Theta functions, Gaussian series, and spatially periodic solutions of the Korteweg–de Vries equation

John P. Boyd
Department of Atmospheric and Oceanic Science, University of Michigan, Ann Arbor, Michigan 48109

(Received 31 July 1981; accepted for publication 4 September 1981)

It has been shown by Novikov [Funct. Anal. Appl. **8**, 236 (1974)], Dubrovin *et al.* [Russian Math. Surveys **31**, 59 (1976)], Lax [Commun. Pure Appl. Math. **28**, 141 (1975)], McKean and van Moerbeke [Inv. Math. **30**, 217 (1975)], and others that the nonlinear evolution equations which admit solitary waves also have spatially periodic exact solutions ("polycnoidal waves") which can be expressed in terms of multidimensional Riemann theta functions. Here, it is shown that via Poisson summation, the Fourier series that define the theta functions can be transformed into an infinite series of Gaussian functions. Because the lowest terms of the Gaussian series generate the usual solitary waves, it is possible to intimately explore the relationship between solitary waves and these spatially periodic "polycnoidal" waves. Also, by using the Gaussian series, one can perturbatively calculate phase velocities and wave structure for the "polycnoidal" wave even in the strongly nonlinear regime for which the soliton (or multisoliton) is the lowest order approximation. It is further shown that the Fourier series and the complementary Gaussian series both converge so rapidly in the intermediate regime of moderate nonlinearity that one may loosely state that a solitary wave is almost a linear wave, and a linear wave almost a soliton. Thus, by using both series together, one can obtain a very complete description of these stable, finite amplitude, periodic solutions. For expository simplicity, this first discussion of the Gaussian series approach to "polycnoidal" waves will concentrate on the most elementary example: the ordinary "cnoidal" wave of the Korteweg–de Vries equation. The great virtue of the Poisson method, however, is that it extends almost trivially to other equations (the Nonlinear Schrödinger equation, the Sine–Gordon equation, and a multitude of others) and also to periodic solutions of these equations that are describable in terms of higher dimensional theta functions ("polycnoidal" waves). The next to last section proves a number of generalizations of the theorems of Hirota [Prog. Theor. Phys. **52**, 1498 (1974)] applicable both to "cnoidal" and "polycnoidal" solutions without restriction, and explains how these extensions will work.

PACS numbers: 02.60. + y

1. INTRODUCTION

The first exact, nonlinear, spatially periodic solutions to an evolution equation of the class discussed here were obtained by Korteweg and de Vries[1] 85 years ago for the equation that now bears their name. They showed that their equation, henceforth referred to by the abbreviation KdV, has steadily translating waves that can be mathematically described by the elliptic cosine function cn(x; m). Since it was like a function whose abbreviation is "cn," they called these waves "cnoidal" waves. The nonlinear, spatially periodic solutions discussed here are generalizations of these cnoidal waves that, in the absence of any generally accepted terminology, will be referred to as "polycnoidal" waves in the rest of the paper. The reason that these generalized waves are important is that it appears that *any* spatially periodic solution of the KdV equation—or a number of other equations in the same class—can be approximated to any chosen degree of accuracy for any chosen finite time interval by an appropriate "polycnoidal" wave. Thus, to understand these generalized cnoidal waves is also to understand the general spatially periodic solution to the Korteweg–de Vries and other evolution equations.

The elliptic cosine depends on a parameter m (the "modulus"); Korteweg and de Vries showed that the limit $m = 0$ corresponds to a linear wave and the elliptic cosine reduces to the ordinary cosine. In the limit $m \rightarrow 1$, the spatial period of the wave becomes infinite, the elliptic cosine becomes the hyperbolic secant function, and the cnoidal wave becomes the solitary wave discovered observationally in 1831[2]: an isolated, steadily translating, finite amplitude peak of permanent form. Equivalently, if one rescales the variables so as to keep the spatial period fixed, the solitary wave or "soliton" corresponds to the limit of infinitely large amplitude (the spatial structure of the soliton tends toward that of a delta function) while the linear wave as usual is the limit of infinitesimal amplitude.

"Polycnoidal" waves also tend to linear waves and solitons in appropriate limits, and the relationship between these limits and the actual polycnoidal waves of intermediate amplitude is one of the major themes of this paper. Before discussing how we propose to explore these relationships, it is appropriate (and necessary) to briefly review the major developments in the theory of the Korteweg–de Vries equation.

Most work on this equation has studied its solutions subject to two different species of initial/boundary conditions: (i) the unbounded problem in which $x \in [-\infty, \infty]$ and the initial condition is localized, i.e., is exponentially small everywhere outside of a finite interval, and (ii) the spatially periodic problem in which both the initial condition and the

 0022-2488/82/030375-13$02.50

solution for all later times are required to be periodic functions of x. Korteweg and de Vries were able to derive a solution for the unbounded problem (the solitary wave) as a limit of their class of solutions for the periodic problem, but in later work, these two problems have represented completely separate lines of development.

Little progress with either was made until 1967, when Gardner *et al.*[3] showed the unbounded problem, although the KdV equation itself is nonlinear, could be solved exactly through a sequence of solving only linear equations: the so-called inverse scattering method. This procedure was to the study of nonlinear equations what the Rosetta Stone was to Egyptology, and it was subsequently extended to a large class of other equations, including the Nonlinear Schrödinger equation, the sine–Gordon equation, and many others. Although for simplicity, the discussion here will concentrate on the KdV equation, the ideas and techniques explained here extend to all members of this class of "exactly integrable" nonlinear equations.

The inverse scattering analysis showed that the general solution to the KdV equation (and its fellows) consists of two parts: a finite number of solitary waves or "solitons," which are permanent, isolated, finite amplitude waves, plus "radiation," which is used as a catch-all term to describe the miscellaneous peaks and ripples that eventually disperse so that the solitons are the sole asymptotic solution as $t \to \infty$. The reason for the name "soliton," with its connotation of particlelike rather than wavelike properties, is that when solitons collide, they eventually emerge from collision unchanged in shape, size, or speed except for a phase-shift.

Unfortunately, although inverse scattering is so useful for theoretical and qualitative purposes, researchers have found that for obtaining quantitative results and case studies, the steps of the inverse scattering method are so cumbersome that it is easier to numerically integrate the KdV equation directly using a conventional time-marching scheme. Because of this clumsiness, Hirota[4,5] introduced an alternative approach to the unbounded problem which is the direct ancestor of the methods to be used here.

Hirota's technique is based on a logarithmic transformation of the dependent variable to give a transformed version of the KdV equation which, to avoid confusion, will be referred to as the Hirota–Korteweg–de Vries equation or H–KdV for short. Hirota showed that although the H–KdV equation is nonlinear, it is possible to construct exact solutions by adding finite sums of exponentials from which the exact, multisoliton solutions of the KdV equation can be obtained through the inverse transformation. Hirota's method is very simple and involves nothing more exotic than differentiation, logarithms, and exponentials, but it has the weakness of excluding the "radiation" part of the general KdV solution. However, for smooth, large-amplitude initial conditions, almost all of the initial energy goes into the solitons anyway (referred to as the principle of "soliton dominance" in Boyd[6]), so this restriction is not fatal, and Hirota's method is still actively used in research today (for example, Ma and Redekopp[7]) even when the exact inverse scattering algorithm is known for the same equation (Ma[8]).

Meanwhile, independent groups of American[9] and Russian[10,11] mathematicians have developed an analog of the inverse scattering method for obtaining a class of generalizations of the cnoidal waves for (ii), the spatially periodic problem. Unfortunately, it is again true that the formal solution, which will be here called the "Hill's spectrum method," is as cumbersome computationally as it is powerful theoretically.

However, the author and Nakamura[12] independently realized that there is an alternative. The transformation that relates the theta functions to the actual solution of the Korteweg–de Vries equation is exactly that made by Hirota—in other words, the theta functions are exact solutions not of the KdV equation itself but rather of Hirota's transformed version, the H–KdV equation. This suggests what Nakamura calls the "direct method": computing nonlinear phase speed corrections and theta function parameters via direct substitution of the theta series into the H–KdV equation and matching of Fourier series coefficients. Thus, there are now two approaches—the formal exact method based on quantum mechanics potential theory and a more heuristic but vastly simpler alternative based on solving Hirota's transformed equations—for both the infinite and periodic spatial domains.

In later work, Nakamura, Hirota, Ito, and Matsuno[13–15] have greatly extended this direct theta function procedure and their papers are highly original and a treasurehouse of useful information. However, they work exclusively with the Fourier series representation of the theta functions and omit all mention of what are the principal themes of this paper: that a direct method using the Gaussian series representation is not only possible but is more useful than its Fourier series counterpart.

The reason for the greater usefulness of the Gaussian series is that one can obtain most of the information derivable from the theta–Fourier series by directly attacking the Korteweg–de Vries equation via the singular perturbation technique known variously as the method of strained parameters or the method of multiple scales.[16,17] Because the multiple-scales approach is conceptually useful also in understanding precisely what a "polycnoidal" wave is, some multiple scales calculations for polycnoidal waves are given in Appendix B.

The Gaussian series representation, however, is directly tied to the fact that the exact solutions of the H–KdV equation are theta functions, and it gives results that cannot be reproduced by any conventional perturbation scheme. As a consequence, the rest of the paper will concentrate upon the Gaussian representation of the theta function, except for the appendices.

Sections 2–5 describe the simplest example: the ordinary cnoidal wave for the Korteweg–de Vries equation. Despite the fact that this was solved 85 years ago, the theta–Gaussian method will nonetheless yield some new results. Later sections of the paper and future work now in progress will deal with polycnoidal waves and other evolution equations. But the point in beginning with this simple example is that the ideas explained through it are the key to everything else.

2.THE HIROTA–KORTEWEG–DE VRIES EQUATION AND THE THETA FUNCTION

The Korteweg–de Vries equation itself is

$$u_t + uu_x + u_{xxx} = 0. \tag{2.1}$$

The transformation to the H–KdV equation is made in two stages. First, set $u = p_x$ and then integrate (2.1) once with respect to x to obtain

$$p_t + 1/2p_x^2 + p_{xxx} = A, \tag{2.2}$$

where A is an arbitrary constant of integration. In deriving the multiple soliton solutions, one can set $A = 0$ without loss of generality, but as noted by Nakamura,[12] this is absolutely disastrous for the polycnoidal wave. Instead, A must be computed as a function of the parameters in the same way as the nonlinear phase speed. The second step is to introduce the nonlinear transformation

$$p = 12(\ln F)_x. \tag{2.3}$$

Substituting (2.3) in (2.2), on finds that all third and fourth degree terms in F identically cancel to leave the H–KdV equation:

$$F(F_t + F_{xxx})_x - F_x(F_t + F_{xxx}) + 3(F_{xx}^2 - F_x F_{xxx}) = AF^2. \tag{2.4}$$

Hirota showed that (2.4) and indeed all his transformed equations can be expressed more compactly by using certain bilinear operators, but this alternative version of (2.4) will be deferred until Sec. 6, where Hirota's bilinear operators will be useful in proving certain theorems. The solution of the KdV equation is

$$u = 12(\ln F)_{xx}, \tag{2.5}$$

where F is a theta function.

The argument of the theta function is

$$X = x - ct + \phi, \tag{2.6}$$

where c is the phase speed and ϕ a constant phase factor. For higher-dimensional theta functions, we have additional arguments of the form $Y = k_2(x - c_2t) + \phi_2$, $Z = k_3(x - c_3t) + \phi_3$, as one would anticipate from the fact that the polycnoidal waves reduce to multiple solitary waves in the appropriate limit. Since the N-soliton solution consists of N distinct peaks each with its own width and phase speed, it follows that a function of N variables of the form of X, Y, Z, etc., are necessary to provide a compact description of the waves. By rescaling via a theorem proved in Sec. 7, one can always set one of the wave numbers k equal to 1 without loss of generality, and this has been done in (2.6).

For the one-dimensional case, one can take the theta function to be

$$\theta_4(X; q) = 1 + \sum_{n=1}^{\infty} (-1)^n q^{n^2} \cos(2nX), \tag{2.7}$$

where q is a constant known as the "nome." When working with theta functions, the nome is a more convenient measure of the "ellipticity" of the elliptic cosine than the modulus m; the relationship between them is given in Appendix A. However, q has one thing in common with m: the limit $q \to 0$ again gives the linear wave while $q \to 1$ gives the solitary wave. From the form of the Fourier series in (2.7), one can show that it converges uniformly and absolutely for all $q < 1$, but it is self-evident that the rate of convergence becomes poorer and poorer as the soliton limit is approached. Consequently, it is the alternative Gaussian series for the theta function that one must use to explore the relationship between solitons and cnoidal waves.

Nonetheless, even from a perspective focused entirely on Fourier series, the theta series (2.7) converges much more rapidly than that for $u(x)$ itself, which is

$$u = 12(\ln \theta_4)_{XX} = 96q \sum_{n=1}^{\infty} \frac{nq^{n-1}}{1 - q^{2n}} \cos(2nX), \tag{2.8}$$

as obtained by differentiating the known[18] series for the logarithmic derivative of the theta function. The elliptic cosine, whose square gives $u(x)$ (see Appendix A for details), is a meromorphic function, and this alone[19] is enough to prove that the coefficients of the Fourier series must be asymptotically $O(q^n)$ for some constant q, $|q| < 1$. [This, in fact, is true for the hyperelliptic functions that give $u(x, t)$ for the polycnoidal waves as well.] The theta functions, however, whether in one or many dimensions, are entire functions. For θ_4, we see that the Fourier coefficients are $O(q^{n^2})$ so that 10 terms of the theta Fourier series give the same accuracy as 100 terms of the Fourier series for cn^2. It is precisely this very rapid convergence for the entire function as opposed to the meromorphic function that led C. G. Jacobi to introduce the theta functions in the first place and build his entire approach to elliptic functions around them. It is precisely because the Gaussian series for the theta functions shares this same very rapid convergence that it will be shown to be a powerful tool for understanding strongly nonlinear, spatially periodic waves.

Poisson summation[20] of (2.7) gives the Gaussian series, which may be written in either of the two forms:

$$\theta_4(X; q) = \begin{cases} 2s^{1/2}\exp\left\{-sX^2/\pi\right\} \sum_{n=0}^{\infty} q'^{(n+1/2)^2} \cosh[(2n+1)sX] & (2.9) \\ s^{1/2} \sum_{-\infty}^{\infty} \exp[-s(X - \frac{1}{2}\pi(2n+1))^2/\pi]. & (2.10) \end{cases}$$

where q', the "complementary nome," and s are defined by

$$q' \equiv e^{\pi^2/\ln q}, \tag{2.11}$$

$$s = -\pi/\ln q. \tag{2.12}$$

The relationships between q, q', and s can be expressed more symmetrically in the form

$$q = e^{-\pi/s}, \tag{2.13}$$

$$q' = e^{-\pi s}. \tag{2.14}$$

As $q \to 1$, $q' \to 0$ and vice versa so that the Gaussian series [(2.9) or (2.10)] and the Fourier series (2.7) are indeed complementary, with one converging rapidly in the parameter range where the other converges slowly.

The first form of the Gaussian series, (2.9), which can be obtained from (2.10) by multiplying out the exponents in (2.10), extracting the common factor of $\exp(-sX^2/\pi)$, and combining exponentials into hyperbolic cosines, is the one that is most closely analogous to the Fourier series. Because

q' appears explicitly in (2.9), this form is the one that is useful for practical perturbative calculations.

The second form, (2.10), is more useful conceptually: it shows that the theta function can be represented as a series of Gaussians of identical size and shape spaced at intervals of π on the whole interval $X\in[-\infty, \infty]$. For large q' (small q), the Gaussians overlap heavily and the shape is roughly that of an ordinary cosine function with symmetrical crests and troughs. For small q'—the near-soliton-limit—the Gaussians are well separated and the theta function takes the form of sharp, narrow peaks separated by broad, flat troughs.

In the limit $q'\to 0$, only the two Gaussians at $X = \pm\pi/2$ are significant because the contributions of all the other terms in the series are exponentially small on the interval $X\in[-\pi/2, \pi/2]$. This suggests that—despite all appearances to the contrary—this limiting solution of two Gaussians is somehow equivalent to the single soliton solution of Hirota, which is the sum of a constant plus an exponential. In the next section, we shall see that this is indeed the case and explore how the transition from soliton to cnoidal wave is made.

3. THE BI-GAUSSIAN SOLITON AND THE PROBLEM OF PERIODICITY

A. The bi-Gaussian soliton

By direct substitution, one can show

$$\Theta(X) = e^{-(L/2)X^2}\cosh(sX) \tag{3.1}$$

is an *exact* solution of the Hirota–Korteweg–de Vries equation (2.4) for *arbitrary* values of s and L provided that

$$c = 4s^2 - 12L, \tag{3.2}$$

$$A = 2L(2s^2 - 3L), \tag{3.3}$$

where c is the phase speed, A the constant of integration in the H–KdV equation, and $X = x - ct + \phi$ as defined by (2.6) with ϕ arbitrary. $\Theta(X)$ is[21] the sum of two Gaussians with peaks at $X = \pm s/L$, so it will be referred to as the bi-Gaussian soliton solution. Taking the second logarithmic derivative gives

$$u(x,t) = -12L + 12s^2\,\mathrm{sech}^2[s(x-ct)]. \tag{3.4}$$

As mentioned earlier, the theta functions of the proceeding section are functions of but a single parameter—and so are the one-soliton solutions of the KdV equation in the form usually given—whereas the bi-Gaussian $\Theta(X;s,L)$ contains *two* independent parameters. However, one can see from (3.4) that the extra parameter corresponds merely to the freedom (mathematically, if not physically!) to change mean sea level in the Korteweg–de Vries equation. In mathematical terms, it is trivial to prove that if $u(x-ct)$ is any solution to the KdV equation, than

$$\tilde{u} \equiv a + u(x-(c+a)t) \tag{3.5}$$

is also a solution. Thus, cnoidal waves and the one-soliton solution are really two-parameter families, but one of the parameters, the constant a in (3.5), is trivial, and only the nome q is significant. However, with the bi-Gaussian, we can obtain the full two-parameter single-soliton family by varying s and L.

Hirota's solution is the special case $L = 0$ that gives

$$\Theta = \cosh(sX). \tag{3.6}$$

This gives the usual soliton with zero mean sea level, but it is not quite in standard form itself. However, one can show[22] that any solution of the H–KdV equation is still a solution if multiplied by an arbitrary constant times an exponential whose argument is linear in X. Multiplying (3.6) by 2 $\exp(sX)$ gives

$$F = 1 + \exp(2sX), \tag{3.7}$$

which is the form in which Hirota's 1-soliton solution is usually given.

Unfortunately, to extend this bi-Gaussian function into the infinite series for the cnoidal wave, we must take a nonzero value of L. This in turn means accepting the annoying complication of a shift in mean sea level via the term $-12L$ in (3.4) and an identical shift also in the phase speed (3.2). To have generalized Hirota's one-parameter solution (3.7) to a two-parameter one is not in itself a very useful accomplishment. However, it is the bi-Gaussian—not (3.7)—that generates the cnoidal wave, so one must understand the shifts introduced by L to make correct comparisons between exact and approximate cnoidal wave solutions as shall be done in Sec. 5.

B. Periodicity

If one approached the problem of computing approximate periodic solutions to the KdV equation with no knowledge of theta functions—but a knowledge of solitons—one heuristic approach would be to approximate the cnoidal wave by an infinite series of evenly spaced hyperbolic secant functions, i.e.,

$$u(x,t) \doteq 12s^2 \sum_{n=-\infty}^{\infty} \mathrm{sech}^2(s(X-n\pi)). \tag{3.8}$$

When s is large, the soliton peaks are narrow and well-separated with little overlap, so (3.8), with c given by the usual soliton formula (3.2), is indeed a consistent first approximation to the cnoidal wave with an error which decreases exponentially fast as s increases.

The only problem is that there is no particularly good way to calculate higher-order corrections—to explore precisely how the periodicity has altered the cnoidal wave from the soliton. One could substitute (3.8) into the Korteweg–de Vries equation, but the inhomogeneous terms even at lowest order would involve fourth powers of reciprocal hyperbolic functions, and inverting the linear part of the Korteweg–de Vries equation requires inverting a partial differential operator. Furthermore, the error in neglecting higher values of n in the series (3.8) is an exponential function of s, but algebraic powers of s would also appear. In short, perturbative theory using (3.8) as the lowest approximation would be a horrible mess, requiring great analytical ingenuity to obtain even the first and second corrections. Furthermore, the shape of u as a function of X, as well as the nonlinear phase speed c, would both have to be corrected order-by-order.

If we apply this same heuristic philosophy to the

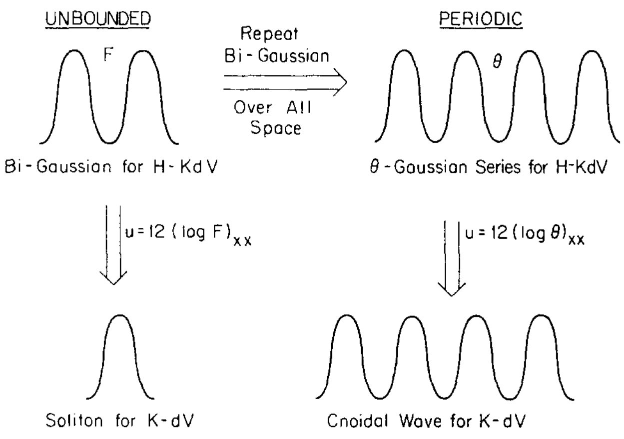

FIG. 1. Schematic diagram showing the relationship between the bi-Gaussian and theta functions solutions to the H–KdV equation. The left side shows the situation when the domain is unbounded: the solution to the H–KdV equation has just two peaks on all of $X \in [-\infty, \infty]$, and the second logarithmic derivative of this gives a single peak which is the usual solitary wave. When bi-Gaussian pattern is repeated with even spacing over all X, it generates the Gaussian series of the theta function. This, as shown on the right, is a periodic solution of the H–KdV equation and its second logarithmic derivative gives the usual cnoidal wave.

H–KdV equation, using the bi-Gaussian soliton (3.1) with

$$L = (2/\pi)s, \tag{3.9}$$

so that the peaks of the Gaussians are π units apart, we obviously obtain an infinite series of evenly spaced Gaussians as indicated schematically in Fig. 1. But we have already seen in the previous section that such a series of Gaussians is an *exact* representation of the theta function. Consequently, by taking the single soliton solution and repeating it with even spacing over the whole spatial domain, we obtain the *exact* solution for the spatially periodic problem—but only when we work through the H–KdV equation, Hirota's transformed equation, rather than through the KdV equation itself. The series of hyperbolic secant functions in (3.8) is only approximate.

The same strategy works equally well to generate periodic extensions of the multiple soliton solutions. The double soliton, for example, is given by the sum of four evenly spaced Gaussians forming a square in the XY plane where $X = x - c_1 t + \phi_1$, $Y = k_2(x - c_2 t) + \phi_2$, and where each Gaussian is now an exponential whose argument is a second-degree polynomial in both X and Y. Repeating this basic four-Gaussian unit over the whole XY plane with even spacing gives the Gaussian series for the two-dimensional theta function, and the second logarithmic derivative of this gives the double cnoidal wave solution to the KdV equation.

Because only the nonlinear phase speed (or speeds) need be corrected, it is trivial in principle (although the algebra can become tedious) to calculate the cnoidal and polycnoidal waves to any order. For the cnoidal wave, for example, one obtains at each order two linear equations in two unknowns which can be solved to obtain c and A to that order, with the expansion proceeding in powers of the complementary nome q': in Sec. 4, we shall show how one can obtain the full infinite series for c almost trivially.

Before we do this, however, one final point must be made; Hirota's single soliton solution for the H–KdV equation, (3.7), is spatially *unbounded* even though, thanks to the magic of the logarithmic derivative, it gives the usual sech^2 solution when the transformation back to the KdV equation is made. Consequently, it is not possible to add evenly spaced Hirota functions of the form of (3.7) to obtain any sort of meaningful series. It was therefore absolutely necessary to generalize Hirota's solution to the bi-Gaussian so as to obtain a form for the single soliton which would be *spatially bounded* and *localized* for the Hirota–Korteweg–de Vries equation, too. Only then would it be possible to repeat the single-soliton solution over the whole domain with even spacing to obtain a solution that is manifestly periodic and solves the H–KdV equation exactly.

It is the bi-Gaussian, and not the simple exponential solution of Hirota, that is the proper generator of the cnoidal wave.

4. THE RESIDUAL EQUATIONS AND THE EXACT SOLUTION FOR THE CNOIDAL WAVE

A. The residual equations

The first step in obtaining the full series solution is to rewrite the Gaussian series (2.9) as

$$\theta_4(X;q) = s^{1/2} e^{-sX^2/\pi} \sum_{\substack{-\infty \\ \text{[half integers]}}}^{\infty} q'^{n^2} e^{2nsX}, \tag{4.1}$$

where the hyperbolic cosines have been broken up into pairs of exponentials and where, in order to eliminate factors of $\frac{1}{2}$, the sum is taken over all the "half integers" $\{\ldots, -\frac{5}{2}, -\frac{3}{2}, \frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots\}$. Substituting this into the H–KdV equation gives the residual

$$\rho \equiv s e^{-2sX^2/\pi} \sum_{\substack{-\infty \\ \text{[half integers]}}}^{\infty} \sum_{-\infty}^{\infty} q'^{n^2 + n'^2} \zeta(n - n', c, A) e^{2(n+n')sX}. \tag{4.2}$$

We must solve for c and A such that $\rho \equiv 0$.

One might expect that, since the H–KdV equation (2.4) involves differentiations of up to fourth order with respect to x and first order with respect to time, the Gaussian in (4.1) would cause $\zeta(n - n'; c, A)$ to be a polynomial of fourth degree in x and first degree in t. In fact, as shall be proven in Sec. 6, because of cancellations ζ is *independent* of both x and t, just as if we had substituted the Fourier series for θ_4 into the H–KdV equation instead.

The second important property of ζ—also noted by Nakamura for its Fourier series equivalent—is that it is a function only of the *difference* $(n - n')$, and not of n and n' separately. This property, again proved in Sec. 6, is true for all of Hirota's transformed nonlinear evolution equations, and is sufficient—without specification of the precise form of $\zeta(n - n'; c, A)$—to prove that the residual is the sum of two theta functions. Consequently, the exact definition of ζ will be deliberately postponed to the next subsection.

Keeping these two properties of ζ in mind, the next step is to define $j = n + n'$ and collect terms in $\exp(2jsX)$ to obtain

$$\rho = s e^{-2sX^2/\pi} \sum_{\substack{j = -\infty \\ \text{[integers]}}}^{\infty} R_j e^{2jsX}, \tag{4.3}$$

where

$$R_j \equiv \sum_{\substack{n=-\infty \\ \text{[half integers]}}}^{\infty} q'^{n^2+(j-n)^2}\zeta(2n-j;c,A). \qquad (4.4)$$

The residual ρ vanishes if and only if all the $R_j=0$, where j is any integer, so we seem to have rather more equations (an infinite number!) than we have unknowns (2, c and A). However, it is easy to prove that all even R_j are multiples of R_0 and all the odd R_j are multiples of R_1.

For brevity, we shall deal only with the even R_j. From (4.4)

$$R_{2J} \equiv \sum_{\substack{n=-\infty \\ \text{[half integers]}}}^{\infty} q'^{n^2+(n-2J)^2}\zeta(2[n-J]). \qquad (4.5)$$

Defining

$$N \equiv n - J \qquad (4.6)$$

gives

$$R_{2J} = \sum_{\substack{n=-\infty \\ \text{[half integers]}}}^{\infty} q'^{(N+J)^2+(N-J)^2}\zeta(2N) \qquad (4.7)$$

$$= q^{2J^2} \sum_{\substack{n=-\infty \\ \text{[half integers]}}}^{\infty} q'^{2N^2}\zeta(2N) \qquad (4.8)$$

$$= q'^{2J^2}R_0 \quad \text{for all integers } J. \qquad (4.9)$$

Similarly, one can show

$$R_{2J+1} = q'^{2J^2+2J}R_1. \qquad (4.10)$$

By using (4.1) and the equivalent series for $\theta_3(x;q)$ $[\equiv\theta_4(x+\pi/2;q)]$, which is identical in form with (4.1) except that the sum is taken over all integers instead of half integers, one can rewrite (4.2) as

$$\rho = [R_0\theta_3(2X;q'^2) + R_1\theta_4(2X;q'^2)]se^{-sX^2/\pi}. \qquad (4.11)$$

Thus, the solution of the cnoidal wave has been reduced to solving two equations in two unknowns

$$R_0(c,A) = 0, \qquad (4.12)$$

$$R_1(c,A) = 0. \qquad (4.13)$$

For the Korteweg–de Vries equations, these coupled equations are *linear*; for the Boussinesq equation of Nakamura,[12] these equations are quadratic in c, but can still be solved explicitly.

B. Theta matrix doubling

One important aspect of (4.11) that is independent of the nonlinear evolution equation is that all the quantities involved—R_0, R_1, the theta functions, c, and A—are functions of q'^2 rather than q'. In particular, the series that define R_0 and R_1, are series in q'^{2n^2}, and thus converge much more rapidly than even the fast-converging series for the theta functions themselves, whose coefficients are q'^{n^2}. This same phenomenon holds for the polycnoidal waves, too, where it will be called "theta matrix doubling" for notational reasons explained in Sec. 7. It is this very rapid convergence that makes perturbative calculations for two- and three-dimensional theta functions feasible.

C. The exact solution for the Korteweg–de Vries equation

For the Korteweg–de Vries equation, one can show by direct substitution of (4.1) into the H–KdV equation (3.4) or by using general theorems of Hirota's bilinear operations proved in Sec. 6 that

$$\zeta(n-n';c,A) = \left\{\frac{48s^2}{\pi^2} - 2A + \frac{4sc}{\pi}\right\} + 16s^4(n-n')^4 - 4cs^2(n-n')^2 - \frac{96s^3}{\pi}(n-n')^2, \qquad (4.14)$$

which depends on n and n' via

$$\Delta = (n-n')^2, \qquad (4.15)$$

and (4.12) and (4.13) become the matrix equation

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} c \\ A \end{vmatrix} = \begin{vmatrix} \Omega_1 \\ \Omega_2 \end{vmatrix}, \qquad (4.16)$$

where

$$A_{11} = \frac{4s}{\pi}\Bigg(\sum_{\substack{n=-\infty \\ \text{[half integers]}}}^{\infty} q'^{2n^2}\Bigg) - 16s^2\Bigg(\sum_{\substack{n=-\infty \\ \text{[half integers]}}}^{\infty} n^2 q'^{2n^2}\Bigg), \qquad (4.17)$$

$$A_{12} = -2\Bigg(\sum_{\substack{n=-\infty \\ \text{[half integers]}}}^{\infty} q'^{2n^2}\Bigg), \qquad (4.18)$$

$$\Omega_1 = 16s^2 \sum_{\substack{n=-\infty \\ \text{[half integers]}}}^{\infty} \left(\frac{3}{\pi^2} - \frac{24s}{\pi}[n^2] + 16s^2[n^4]\right)q'^{2n^2}, \qquad (4.19)$$

and where the elements of the second row (corresponding to R_1) are, after multiplication by $q'^{1/2}$, identical with those of the first row except that (i) the sums are taken over the *integers* instead of the half integers and (ii) the terms $n=0$ are taken with a factor of $\frac{1}{2}$ as in (4.15). A more compact description can be obtained by defining

$$H(q') \equiv \sum_{\substack{n=-\infty \\ \text{[half integers]}}}^{\infty} q'^{2n^2} = \frac{-\pi^{1/2}}{[\ln q]^{1/2}}\theta_4(0;q^{1/2}), \qquad (4.20)$$

$$I(q') \equiv \sum_{\substack{n=-\infty \\ \text{[integers]}}}^{\infty} q'^{2n^2} = \frac{-\pi^{1/2}}{[\ln q]^{1/2}}\theta_3(0;q^{1/2}), \qquad (4.21)$$

where the right-hand sides of (4.20) and (4.21) follow by evaluating (4.1) and the corresponding series for θ_3 at $X=0$ and then using the usual relationship between q and q', (2.11). The fact that q appears as $q^{1/2}$ is a consequence of the fact that the series defining $H(q')$ and $I(q')$ are theta-function series in q'^2 rather than q' itself as explained in the previous subsection. Letting a subscripted q' denote differentiation with respect to q', we have

$$A_{11} = (4s/\pi)H - 8s^2q'H_{q'}, \qquad (4.22)$$

$$A_{12} = -2H, \qquad (4.23)$$

$$\Omega_1 = 16s^2\Big(\frac{3H}{\pi^2} - \frac{12sq'}{\pi}H_{q'} + 4s^2q'[q'H_{q'}]_{q'}\Big), \quad (4.24)$$

$$\Lambda_{21} = \frac{4s}{\pi}I - 8s^2q'I_{q'} - \frac{2s}{\pi}, \quad (4.25)$$

$$\Lambda_{22} = -2I + 1, \quad (4.26)$$

$$\Omega_2 = 16s^2\Big(\frac{3}{\pi^2}I - \frac{12sq'}{\pi}I_{q'} + 4s^2q'[q'I_{q'}]_{q'}\Big) - \frac{24s^2}{\pi^2}, \quad (4.27)$$

where s and q' are related via $q' = \exp(-\pi s)$.

One finally obtains

$$c = \frac{(\Omega_1\Lambda_{22} - \Omega_2\Lambda_{12})}{\Lambda_{11}\Lambda_{22} - \Lambda_{12}\Lambda_{21}} \quad (4.28)$$

and a similar expression for A, which is not of physical interest.

5. COMPARISON OF THE EXACT AND APPROXIMATE SOLUTIONS

The two series representations of the theta function are complementary with the Fourier series converging more rapidly for small q and the Gaussian series more rapidly for small q'. The worst possible case is when

$$q = q', \quad (5.1)$$

because then both converge equally well or badly. Consequently, we can limit our attention to this single, "worst case" value of q' which is

$$q' = e^{-\pi} = 0.0432. \quad (5.2)$$

The astonishingly small value of q' makes one feel like cheering. As noted earlier, the theta series converge much faster than an ordinary geometric series because the terms are proportional to q'^{n^2} rather than q'^n, but q' is so small that even the lowest-order approximation is very accurate. The approximate solution to $O(q'^2)$ is

$$u = -24s/\pi + 12\,\mathrm{sech}^2(sX)[1 - 8q'^2\cosh^2(sX) + 16q'^2\cosh^4(sX)] + O(q'^4), \quad (5.3)$$

$$c = -24s/\pi + 4s^2 - 96s^2q'^2, \quad (5.4)$$

where $q' = \exp(-\pi s)$ with

$$s = 1 \text{ at } q = q' = 0.0432. \quad (5.5)$$

The corresponding exact solution in terms of elliptic function is (at $s = 1$)

$$u = -3.81973 + 8.35918cn^2(1.1804X; m = \tfrac{1}{2}), \quad (5.6)$$

$$c = -3.81973, \quad (5.7)$$

where m is the modulus.[23] Although we have been mostly concerned with the Gaussian series, the Fourier series approximations are still sufficiently interesting to be included in our comparisons and are

$$u \doteq 96q\cos(2X) + 192q^2\cos(4X) + O(q^3), \quad (5.8)$$

$$c \doteq -4 + 96q^2, \quad (5.9)$$

where the first follows from (2.8) and the second is derived in Appendix A.

The lowest-order and second-lowest-order approximations to $u(x, t)$ are compared in Figs. 2 and 3. The agreement between the exact and approximate graphs is remarkable; the lowest-order approximations would be acceptable for most purposes and the second-order approximations are almost indistinguishable from the exact solution.

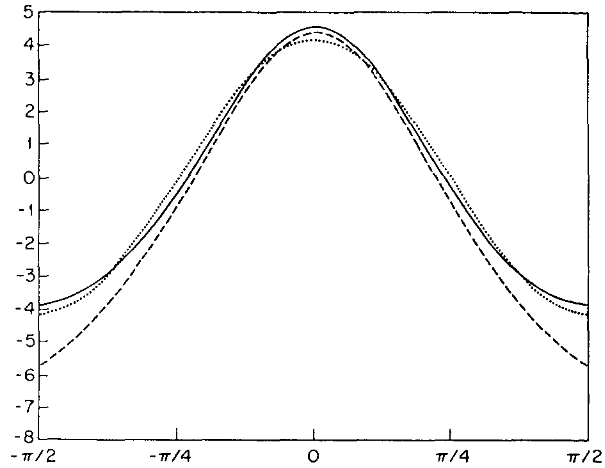

FIG. 2. A comparison of the lowest Fourier approximation (dotted line, a simple cosine function) and the lowest Gaussian series-derived approximation (dashed line, a constant plus the hyperbolic secant squared) with the exact cnoidal wave for $q = q' = 0.0432$.

Similar remarks apply to phase speeds. The lowest-order approximations and their errors are

$$c^{(1)}_{\text{Gaussian}} = -24/\pi + 4 = -3.639, \quad (5.10)$$

Absolute error = 0.181,

Relative error = 4.7%;

$$c^{(1)}_{\text{Fourier}} = -4, \quad (5.11)$$

Absolute error = − 0.180,

Relative error = − 4.7%;

the second-order approximations are

$$c^{(2)}_{\text{Gaussian}} = -3.81871, \quad (5.12)$$

Absolute error = 0.00102,

Relative error = 0.027%;

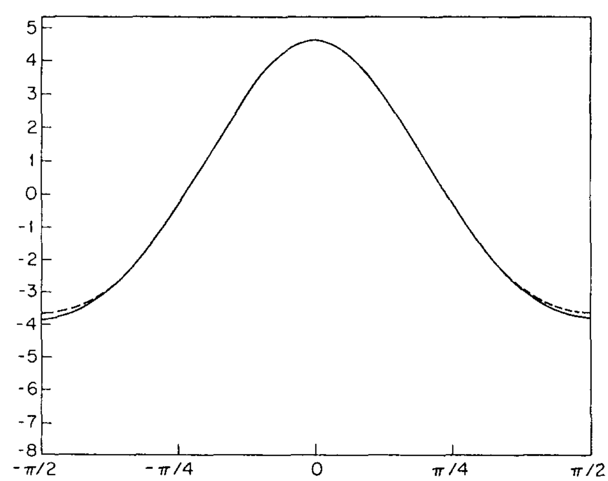

FIG. 3. A comparison of the second Gaussian approximation (5.3) with the exact cnoidal wave for the "worst case" $q = q' = 0.0432$. The second Fourier approximation is indistinguishable from the exact solution to within the thickness of the curve.

$$c^{(2)}_{\text{Fourier}} = -3.8207, \tag{5.13}$$
Absolute error $= -0.00100$,
Relative error $= -0.026\%$.

It is amusing that even in this weakly nonlinear regime—note that $c^{(1)}_{\text{Fourier}}$ is accurate to better than 5%—the solitary wave formula (representing the strongly nonlinear regime) is just as accurate. Thus, the two regimes strongly overlap. It is in this sense that one can say that the solitary wave is "almost a linear wave": it yields an excellent approximation even when q is small enough so that the lowest-order Fourier approximation—a simple cosine function—also gives an accurate approximation. In the same sense in reverse, one can say that the linear wave is "almost a soliton" in that it gives an accurate approximation when the nonlinearity is so strong the wave shape and speed are also accurately given by those of the solitary wave.

This strong overlapping of the linear and solitary wave regimes has two important implications. First, it suggests that perturbation theory will yield useful, understandable results for the "polycnoidal" wave also. Obviously, if one needed to carry the expansion to high order in N different parameters, perturbation theory would be pointless, and one would learn as much—or as little—with much less work by staring at films of numerical integrations of the KdV equations. The accuracy and overlap of the expansions for the ordinary cnoidal wave suggest that this will not be the case; suggest instead that the lowest- or second-lowest-order perturbation theory will be more than adequate.

The other implication is conceptual. The phrase "solitary wave" has the obvious connotation of a single, isolated wave peak. What has been shown here, however, is that a wave that to the eye looks like an ordinary linear cosine function—and as Fig. 2 shows, is well approximated by a cosine function—may nevertheless be accurately modeled by a solitary wave. The isolation of a wave or a wave peak from its fellows is not an essential ingredient either in the balance between nonlinearity and dispersion, which allows the solitary wave and cnoidal wave to exist as stable, permanent forms, or in the mathematical approximation of the wave by the characteristic sech^2 shape and speed of the soliton. Thus, the intuitive equivalence of "solitary" with "isolated" has been shown here to obscure the fact that such an isolated peak and a not-very-steep cnoidal wave are essentially the same thing.

6. HIROTA'S BILINEAR OPERATORS AND SOME THEOREMS ABOUT THEM

Hirota[4] showed that his transformed nonlinear evolution equations could always be expressed in terms of the bilinear operators defined by

$$D^n_x D^m_t(F\cdot G) \equiv \left[\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^n \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^m F(x,t)G(x',t')\right]_{\substack{x'=x\\ t'=t}} \tag{6.1}$$

where the notation indicates that x' and t' are to be replaced by x and t after the differentiations have been performed. The Hirota–Korteweg–de Vries equation, for example, is

$$(D^4_x + D_x D_t)(F\cdot F) = 2AF^2 \quad [\text{H–KdV}], \tag{6.2}$$

which is completely equivalent to (2.4).

Hirota proved a great many theorems and corollaries about the action of these bilinear operators on exponentials with linear arguments. In particular

$$D^n_x D^m_t(e^{(kx+wt)}\cdot e^{(k'x+w't)}) = (k-k')^n(w-w')^m e^{(k+k')x+(w+w')t}. \tag{6.3}$$

Notice that the result depends only on the *difference* between k and k' and between w and w'. When a theta function Fourier series such as

$$\theta_3 = \sum_{\substack{n=-\infty\\ [\text{integers}]}}^{\infty} q^{n^2}\exp[2\pi i n(kx+wt)] \tag{6.4}$$

is substituted into (6.2), the result is to generate a doubly infinite series of cross terms of the form of (6.3) so that the residual is

$$\bar\rho = \sum_{n=-\infty}^{\infty}\sum_{n'=-\infty}^{\infty} q^{n^2+n'^2}\bar\zeta(n-n';c,A)e^{2\pi i(n+n')X}, \tag{6.5}$$

with $X = kx + wt$, which is identical[24] in form to (4.2) except that it is a Fourier series instead of a Gaussian series. Hirota's theorem (6.3) gives ($w = -kc$)

$$\bar\zeta(n-n'c,A) = (2\pi ik)^4(n-n')^4 + (2\pi ik)(2\pi iw)(n-n')^2 - 2A, \tag{6.6}$$

which is a function only of the *difference* $n - n'$. As shown in Sec. 4, this property, that ζ (or $\bar\zeta$) is a function only of $n - n'$, is sufficient to prove that the vanishing of two (possibly nonlinear) equations in c and A is sufficient to give ρ (or $\bar\rho$)$\equiv 0$. Nakamura has shown that this generalizes to higher-dimensional theta functions, too—an N-dimensional Fourier series yields 2^N nonlinear equations in the phase speeds, A and the parameters of the theta function series that are sufficient to determine the whole solution. Furthermore, because the fact that $\bar\zeta$ is a function only of the difference $n - n'$ is a direct consequence of the fact that the bilinear operators yield results that depend only on $(k - k')$ and $(w - w')$ as shown in (6.3), it follows that the theta function solutions to *all* of Hirota's transformed equations expressed in terms of D_x and D_t must also be reducible to 2^N nonlinear algebraic equations in the parameters.

Our goal in this section is to prove a generalization of (6.3) which is applicable to Gaussian series. We shall find that, just as for Fourier series, the action of D_t and D_x on a pair of Gaussians depends only on the difference in their arguments. This in turn immediately implies that when an N-dimensional Gaussian theta function is substituted into any of Hirota's equations, the problem again reduces to 2^N nonlinear algebraic equations.

Before stating and proving the central result, it is important to note one powerful simplication: all terms in a given theta–Gaussian series have *identical second degree arguments* in the exponentials, and thus differ only in the terms in the exponentials which are *linear* in x and t. This is because all the terms in the theta series are Gaussians of *identical* shape, differing only in the location of the peak. Consequently, one can always factor out the common second-degree exponential as was done explicitly for the one-dimensional theta function in (2.9). The same will be done in the results below.

Theorem: Let D_x, D_t be the usual Hirota bilinear operators defined by (6.1) above, and let

$$F \equiv \exp[-(\alpha/2)x^2 - \beta xt - (\gamma/2)t^2]\exp(kx + wt), \tag{6.7}$$

$$G \equiv \exp[-(\alpha/2)x^2 - \beta xt - (\gamma/2)t^2]\exp(k'x + w't); \tag{6.8}$$

then

(i)

$$D_x^n(F\cdot G) = \alpha^{n/2} H_n\left[\frac{k-k'}{2\alpha^{1/2}}\right]FG, \tag{6.9}$$

where $H_n(y)$ is the usual Hermite polynomial. $D_t^n(F\cdot G)$ is given by (6.9) also with α replaced by γ and $(k-k')$ by $(w-w')$.

(ii)

$$\exp[\delta D_x](F\cdot G) = \exp[-\alpha\delta^2 + \delta(k-k')]FG. \tag{6.10}$$

(iii) Defining

$$Q_n^m \equiv D_x^n D_t^m(F\cdot G), \tag{6.11}$$

Q_n^m is determined by the recursion

$$Q_n^{m+1} = -2n\beta Q_{n-1}^m + (w-w')Q_n^m - 2\gamma m Q_n^{m-1}, \tag{6.12}$$

where the starting values are

$$Q_n^{-1} \equiv 0 \text{ for all } n, \tag{6.13}$$

$$Q_n^0 = \alpha^{n/2}H_n\left[\frac{k-k'}{2\alpha^{1/2}}\right] \qquad \text{[as given by (i)]}. \tag{6.14}$$

Proof: The demonstration of (i) is inductive. We use Q_n^m as defined in (6.11) except that we do *not* set $x' = x$, $t' = t$ until after obtaining the general recursion. We also temporarily drop the superscript, which is understood to be 0.

Define

$$Q_0 \equiv FG. \tag{6.15}$$

By explicit differentiation

$$Q_1 = D_x(F\cdot G) = [-\alpha(x-x') - \beta(t-t') + (k-k')]FG \tag{6.16}$$

To proceed to the next order, we can use the Leibnitz product-of-a-derivative rule after replacing FG by Q_0 to obtain

$$\begin{aligned} Q_2 &= D_x^2(F\cdot G) \\ &= [-\alpha(x-x') - \beta(t-t') + (k-k')]D_x(Q_0) \qquad (6.17) \\ &\quad + Q_0 D_x[-\alpha(x-x') - \beta(t-t') + (k-k')] \\ &= [-\alpha(x-x') - \beta(t-t') + (k-k')]Q_1 - 2\alpha Q_0. \end{aligned} \tag{6.18}$$

Let us now suppose that the recursion relation

$$Q_{n+1} = [-\alpha(x-x') - \beta(t-t') + (k-k')]Q_n - 2n\alpha Q_{n-1} \tag{6.19}$$

holds for a given n; we have already shown that it is true for $n = 1$. To demonstrate that it must also hold with $n \to n+1$, again apply D_x to (6.19) and invoke the Leibnitz rule to obtain

$$\begin{aligned} Q_{n+2} &= Q_n D_x[-\alpha(x-x') - \beta(t-t') + (k-k')] \\ &\quad + [-\alpha(x-x') - \beta(t-t') + (k-k')]D_x Q_n \\ &\quad - 2n\alpha D_x Q_{n-1} \qquad (6.20) \\ &= -2\alpha Q_n + [-\alpha(x-x') - \beta(t-t') \\ &\quad + (k-k')]Q_{n+1} - 2n\alpha Q_n, \end{aligned} \tag{6.21}$$

which is identical with (6.19) except for the replacement of n by $n+1$. Therefore, by induction, the recursion (6.19) must hold for all n. Setting $x = x'$ and $t = t'$ simplifies it to

$$Q_{n+1} = (k-k')Q_n - 2n\alpha Q_{n-1}. \tag{6.22}$$

The Hermite polynomials $H_n(y)$ satisfy the recursion relation

$$H_{n+1} = 2yH_n - 2nH_{n-1}. \tag{6.23}$$

It is trivial to show that

$$Q_n = \alpha^{n/2}H_n\left[\frac{k-k'}{2\alpha^{1/2}}\right] \tag{6.24}$$

satisfies (6.22) and the starting values (6.15) and (6.16) by directly substituting (6.24) into (6.22) and then employing the Hermite recursion (6.23) with the starting values $H_0 = 1$ and $H_1 = 2y$.

Part (ii) is derived by interpreting the exponential of an operator as the power series

$$\exp[\delta D_x] \equiv \sum_{n=0}^{\infty} \frac{\delta^n D_x^n}{n!}, \tag{6.25}$$

applying (i) term by term, and then equating the result to the generating function of the Hermite polynomials

$$e^{-t^2+2ty} = \sum_{n=0}^{\infty} \frac{t^n}{n!}H_n(y), \tag{6.26}$$

with $t = \delta\alpha^{1/2}$ and $y = (k-k')/2\alpha^{1/2}$.

An alternative proof can be obtained by using Hirota's result[4] that $\exp(\delta D_x)(F\cdot G) = F(x+\delta)G(x-\delta)$ for any F, G, and specializing it to the case when F and G are both Gaussians.

Part (iii) is proved by induction and use of the Liebnitz rule exactly as for (i) with the addition of the identity of $dH_n/dy = 2nH_{n-1}$, so details will be omitted. A simple, closed form solution for (6.12) has not been found; however, since Hirota's various evolution equations involve only first or second mixed derivatives, such a general solution is not really needed for the theory of polycnoidal waves.

7. MULTIDIMENSIONAL THETA FUNCTIONS

The general Riemann theta function of "reduced half integer characteristic" and dimension N is defined by

$$\begin{aligned} \theta\begin{bmatrix}\epsilon \\ \epsilon'\end{bmatrix}(\delta, \mathbf{T}) &= \sum_n \exp\left\{\pi i\left[\sum_{i=1}^N\sum_{j=1}^N T_{ij}\left(n_i + \frac{\epsilon_i}{2}\right)\right.\right. \\ &\quad \left.\left.\times\left(n_j + \frac{\epsilon_j}{2}\right) + 2\sum_{i=1}^N\left(n_i + \frac{\epsilon_i}{2}\right)\left(\zeta_i + \frac{\epsilon_i'}{2}\right)\right]\right\}. \end{aligned} \tag{7.1}$$

$\mathbf{T}$ is the $N \times N$ symmetric square matrix, called the "theta matrix," with positive definite *imaginary* part whose elements are written T_{ij}. ζ is the N-dimensional vector of dependent variables; in applications to polycnoidal waves

$$\zeta = \begin{pmatrix} \zeta_1 \\ \xi_2 \\ \vdots \\ \zeta_N \end{pmatrix} = \begin{pmatrix} k_1(x - c_1 t) + \phi_1 \\ k_2(x - c_2 t) + \phi_2 \\ \vdots \\ k_N(x - c_N t) + \phi_N \end{pmatrix}, \tag{7.2}$$

where the k_i are wave numbers, the c_i phase speeds, and the ϕ_i are constant phase factors. The quantity $[{\epsilon \atop \epsilon'}]$ consists of two N-dimensional row vectors written one above the other, where each element is either 1 or 0 and is known as the "characteristic" of the theta function,

$$\begin{bmatrix} \epsilon \\ \epsilon' \end{bmatrix} = \begin{pmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_N \\ \epsilon_1' & \epsilon_2' & \cdots & \epsilon_N' \end{pmatrix}. \tag{7.3}$$

Since there are a total of $2N$ matrix elements in the characteristic, each of which can independently take 2 values, there are a total of 4^N linearly independent theta functions with reduced half-integer characteristics. Note that it is conventional to define the multidimensional theta function so that it is periodic with period 2 versus the period of π that is conventional for the one-dimensional Jacobian theta functions θ_3 and θ_4.

Fortunately, it is always sufficient to take $\Theta\,[{0 \atop 0}](\zeta; \mathbf{T})$ as the solution of any of Hirota's differential equations, where $\mathbf{0}$ is the N-dimensional vector whose elements are 0. Hirota and Ito[15] have shown that when the Fourier series of $\Theta\,[{0 \atop 0}]$ $(\zeta, \mathbf{T})$ is substituted into one of the evolution equations, the result is a residual of the form

$$\tilde{\rho} = \sum_{\epsilon_1=0}^{1} \sum_{\epsilon_2=0}^{1} \cdots \sum_{\epsilon_N=0}^{1} R(\epsilon)\theta \begin{bmatrix} \epsilon \\ 0 \end{bmatrix}(2\zeta, 2\mathbf{T}). \tag{7.4}$$

There are a total of 2^N reduced half-integer theta functions with $\epsilon' \equiv 0$, and thus there are 2^N terms in (7.4) and 2^N nonlinear equations

$$R(\epsilon_1, \epsilon_2, ..., \epsilon_N) = 0, \quad \epsilon_i = 0 \text{ or } 1 \tag{7.5}$$

that must be solved to obtain the N phase speeds c_N, the constant of integration A in the H–KdV equation, and the $N(N-1)/2$ off-diagonal theta elements. (Recall $T_{ij} = T_{ji}$.) For $N \geqslant 3$, this gives more equations than unknowns; Hirota and Ito[15] have shown numerically that for $N = 3$, one of the eight equations (7.5) is redundant, and that one must solve seven equations in seven unknowns. Presumably something similar happens for large N although an analytical proof is lacking.

The diagonal elements play a role analogous to that of the nome for ordinary elliptic functions. In one dimension

$$q = e^{\pi i T_{11}}, \tag{7.6}$$

where T_{11} is positive imaginary. In more dimensions, one can define "nomes" via

$$q_j = e^{\pi i T_{jj}}, \quad j = 1, ..., N \tag{7.7}$$

and obtain perturbative solutions in the form of an N-dimensional power series in the q_j.

The "theta constants" are defined by

$$\theta \begin{bmatrix} \epsilon \\ \epsilon' \end{bmatrix}(T) \equiv \theta \begin{bmatrix} \epsilon \\ \epsilon' \end{bmatrix}(0, \mathbf{T}). \tag{7.8}$$

Hirota and Ito,[15] without calling attention to the fact, show that the residual equations (7.5) can be expressed entirely in terms of theta functions of the wave numbers $k_1, ..., k_N$ [when the Hirota equation involves $\exp(\delta D_x)$] and the theta constants $\theta\,[{\epsilon \atop 0}](2\mathrm{T})$ and their derivatives with respect to the *diagonal* matrix elements. This is analogous to Sec. 4 where the solution for the cnoidal was expressed in terms of the functions $H(q')$ and $I(q')$—which effectively *are* theta constants—and their first two derivatives with respect to q'. In general, if the highest power of $D_x^n D_t^m$ is $m + n = J$, the residuals $R(\epsilon)$ will involve differentiations with respect to T_{jj} of up to order $J/2$; note that only *even* values of $(m + n)$ occur in Hirota's equations.

The significance of this (besides the fact that it provides a simple and compact description of the residuals) is that it implies that one can apply Poisson summation directly to the residuals $R(\epsilon)$ to transform them into nonlinear equations in the complementary nomes q'_1, q'_2, etc.

The only flaw with this is that "theta-matrix doubling" occurs; the theta constants will appear in the residual equations $R(\epsilon)$ have double the theta matrix of the theta function that solves the H–KdV equation. In terms of the nomes, this means that the perturbation series for $c_1, ..., c_N$ and so on are functions of the *squares* of $q_1, q_2, ..., q_N$ rather than the nomes themselves. This implies very rapid convergence, of course, and is the reason that the first and second approximations to c in the one-dimensional case were seen in Sec. 5 to give such remarkable accuracy.

However, the Poisson summation for the theta function is given by

$$\theta \begin{bmatrix} 0 \\ 0 \end{bmatrix}(\zeta, \mathbf{T}) = \frac{\pi^{N/2}}{|\det T|^{1/2}}\, \theta \begin{bmatrix} 0 \\ 0 \end{bmatrix}(\mathbf{T}^{-1}\zeta, \mathbf{T}^{-1}), \tag{7.9}$$

which shows that doubling the theta matrix halves that of the Poisson sum. In other words, by applying Poisson summation directly to the residual equations $R(\epsilon)$, we pay for the rapidly-converging series in q^2 by obtaining slowly-convergent series in $(q')^{1/2}$ upon Poisson summation. Equations (4.20) and (4.21) show that exactly the same happens in reverse when we attempt to write theta constant series in q' directly in terms of those in q.

Thus, the best approach is to substitute *separately* the Fourier series and the Gaussian series into the H–KdV equations and its fellows, to obtain a series for c in q^2 from the Fourier series and another in q'^2 from the Gaussian series.

When performing this direct substitution for the Gaussian series, it is convenient to use the freedom to shift the phase factors in $\zeta_1, \zeta_2, ..., \zeta_N$ to replace $\theta\,[{0 \atop 0}](\zeta, \mathbf{T})$ by $\theta\,[{0 \atop 1}](\zeta, \mathbf{T}) = \theta\,[{0 \atop 0}](\zeta + \frac{1}{2}, \mathbf{T})$ because this has the simpler Poisson sum

$$\theta \begin{bmatrix} 0 \\ 1 \end{bmatrix}(\zeta, \mathbf{T}) = \frac{\pi^{N/2}}{|\det T|^{1/2}} \times \sum_n \exp\left\{ -\pi \sum_{i=1}^{N} \sum_{j=1}^{N} (\zeta_i + n_i)(\zeta_j + n_j) S_{ij} \right\}, \tag{7.10}$$

where the sum is taken over all possible half integers $\{-\infty, ..., -\frac{5}{2}, -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, ..., \infty\}$ in each of the N sum variables

$n_1, \ldots, n_N$, and where

$$S = I_m(\mathbf{T}^{-1}). \quad (7.11)$$

In one dimension, $\theta\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right](\zeta, \mathrm{T}) = \theta_4(x; q)$, which was what we used in Secs. 2–5.

Again, however, the substitution of the series reduces the problem—after using the theorems of Sec. 6—to solving 2^N nonlinear equations in $N^2/2 + N/2 + 1$ unknowns, implying some redundancy in the residual equations, just as when the theta Fourier series was used. Again, these equations can be solved perturbatively in the N complementary nome variables defined by

$$q_j' = e^{-\pi s_{jj}}. \quad (7.12)$$

The lowest 2^N terms in (7.10), those with $n_j = \pm 1/2$ for all j, generate the N-soliton solution of the KdV equation just as the bi-Gaussian ($N = 1$) was shown to generate the single soliton in Sec. 3. For $N = 2$, we have a "tetra-Gaussian" whose four peaks form a rectangle in the ζ_1-ζ_2 plane; for $N = 3$, an "octo-Gaussian" whose eight centers form the corners of a cube in ζ_1–ζ_2–ζ_3 space. And so it goes.

We close with two elementary theorems which simplify the calculations and have already been used above.

Theorem:

(i) If $U(\zeta_1, \zeta_2, \ldots, \zeta_N)$ is a solution of the KdV equation where $\zeta_j = k_j(x - c_j t) + \phi_j$ as in (7.2), then

$$V(x, t) = p + U(\bar{\zeta}_1, \bar{\zeta}_2, \ldots, \bar{\zeta}_N) \quad (7.13)$$

is also a solution of the KdV equation provided

$$\bar{\zeta}_j(k_j, \phi_j, c_j) = \zeta_j(k_j, \phi_j, c_j + p) \quad (7.14)$$

for all j where p is an arbitrary constant.

(ii) If $u(x, t)$ is a solution of the KdV equation, then

$$\tilde{v}(x, t) \equiv \lambda^2 u(\lambda x, \lambda^3 t) \quad (7.15)$$

is also a solution.

The proofs are elementary and are not given here. The first theorem allows one to choose "mean sea level," i.e., $\int_{-\pi/2}^{\pi/2} u(x, t)\, dx$, to be whatever one wishes. Note that the theta-function solution normally picks its own "mean sea level," which is generally different from zero.

The second theorem allows us to take the periodicity interval to be π or 2 or whatever is convenient. It also permits us to set one of the wave numbers $k_1 = 1$ without loss of generality.

8. CONCLUSIONS AND SUMMARY

The theory of polycnoidal waves for the Korteweg–de Vries and other evolution equations is built upon four fundamental ideas. The first was the recognition by Lax and Novikov that the KdV equation had a class of generalized cnoidal wave solutions, here dubbed polycnoidal waves, that could be used to approximate an arbitrary, spatially-periodic initial condition and that could be formally calculated from the spectrum of Hill's equation. The second was the independent discovery by Akira Nakamura and the author that Hirota's "direct method" was just as useful for polycnoidal waves as for the multiple-soliton solutions of which they are generalizations. Nakamura, Hirota, and Ito subsequently refined the "direct method" using Fourier series to a high art.

The third is the discovery, first presented here, that the direct method can also be applied using the alternative Gaussian series for the theta functions to make it possible to explore strongly nonlinear polycnoidal waves and their relationship with multiple solitary waves.

The fourth, also presented here for the first time, is that by using these two different series—the theta Fourier expansion and the theta Gaussian expansion—in a complementary way, the former for small amplitude and the latter for large amplitude, one can obtain rapidly convergent perturbation series—a couple of terms are enough—to approximate the polycnoidal waves over the whole range of parameter space.

There are some additional complications, such as the way soliton phase shifts enter the formalism, which arise for polycnoidal waves of dimension $N > 2$. Since this present work is already lengthy, the actual polycnoidal wave calculations will be described in a later work. Here, however, the full mathematical machinery to perform these calculations has been presented with a very thorough discussion of the ordinary cnoidal wave to illustrate both how to use the formalism and also why perturbation theory should be useful even for the more complicated cases. As we saw in Sec. 5, just two terms of the perturbation series in q (derived from the theta Fourier series) and two terms of the series in q' (derived from the theta Gaussian series) were sufficient to give the cnoidal wave phase speed c to within a relative error of 0.027% for all possible values of q and q'.

The path for future research is to simply follow up this initial success by explicit calculations for $N > 2$, concentrating particularly on the Gaussian series approach that is the central theme of this work. For the Fourier series results of Nakamura and Hirota can be obtained another way as explained in Appendix B—an even more direct method than Hirota's. However, at present there seems to be no alternative to the theta Gaussian series for exploring strongly nonlinear polycnoidal waves and their close relationship with multiple solitons.

ACKNOWLEDGMENT

This work was supported by NSF Grant OCE-79-09191.

APPENDIX A: THE EXACT SOLUTION FOR THE CNOIDAL WAVE

By using identities 17.2.13 and 17.2.11 of the NBS Handbook,[23] one can show that

$$12[\ln \theta_4]_{XX} = \delta + \epsilon\, cn^2(2K(m)X/\pi; m), \quad (A1)$$

where $K(m)$ is the complete elliptic integral with m related to q by

$$q = e^{-\pi K(1-m)/K(m)} \quad (A2)$$

and with the phase speed c given by ($X = x - ct$)

$$c = \delta + [(2m - 1)/3m]\epsilon, \quad (A3)$$

with

$$\delta = [-48K(m)/\pi^2][(m - 1)K(m) + E(m)], \quad (A4)$$

$$\epsilon = -48mK^2(m)/\pi^2. \quad (A5)$$

The phase speed is obtained by substituting the right-hand side of (A1) directly into the KdV equation and using elliptic function identities.[23] For small q, one has the approximations

$$m = 16q(1 - 8q + 44q^2), \quad \text{(A6)}$$

$$K(m(q)) \doteq (\pi/2)(1 + 4q + 4q^2), \quad \text{(A7)}$$

$$E(m(q)) \doteq (\pi/2)(1 - 4q + 20q^2). \quad \text{(A8)}$$

Substituting these small q approximations into (A4) and (A5) gives the approximation for c given by (5.9). The corresponding approximation for large q is derived directly by solving the residual equations in Sec. 4, and then Taylor-expanding (4.28), which is actually a rational function of q'.

When carrying out the analysis in terms of theta functions, it is convenient to take q or q' as the perturbation parameter. In analyzing observations or laboratory experiments, one would probably take ϵ as the fundamental quantity since this is what is most easily measured. (It is the difference between the peak and the trough.) In applying multiple-scales perturbation theory to the KdV equation as in Appendix B, one would normally fix the Fourier coefficient of $\cos(2nx)$ (the linear wave) at a certain value a and use that as the perturbation parameter.

Fortunately, it is easy to relate these different measures of the nonlinearity of the wave to each other by using the Fourier series for $u(x, t) = 12(\ln \theta_4)_{xx}$ given earlier:

$$u(x, t) = 96q \sum_{n=1}^{\infty} \frac{nq^{n-1}}{1 - q^{2n}} \cos(2nx). \quad \text{(A9)}$$

The coefficient of $\cos(2x)$ is

$$a = 96q/(1 - q^2), \quad \text{(A10)}$$

which is trivially solved to give q as a function of a. Since $\epsilon = u(0, 0) - u(\pi/2, 0)$, we can evaluate (A9) at these values of x and subtract to obtain

$$\epsilon = 192q \sum_{n=0}^{\infty} \frac{(2n+1)q^{2n}}{1 - q^{2(2n+1)}}, \quad \text{(A11)}$$

which can be easily reverted term-by-term to give a series expansion for $q(\epsilon)$.

APPENDIX B: THE METHOD OF MULTIPLE SCALES AND POLYCNOIDAL WAVES

In the near linear regime, one can bypass the use of the Fourier series for the theta functions by using a much more general technique[16,17] known variously as the "method of strained parameters" or "method of multiple scales," which can be applied directly to Korteweg–de Vries equation. The amplitude of the lowest harmonic, a, is assumed to be a small parameter. It is further assumed that (i) the wave is steadily translating at a phase speed c and (ii) that $u(x - ct)$ can be expanded as a power series in a. One can then substitute the power series into the differential equation, match powers of a and solve the perturbation equations order-by-order. However, there is one modest complication: the phase speed c is usually altered by the nonlinearity, so it is necessary to assume c can also be expanded in a power series in a. The technique is very similar to the usual Rayleigh–Schrödinger perturbation theory of quantum mechanics with c playing the role of the eigenvalue. A full description with many, many examples and problems is given in the texts by Nayfeh[16] and Bender and Orszag.[17] It is not exactly a new idea; Stokes applied it to water waves in 1847.

Because this algorithm is so straightforward, it is not only easy to do by hand but also simple to program for a computer. Using the algebraic manipulation language REDUCE 2, which can explicitly multiply and differentiate Fourier series and manipulate trigonometric identities, a short program was written by the author to calculate single and double cnoidal waves. For the expenditure of \$1.50 (about 8 sec of CPU time on the University of Michigan Amdahl), the following was obtained to fifth order for the single cnoidal wave ($X = x - ct$):

$$c = -4 + a^2/96 + a^4/884\,736, \quad \text{(B1)}$$

$$\begin{aligned} u(x, t) = {} & a\cos(2X) + (a^2/48 - a^4/221\,184)\cos(4X) \\ & + (a^3/3072 - a^5/9\,437\,184)\cos(6x) \\ & + (a^4/221\,184)\cos(8X) \\ & + (5a^5/84\,934\,656)\cos(10X). \end{aligned} \quad \text{(B2)}$$

By Taylor-expanding the relationship between a and the elliptic function nome q, which was shown in Appendix A to be

$$a = 96q/(1 - q^2) = 96q(1 + q^2 + q^4 + \cdots), \quad \text{(B3)}$$

one can recast the expansion in powers of q and thus duplicate the results of the "direct Fourier series" method of Nakamura, Hirota, and others:

$$c = -4 + 96q^2 + 288q^4 + \cdots, \quad \text{(B4)}$$

$$\begin{aligned} u(x, t) = 96q[& (1 + q^2 + q^4)\cos(2x) \\ & + 2q\cos(4x) + 3q^2\cos(6x) \\ & + 4q^3\cos(8x) + 5q^4\cos(10x)]. \end{aligned} \quad \text{(B5)}$$

In the body of this paper, only terms through q^2 were kept because these give more than enough accuracy; the expansions have been carried to higher order here simply to make the point that it is easy to calculate Fourier series via the method of multiple scales and that a theta function approach is not really needed except in the opposite near-soliton regime of strong nonlinearity.

In a similar way, one can calculate double cnoidal waves by starting with the lowest order approximation

$$u(x, t) \doteq a(\cos(2X) + b\cos(MY)), \quad \text{(B6)}$$

where again a is a small parameter and b is $O(1)$, with

$$X = x - c_1 t + \phi_1, \quad \text{(B7)}$$

$$Y = x - c_2 t + \phi_2. \quad \text{(B8)}$$

M may take on any constant value; however, unless $M/2$ is a rational number, the wave will be "almost periodic" in x in the formal mathematical sense of the word as opposed to truly periodic. But Novikov[10] and Dubrovin *et al.*[11] have emphasized that the polycnoidal wave may indeed be almost periodic in space.

The most reasonable value of M, however, is $M = 4$ so that the second component is the second harmonic of the first. Unless the initial condition is rather peculiar, the second harmonic is usually the largest Fourier component after the fundamental. Consequently, in applying the double cnoidal wave to model events in a laboratory tank or the real world, $M = 4$ is the case in which one would be chiefly interested. For simplicity, the results below are therefore confined to this case. The computer program, however, can solve the double cnoidal wave with M as a purely symbolic parameter with no numerical value specified.

One obtains

$$c_1 = -4 + a^2/96, \tag{B9}$$

$$c_2 = -16 + a^2 b^2/384, \tag{B10}$$

$$\begin{aligned} u(x,t) = {} & a[\cos(2X) + b\cos(4Y)] \\ & + a^2[(1/48)\cos(4X) + (b^2/192)\cos(8Y) \\ & + (b/48)\cos(2X+4Y) - (b/48)\cos(2X-4Y)] \\ & + a^3[(1/3072)\cos(6X) + (b^3/49152)\cos(12Y) \\ & + (b/2592)\cos(4X+4Y) \\ & + (25b^2/165888)\cos(2X+8Y) \\ & - (b^2/2048)\cos(2X-8Y)]. \end{aligned} \tag{B11}$$

This run also cost a mere \$1.50. It is trivial to generalize the algorithm to triple and higher cnoidal waves; in point of fact, the same program with changes in only two statements was used to compute both the single and double cnoidal wave results given above.

The message is clear. The theta Fourier series is not essential in understanding spatially periodic solutions of the KdV and other evolution equations; it is the theta Gaussian series that contains treasure.

[1]D. J. Korteweg and G. de Vries, Philos. Mag. **39**, 422 (1895).
[2]J. Scott-Russell, Proc. Roy. Soc. Edinburgh **319** (1844).
[3]C. S. Gardner, J. M Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[4]R. Hirota, Prog. Theor. Phys. **52**, 1498 (1974).
[5]Ryogo Hirota, Phys. Rev. Lett. **27**, 1192 (1971).
[6]J. P. Boyd, J. Phys. Oceangr. **10**, 1699 (1980).
[7]Y.-C. Ma and L. G. Redekopp, Phys. Fluids **22**, 1872 (1979).
[8]Y.-C. Ma, Stud. Appl. Math. 59, **201** (1978).
[9]P. D. Lax, Commun. Pure Appl. Math. **28**, 141 (1975); H. P. McKean and P. van Moerbeke, Inv. Math **30**, 217 (1975).
[10]S. P. Novikov, Funct. Anal. Appl. **8**, 236 (1974).
[11]B. A. Dubrovin, V. B. Matreev, and S. P. Novikov, Russian Math. Surveys **31**, 59 (1976).
[12]Akira Nakamura, J. Phys. Soc. Japan **47**, 1701 (1979).
[13]Akira Nakamura, J. Phys. Soc. Japan **48**, 1365 (1980).
[14]Akira Nakamura and Yoshimasa Matsuno, J. Phys. Soc. Japan **48**, 653 (1980).
[15]Ryogo Hirota and Masaaki Ito, J. Phys. Soc. Japan **50**, 338 (1981).
[16]A. H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973).
[17]C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).
[18]Richard Bellman, *A Brief Introduction to Theta Functions* (Holt, Rinehart, and Winston, New York, 1961), p. 45.

[19]The Fourier series of a singular function which is analytic on the real axis has a finite strip of convergence in the complex plane $\{\text{all } x;\ -\tau < y < \tau\}$, where $z = x + iy$ and the singularity of $f(z)$ nearest the real axis lies on $y = \pm\tau$. Since $|\cos(nz)|$ is bounded by $\exp(ny)$, it follows that the Fourier series can converge inside the strip bounded by the lines on which $f(z)$ is singular and diverge outside it only if the Fourier coefficients a_n satisfy the inequality $\exp[-n(\tau+\epsilon)] < |a_n| < \exp[-n(\tau-\epsilon)]$. Subtracting off the poles on $y = \pm\tau$, using the known Fourier expansion for $1/(a + \cos z)$ and noting that $f(z) - 1/(a + \cos z)$ has a larger strip of convergence than $f(z)$ can be used to establish that $a_n \sim O(q^n)$ precisely, where $q = \exp(-\tau)$.

[20]Richard Bellman, Ref. 18, pp. 7–12; Philip M. Morse and Herman Feshbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953), pp. 466–467.

[21]Strictly speaking, $\Theta(X)$ must be multiplied by an X-independent constant to write it as the sum of two Gaussians. However, a constant times a solution of the H-KdV equation is also a solution, so this multiplicative constant has no physical significance.
[22]This is a corollary of a theorem which will be proved in Sec. 6 although it has been known to soliton specialists for years.

[23]Milton Abramowitz and Irene Stegun, Eds., *Handbook of Mathematical Functions* (Dover, New York, 1965).

[24]For notational simplicity, θ_3 is used here instead of θ_4, and the Fourier sum must be taken over the integers instead of the half-integers. These minor differences are irrelevant to the analogy between (4.2) and (6.5), however.