What Is a Soliton?

by Peter S. Lomdahl

About thirty years ago a remarkable discovery was made here in Los Alamos. Enrico Fermi, John Pasta, and Stan Ulam were calculating the flow of energy in a one-dimensional lattice consisting of equal masses connected by nonlinear springs. They conjectured that energy initially put into a long-wavelength mode of the system would eventually be "thermalized," that is, be shared among all modes of the system. This conjecture was based on the expectation that the nonlinearities in the system would transfer energy into higher harmonic modes. Much to their surprise the system did not thermalize but rather exhibited energy sharing among the few lowest modes and long-time near recurrences of the initial state.

This discovery remained largely a mystery until Norman Zabusky and Martin Kruskal started to investigate the system again in the early sixties. The fact that only the lowest order (long-wavelength) modes of the discrete Fermi-Pasta-Ulam lattice were "active" led them in a continuum approximation to the study of the nonlinear partial differential equation

$$\frac{\partial u}{\partial t} + u\,\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0 \ . \qquad (1)$$

This equation (the KdV equation) had been derived in 1885 by Korteweg and de Vries to describe long-wave propagation on shallow water. But until recently its properties were not well understood.

From a detailed numerical study Zabusky and Kruskal found that stable pulse-like waves could exist in a system described by the KdV equation. A remarkable quality of these solitary waves was that they could collide with each other and yet preserve their shapes and speeds after the collision. This particle-like nature led Zabusky and Kruskal to name such waves *solitons.* The first success of the soliton concept was explaining the recurrence in the Fermi-Pasta-Ulam system. From numerical solution of the KdV equation with periodic boundary conditions (representing essentially a ring of coupled nonlinear springs), Zabusky and Kruskal made the following observations. An initial profile representing a long-wavelength excitation would "break up" into a number of solitons, which would propagate around the system with different speeds. The solitons would collide but preserve their individual shapes and speeds. At some instant all of the solitons would collide at the same point, and a near recurrence of the initial profile would occur.

This success was exciting, of course, but the soliton concept proved to have even greater impact. In fact, it stimulated very important progress in the analytic treatment of initial-value problems for nonlinear partial differential equations describing wave propagation. During the past fifteen years a rather complete mathematical description of solitons has been developed. The amount of information on nonlinear wave phenomena obtained through the fruitful collaboration of mathematicians and physicists using this description makes the soliton concept one of the most significant developments in modern mathematical physics.

The nondispersive nature of the soliton solutions to the KdV equation arises not because the effects of dispersion are absent but because they are balanced by nonlinearities in the system. The presence of both phenomena can be appreciated by considering simplified versions of the KdV equation.

Eliminating the nonlinear term $u(\partial u/\partial x)$ yields the linearized version

$$\frac{\partial u}{\partial t} + \frac{\partial^3 u}{\partial x^3} = 0 \ . \qquad (2)$$

The most elementary wave solution of this equation is the harmonic wave

$$u(x,t) = A \exp\left[i(kx + \omega t)\right] \ , \qquad (3)$$

where k is the wave number and ω is the angular frequency. In order for the displacement $u(x,t)$ given by Eq. 3 to be a solution of Eq. 2, ω and k must satisfy the relation

$$\omega = k^3 . \tag{4}$$

Such a "dispersion relation" is a very handy algebraic description of a linear system since it contains all the characteristics of the original differential equation, Two important concepts connected with the dispersion relation are the phase velocity $v_p = \omega/k$ and the group velocity $v_g = \partial\omega/\partial k$. (For the dispersion relation given by Eq. 4, $v_p = k^2$ and $v_g = 3k^2$). The phase velocity measures how fast a point of constant phase is moving, while the group velocity measures how fast the energy of the wave moves. The waves described by Eq. 2 are said to be dispersive because a wave with large k will have larger phase and group velocities than a wave with small k. Therefore, a wave composed of a superposition of elementary components with different wave numbers (different values of k in Eq. 3) will disperse, or change its form. as it propagates.

Now we eliminate the dispersive term $\partial^3u/\partial x^3$ and consider the equation

$$\partial u/\partial t + u\partial u/\partial x = 0 . \tag{5}$$

This simple nonlinear equation also admits wave solutions, but they are now of the form $u(x,t) = f(x - ut)$, where the function f is arbitrary. (That $f(x - ut)$ is a solution of Eq. 5 is easily verified by substitution.) For waves of this form, the important thing to note is that the velocity of a point of constant displacement u is equal to that displacement. As a result, the wave "breaks"; that is, portions of the wave undergoing greater displacements move faster than, and therefore overtake, those undergoing smaller displacements, This multivaluedness is a result of the nonlinearity and, like dispersion, leads to a change in form as the wave propagates.

A remarkable property of the KdV equation is that dispersion and nonlinearity balance each other and allow wave solutions that propagate without changing form (Fig. 1), An example of such a solution is

$$u(x,t) = 3c\ \text{sech}^2[c^{1/2}(x - ct)/2] , \tag{6}$$

where the velocity c can take any positive value. This is the one-soliton solution of the KdV equation.

Although our discussion may have provided some glimpse of the interplay between dispersion and nonlinearity in the KdV equation, it has not. of course, provided any explanation of how solitons preserve their shapes and speeds after collision. This particle-like property is more than just a mere curiosity; it is of deep mathematical significance, A full understanding of this property requires an extensive mathematical discussion that we will not attempt here. We mention, however, that not all nonlinear partial differential equations have soliton solutions. Those that do are generic and belong to a class for which the general initial-value problem can be solved by a technique called the inverse scattering transform, a brilliant scheme developed by Kruskal and his coworkers in 1967. With this method, which can be viewed as a generalization of the Fourier transform to nonlinear equations, general solutions can be produced through a series of linear calculations. During the solution process it is possible to identify new nonlinear modes—generalized Fourier modes—that are the soliton components of the solution and, in addition, modes that are purely dispersive and therefore often called radiation. Equations that can be solved by the inverse scattering transform are said to be completely integrable.

The manifestation of balance between dispersion and nonlinearity can be quite different from system to system. Other equations thus have soliton solutions that are distinct from the bell-shaped solitons of the KdV equation. An example is the so-called nonlinear Schrodinger (NLS) equation. This equation is generic to all conservative systems that are weakly nonlinear but strongly dispersive. It describes the slow temporal and spatial evolution of the envelope of an almost monochromatic wave train. We present here a heuristic derivation of the NLS equation that shows how it is the natural equation for the evolution of a carrier-wave envelope. Consider a dispersion relation for a harmonic wave that is amplitude dependent:

$$\omega = \omega(k,|E|^2) . \tag{7}$$

Here $E = E(x,t)$ is the slowly varying envelope function of a modulated wave with carrier frequency ω and wave number k. The situation described by Eq. 7 occurs, for example. in nonlinear optical phenomena, where the dielectric constant of the medium depends on the intensity of the electric signal. Other examples include surface waves on deep water, electrostatic plasma waves, and bond-energy transport in proteins.

By expanding Eq. 7 in a Taylor's series about ω_0 and k_0, we obtain

$$\omega - \omega_0 = \frac{\partial\omega}{\partial k}\bigg|_0 (k - k_0) + \frac{1}{2}\frac{\partial^2\omega}{\partial k^2}\bigg|_0 (k - k_0)^2 + \frac{\partial\omega}{\partial(|E|^2)}\bigg|_0 |E|^2 . \tag{8}$$

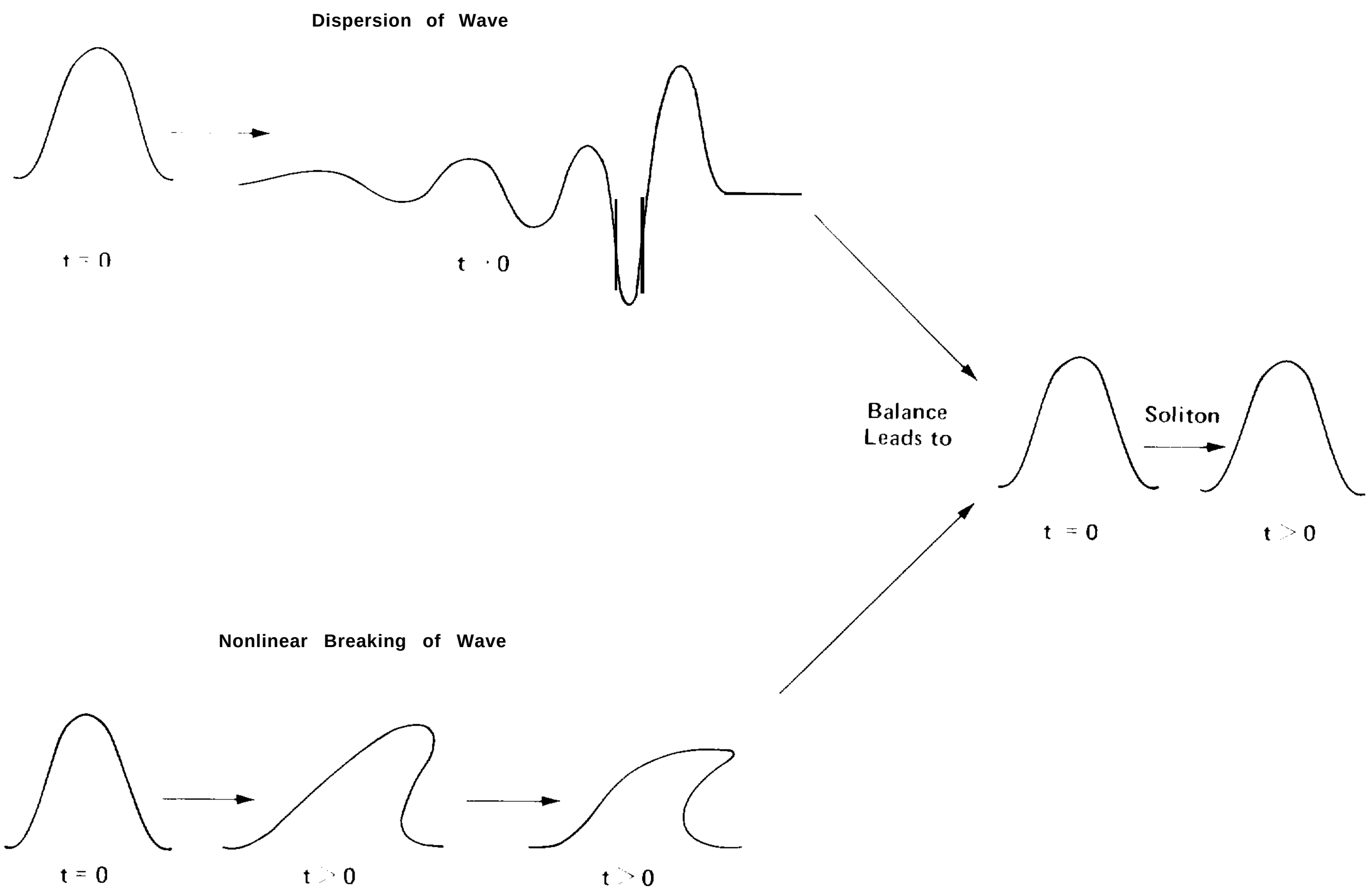

Fig. 1. Two effects, dispersion and breaking, cause the shape of a wave to change as it propagates. For a wave described by the KdV equation, these two effects balance, and the wave—a soliton—propagates without changing shape.

We have expanded only to first order in the nonlinearity but to second order in the dispersion because the first-order dispersion term, as we shall see, only represents undistorted propagation of the wave with the group velocity $v_g = [\partial\omega/\partial k]_0$. If we now substitute the operators $i(\partial/\partial t)$ for $\omega - \omega_0$ and $-i(\partial/\partial x)$ for $k - k_0$ in Eq. 8 and let the resulting expression operate on E, we get

$$i\left[\frac{\partial E}{\partial t} + \frac{\partial\omega}{\partial k}\bigg|_0 \frac{\partial E}{\partial x}\right] + \frac{1}{2}\,\frac{\partial^2\omega}{\partial k^2}\bigg|_0 \frac{\partial^2 E}{\partial x^2} - \frac{\partial\omega}{\partial(|E|^2)}\bigg|_0 |E|^2 E = 0 \,. \qquad (9)$$

This is the nonlinear Schrodinger equation, so called because of its resemblance to the Schrodinger equation even though its derivation often has nothing to do with quantum mechanics. The first term of Eq. 9 represents undistorted propagation of the wave at the group velocity. and the second and third terms represent its linear and nonlinear distortion. respectively. This crude derivation of the NLS equation shows how it arises in systems with amplitude-dependent dispersion relations. but more formal methods are necessary if detail about the coefficients, such as $[\partial\omega/\partial(|E|^2)]_0$, is required.

It is often preferable to express Eq. 9 in a neater form. For this purpose we transform the variables x and t into z and τ, where $z = x - \partial\omega/\partial k|_0 t$ is a coordinate moving with the group velocity and $\tau = 1/2[\partial^2\omega/\partial k^2|_0 t$ is the normalized time. Equation 9 then reduces to

$$i\frac{\partial E}{\partial \tau} + \frac{\partial^2 E}{\partial z^2} + 2\kappa|E|^2E = 0\ , \qquad (1o)$$

where

$$\kappa = -[\partial\omega/\partial(|E|^2)]_0/[\partial^2\omega/\partial k^2]_0\ . \qquad (11)$$

The NLS equation—like the KdV equation—is completely integrable and has soliton solutions. The analytic form for a single-soliton solution is given by

$$E(z,\tau) = 2\eta\ \mathrm{sech}[2\eta(\theta_0 - \eta z - 4\xi\eta\tau)]$$
$$\times \exp\{-2i[\phi_0 + 2(\xi^2 - \eta^2)t + \xi z]\}\ , \qquad (12)$$

where ξ, η, θ_0, and ϕ_0 are free parameters determining the speed, amplitude, initial position, and initial phase, respectively, of the soliton. Figure 2 shows the profile of this soliton.

Any initial excitation for the NLS equation will decompose into solitons and/or dispersive radiation. A monochromatic wave train solution $E(z,\tau) = E(\tau)$ is thus unstable to any z-dependent perturbation and breaks up into separate and localized solitons. This phenomenon is called the Benjamin-Feir instability and is well known to any surfer on the beach who has noticed that every, say, seventh wave is the largest. The NLS equation is in a way more universal than the KdV equation since an almost monochromatic, small-amplitude solution of the KdV equation will evolve according to the NLS equation.

The last type of soliton we mention, which is distinctly different from the KdV or NLS solitons, is one that represents topologically invariant quantities in a system. Such an invariant can be a domain wall or a dislocation in a magnetic crystal or a shift in the bond-alternation pattern in a polymer. The prototype of equations for such solitons is the sine-Gordon equation,

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} = \sin u\ . \qquad (13)$$

Notice that this equation allows for an infinite number of trivial

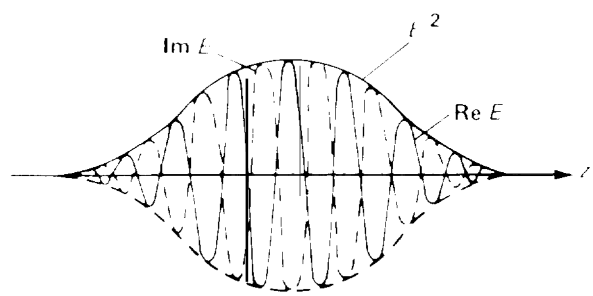

Fig. 2. Profile of a single-soliton solution of the NLS equation.

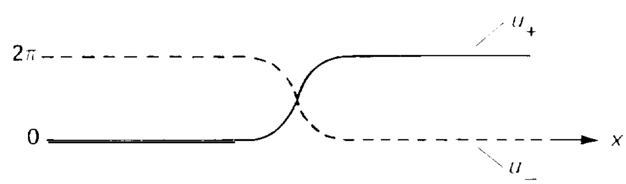

Fig. 3. Profiles of soliton solutions of the sine-Gordon equation.

RELATED TOPICS

solutions, namely $u = 0, \pm 2\pi, \pm 4\pi, \ldots$. Systems with a multitude of such "degenerate ground states" also allow solutions that connect two neighboring ground states. Solutions of this type are often called kinks, and for the sine-Gordon equation they are exact solitons; that is, they collide elastically without generation of dispersive radiation. The analytic form, whose profile is shown in Fig. 3, is given by

$$u_{+}(x,t) = 4 \tan^{-1}\{\exp[\pm(x - x_0 - ct)/(1 - c^2)^{1/2}]\} , \qquad (14)$$

where the solution u– is often called an antikink. The parameter c $(-1 < c < 1)$ determines the velocity and x_0 the initial position, Other equations with degenerate ground states also have kink and antikink solutions, but they are not exact solitons like those of the sine-Gordon equation. It is interesting to note that small-amplitude solutions of the sine-Gordon equation also can be shown to evolve according to the NLS equation.

Equations with soliton solutions are generic, and, although real systems often contain mechanisms (impurities, dissipative forces, and multidimensionality) that destroy exact soliton behavior, they are very useful as a starting point for analysis. In fact, perturbation methods—with the perturbation taking place around the soliton—have been developed to compute the response of the soliton to external forces, damping, etc. Often the result is that the parameters characterizing the soliton (such as velocity and amplitude) are now time dependent, with the time dependence governed by simple ordinary differential equations. The original equations are therefore still very useful. Because the mechanisms that give rise to soliton equations are so prevalent, the suggestion that solitons might arise in biology is not so surprising. The question to be asked is how well a particular biological system satisfies the criteria underlying the soliton equation. ■

Further Reading

The classic paper where the word "soliton" was introduced is "Interaction of 'Solitons' in a Collisionless Plasma and the Recurrence of Initial States" by N. J. Zabusky and M. D. Kruskal in *Physical Review Letters 15(1965):240.* For many references see also "Computational Synergetics and Mathematical Innovation" by Norman J. Zabusky in *Journal of Computational Physics 43(1981):95.*

There are an increasing number of papers on solitons; a good review paper covering the subject to 1973 is "The Soliton: A New Concept in Applied Science" by Alwyn C. Scott, F. Y. F. Chu, and David W. McLaughlin in *Proceedings of the IEEE 61(1973): 1443.*

Good accounts of the subject, together with up-to-date lists of references, can also be found in many textbooks, including the following.

G. L. Lamb, Jr. *Elements of Soliton Theory.* New York: John Wiley & Sons, 1980.

Mark J. Ablowitz and Harvey Segur. *Solitons and the Inverse Scattering Transform.* Philadelphia: Society for Industrial and Applied Mathematics, 1981.

R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, and H. C. Morris. *Solitons and Nonlinear Wave Equations.* New York: Academic Press, 1982.