

(Note: the only pegs in Fig. 6 are at the points A, B, and C.) Suppose we generate some kind of disturbance (let us say for definiteness one of the "packets" illustrated above) at A. It will propagate down to S, and at that point, we will find in general that part of it will be transmitted down the string F toward C, and the rest either transmitted toward B or reflected toward A. Since $E \propto q^2$, a finite amount of energy, call it E_0, will be transmitted down F. By symmetry, if we generate the same disturbance at B, the same will happen but with the roles of A and B interchanged; again part of the wave will be transmitted down F, and the energy transmitted down it will again be E_0. Note that in each case, we could have generated a "negative" pulse at A (or B); then, of course, the pulse traveling down F would simply have the sign of $q(x,t)$ reversed, but since E is proportional not to q but to q^2, the transmitted energy is still E_0.

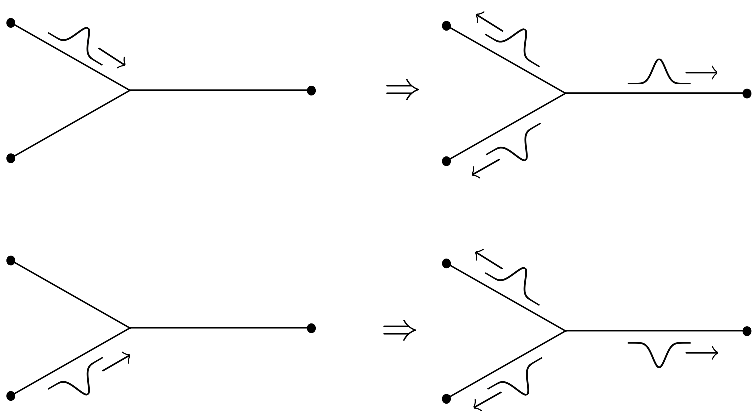
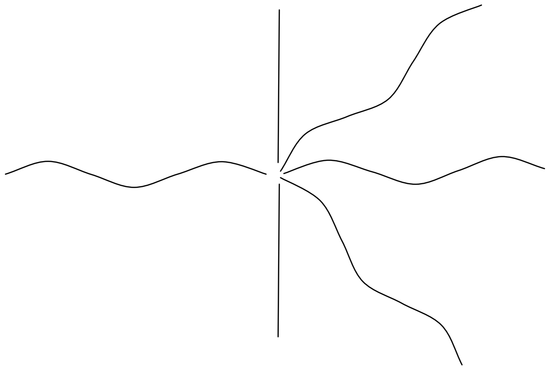


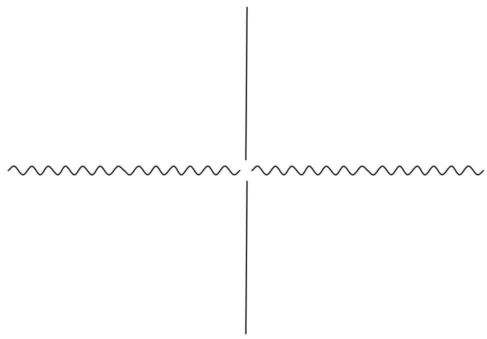
Figure 7

Now let us ask: What happens if we *simultaneously* generate wave packets at A and B? By the principle of superposition, the packet propagating down F must be just the algebraic sum of those generated separately by the disturbances traveling from A and from B. If the packets at A and B are produced "in phase" – i.e., with the same sign – then we find that a packet of twice the height propagates down F. Since $E \propto q^2$, this means that the energy transmitted down F is *four times*, not twice, E_0; i.e., it is *more* than the sum of the energies transmitted when only A was oscillating and when only B was! This phenomena is called "constructive interference." The opposite case, when the packet at B is "out of phase" with that at A (opposite sign) is even more spectacular: in this case the packets traveling down F exactly cancel one another, so $q(x,t)$ is zero on F and no energy is transmitted down it ("destructive interference")! (Since I said earlier that the total energy of a wave is conserved, you might well wonder where the "missing" energy has gone. The answer is that things work out automatically so that destructive interference in F is compensated by constructive interference in A or B, in just such a way that the total energy is indeed conserved.) Can all the energy go down F?

A well-known special case of the phenomenon of interference is the *diffraction* of a wave by an aperture or a solid object – i.e., its deviation from propagation approximately in a straight line (see Fig. 8). The condition for diffraction effects to be appreciable is that the dimensions of the diffracting object should be not much larger than the wavelength of the relevant wave. Thus, for example, a doorway will diffract sound (wavelength ~ 1 m) but not, to any observable degree, visible light (wavelength $\sim 10^{-7}$ m). To diffract visible light efficiently, one needs to use objects such as very fine wire meshes with dimensions of the order of $\sim 10^{-5}$ m or less.



(a) Wavelength much greater than size of aperture: diffraction.



(b) Wavelength much smaller than size of aperture: no diffraction.

Figure 8

The "Young's slits" experiment

Newton, and most of his immediate successors, had imagined light to be a stream of particles, while Huygens and others argued that it was a wave phenomenon. The issue was decisively settled (at least for a time!) by a series of experiments in the early 19th century that showed that light to have the characteristically "wavelike" properties of interference and diffraction. Of these experiments, probably the most spectacular (and the one that is traditionally used, and that we will use later, to introduce some of the basic concepts of quantum mechanics) is the one first performed by Thomas Young in 1802, and usually known as "Young's slits".

The experiment involves a source of light (of a special kind – a simple light bulb will not do; see below), an opaque screen with two slits S_1 and S_2 cut in it, and a final screen that can either be viewed with the naked eye (as Young must have done) or covered with a photographic emulsion. If one of the intermediate slits, say S_2, is blocked, so that light can reach the final screen only through S_1, the pattern seen on this screen is a diffuse blur without any spectacular properties; similarly if S_1 is blocked and only S_2 open. However, if both slits are opened simultaneously, the final screen exhibits a striking pattern of alternating bright and dark lines – a so-called "interference pattern". It seems very difficult to explain this behavior if light is a stream of particles (we shall see just how difficult when we get to quantum mechanics!), but it has an immediate and natural and explanation if light is a wave, as follows.

Let us suppose that the slits S_1 and S_2 are symmetrically positioned relative to the source; then whenever there is a crest of the wave at S_1, there will be one at S_2, and similarly a trough at S_1 will be accompanied by one at S_2 (the technical phrase is that the waves at S_1 and S_2 are "in phase"). Now consider a point X on the final screen. If X happens to be positioned symmetrically with respect to the two slits S_1 and S_2, then the crests of the waves coming from these two slits will arrive simultaneously and will reinforce one another. On the other hand, suppose X is not symmetrically placed, as in Fig. 9, so that the distance to S_1 and S_2 is different. Then the time taken for the wave

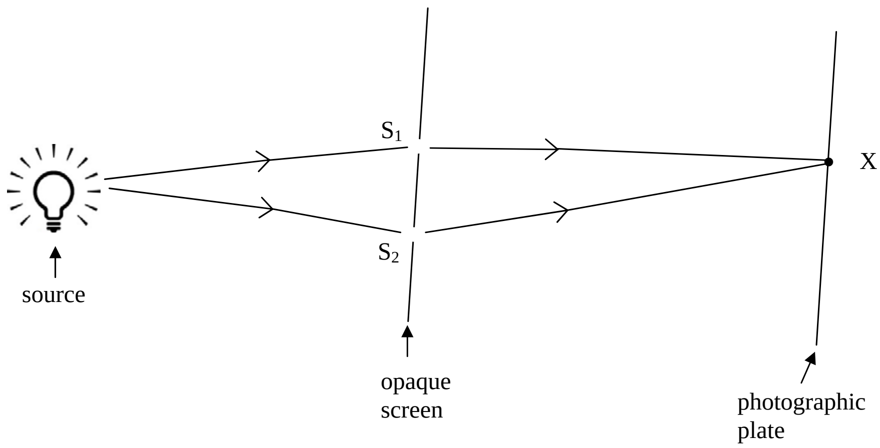


Figure 9

to propagate to X from S_1 is different from that taken from S_2, and in general the crests of the two waves will not arrive simultaneously; in fact, we may even find that a crest for S_1 arrives simultaneously with a trough from S_2 and vice versa. In that case, the two waves will exactly cancel, leaving no disturbance at all at X ("destructive interference"). In fact, the amount of "disturbance" (i.e., presumably, of light) received at X should depend on the difference (call it $\Delta s(X)$) in the path lengths $S_1 - X$ and $S_2 - X$, being a maximum at points X where $\Delta s(X)$ is an integral number of wavelengths λ of the light ($\Delta s(X) = n\lambda, n = 0, \pm 1, \pm 2...$) and zero when $\Delta s(X) = (n + 1/2)\lambda$. Since $\Delta s(X)$ varies continuously with X, we expect a pattern of bright and dark bands, as observed; in fact, with the help of a little elementary geometry, we can work out, from the observed position of the bands, the wavelength λ (which turns out to be of the order of 5×10^{-7}

meters for ordinary visible light).

Two things are worth noticing about the Young's slits experiment: first, the fact that we can get light from a single source arriving at X through both S_1 and S_2 itself depends on the phenomenon of (single-slit) diffraction (cf. Figs. 8a and b, above). Second, for it to work, λ must be pretty much uniquely defined for the source in question ("monochromatic" light, such as is provided, to a good approximation, by a sodium arc lamp): it would be difficult or impossible to see the effect with a standard light bulb, since the emitted light in this case has a wide spread in its values of λ, and so the condition $\Delta s(X) = n\lambda$ is met at different X for different components of the light.

Behavior of waves under Galilean transformation

Consider, e.g., a (nondispersive) wave on a string, or more realistically a sound (air pressure) wave, that is exclusively right-moving. Note that all our calculations of wave behavior have been implicitly done in a frame of reference in which the *eqm. medium is stationary* (e.g., in the case of a sound wave, the frame with respect to which the air is at rest). In such a frame, the general form of the disturbance is:

$$q(x,t) = f(x - ct)$$

(This is just the formal expression of the statement: the wave moves rightward with velocity c without change of shape.) How does it look from a frame S' moving with velocity v in the rightward direction? (If v is negative relative to c, we are moving to the left.)

Under GT we have:

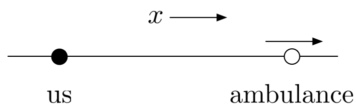
$$x' = x - vt \quad [t' = t] \quad \text{or} \;\; x = x' + vt$$

Hence:

$$q(x', t') \equiv q(x', t) = f(x - ct) = f(x' + vt - ct) \equiv f(x' - (c - v)t).$$

Thus, from the moving frame S', the wave will appear to be traveling with velocity $c - v$ (just as would a mechanical object). (This looks like common sense!) If $v > c$, then the wave will appear to be moving *backward* (just as does a bicycle overtaken by a car or a slower train overtaken by a faster one). By a similar argument, a left-traveling wave will appear to move with velocity $c + v$. Thus, we can ascertain the rest frame of the medium from the consideration that it is the unique frame in which (e.g.) sound moves with equal velocity in both directions (or in 3D, in "all" directions). Note then that this is *unlike* the case of particle mechanics: a definite "privileged" frame *does* exist, namely that in which the medium carrying the wave is at rest.

Doppler effect



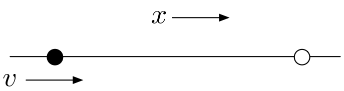
Imagine, e.g., an ambulance moving away from us with velocity v (suppose for definiteness to the right, i.e., toward positive x). Its siren emits sound waves (velocity c with respect to the air) with frequency ν_0. The extra distance that the ambulance has traveled away from us between wave "crests" is $vT_0 = v/\nu_0$, and hence the extra time taken by the crest to reach us is $(v/c)\nu_0^{-1}$. Consequently, the interval between "heard" crests is:

$$T = T_0 + \frac{v}{c} \cdot \frac{1}{\nu_0}, \quad \text{or}$$

$$\frac{1}{\nu} = \frac{1}{\nu_0} + \frac{v}{c} \cdot \frac{1}{\nu_0} \quad \Rightarrow \quad \nu = \frac{\nu_0}{1 + v/c}$$

Note this argument applies for either sign of v: i.e., if $|v|$ denotes the actual speed (irrespective of direction), then the tone of the ambulance moving towards us is $\frac{\nu_0}{1-|v|/c}$, while when it moves away it is $\frac{\nu_0}{1+|v|/c}$. Thus, the pitch drops as it passes us ("Doppler effect").

The same argument should prima facie apply to all kinds of waves, provided that v is measured relative to the ("stationary") medium of propagation. Note that a similar argument applies if the ambulance is at rest and we are moving with respect to the medium (to the right, i.e., v is positive):



In this case, we argue as follows: if we were stationary, then in a time Δt, the number of crests passing us would be $\nu_0 \Delta t$. But in fact, in this time, we have moved a distance $v\Delta t$ in the direction of the ambulance, and therefore passed an extra number of crests equal to $v\Delta t/\lambda = (v/c)\nu_0\Delta t$. Hence the total number of crests is $\nu_0(1 + v/c)\Delta t$, so that ν (the frequency we hear) is related to ν_0 (that emitted) by

$$\nu = \nu_0(1 + v/c)$$

However, if both we and the ambulance are moving at *equal* velocity (in the same sense) with respect to the medium, there is clearly no shift. Since we will need it later in the context of special relativity, I quote for reference the general formula for the case when the source is moving with respect to the medium M with velocity v_s and the receiver is also moving with respect to M with velocity v_r:

$$\nu = \nu_0 \frac{(1 + v_r/c)}{(1 + v_s/c)}$$

This formula can be derived from the above ones by imagining that a fictitious observer who is stationary with respect to M receives and re-emits the sound. Evidently, there is no shift for $v_s = v_r$, as stated.

Finally, what is really fundamental about a wave-type phenomenon?

1. It is a disturbance that is "spread out" in space at any given time: it is meaningless, in general, to ask for the exact "position" of a wave at a given time (although, as we have seen, under certain circumstances, it may be localized in a general region of space).
2. It is represented by an *amplitude* that can be positive or negative, and thus can show the phenomena of *superposition* and *interference*.