10.1098/rspa.2001.0862

THE ROYAL SOCIETY

On the semi-classical limit for the focusing nonlinear Schrödinger equation: sensitivity to analytic properties of the initial data

By S. R. Clarke and P. D. Miller

Department of Mathematics and Statistics, Monash University, Victoria 3800, Australia

Received 23 August 2000; accepted 18 June 2001; published online 30 November 2001

We present the results of a large number of careful numerical experiments carried out to investigate the way that the solution of the integrable focusing nonlinear Schrödinger equation with fixed initial data, when taken to be close to the semi-classical limit, depends on analyticity properties of the data. In particular, we study a family of initial data that have complex singularities at an adjustable distance from the real axis. We also make use of a simple relation that provides the exact solution, at the centre of the wave field, of the elliptic quasilinear system that appears formally as a leading-order model for the semi-classical dynamics. Among other things, we conclude that the semi-classical limit cannot be expected to be continuous with respect to the initial data, even for real analytic data, if there are certain complex singularities present. We argue that, in order to have well-posedness of the semi-classical limit, the correct setting is a physically relevant space of functions with compactly supported Fourier transforms.

Keywords: focusing nonlinear Schrödinger equation; semiclassical limit; non-analyticity

1. Introduction

The initial-value problem for the one-dimensional focusing cubic nonlinear Schrödinger equation

$$\mathrm{i}\hbar\psi_t + \tfrac{1}{2}\hbar^2\psi_{xx} + |\psi|^2\psi = 0, \qquad \psi(x,0) = A(x)\exp(\mathrm{i}S(x)/\hbar), \tag{1.1}$$

where $A(x)$ and $S(x)$ are real-valued amplitude and phase functions with $A(x)$ and $S'(x)$ decaying rapidly as $x \to \pm\infty$, is known to arise in many physical contexts. Most generally, (1.1) emerges from multiscale perturbation theory as the amplitude equation describing the evolution of nearly monochromatic wave packets propagating in conservative one-dimensional systems in the presence of dispersion and weak nonlinearity. The parameter $\hbar$ is a dimensionless measure of the ratio of dispersive effects to nonlinear effects.† The focusing version (1.1) of the nonlinear Schrödinger equation is obtained when the combined effects of dispersion and nonlinearity tend

† This symbol is chosen purely by analogy with the quantum-mechanical interpretation of Schrödinger equations. In this paper, $\bar{}$ is meant to be dimensionless and does not generally refer to Planck's constant.

to destabilize periodic wavetrains. The corresponding instability of (1.1) is the *modulational instability*. This occurs, for example, in the theory of gravity waves on water, and also in the theory of Langmuir waves in plasmas. In recent years, considerable interest originating in the telecommunications industry has been directed toward the problem (1.1) due to its multiple-scales derivation in the modelling of propagation of intense optical pulses in fibres.

On the other hand, the initial-value problem (1.1) also appears occasionally in physics in a non-perturbative way, without the need of appealing to wave packets and multiscale asymptotics. For example (Antanovskii *et al.* 1997), the equivalent system (1.3) to be given below describes the non-equilibrium thermodynamics of a gas with density ρ and momentum distribution μ that

(i) is in a parameter regime where the gas pressure decreases with increasing density, and

(ii) has a free energy function containing the simplest term modelling isotropic dependence on local gradients of the gas density ρ.

The former feature is typical for van der Waals gases supercooled below the critical (condensation) temperature, and the latter feature has been proposed as a phenomenological model for capillarity (Antanovskii 1996). The parameter $\hbar$ is, here, a dimensionless measure of the surface tension.

We are particularly interested in the asymptotic behaviour of solutions of (1.1) for small $\hbar$, particularly in the neighbourhood of fixed values of x and t. We refer to the limit $\hbar \downarrow 0$ as the *semi-classical limit*. Now for $\hbar$ small, it is known that solutions of (1.1) develop features on typical dimensionless length and time-scales of size $O(\hbar)$, and these scales are arbitrarily small as $\hbar$ tends to zero. While this may at first appear to be problematic in those applications when (1.1) is obtained as an envelope equation, in fact semi-classical asymptotics for (1.1) can comfortably coexist with the validity of the multiscale asymptotics leading to (1.1) in certain physically relevant parameter regimes. For example, in a dramatic laboratory experiment carried out by Sudo *et al.* (1989), optical pulses of 100 ps initial duration (this is the x-variable) were launched into a dispersion-shifted optical fibre for which one expects nonlinear processes to dominate the effects of group velocity dispersion for short propagation distances. The idea was to allow the modulational instability to break up each launched pulse into many pulses each of much shorter duration. In this way, a dispersion-shifted optical fibre could conceivably be used as a source of ultrashort pulses for applications in terahertz telecommunications systems. Pulses sufficiently short for such applications cannot be generated using electronic switching technology, which is too slow for fundamental physical reasons. Therefore an optical process is required. In the experiment reported in Sudo *et al.* (1989), the 100 ps electronically generated pulses were observed to develop features on scales of duration 200 fs after propagation in the fibre. This ratio of scales (micro-to-macro) suggests the extremely small value of $\hbar \approx 0.002$ in (1.1). But the carrier wavelength in these experiments was 1319 nm, giving rise to an optical period of 4.41 fs, so that even after the pulse break-up, each small-scale (200 fs) variation of the envelope contains *ca.* 50 optical periods. This is striking experimental evidence of the validity of (1.1) as an envelope equation in a strongly 'semi-classical' context where $\hbar \approx 0.002$ in (1.1).

One way to address the semi-classical limit is to introduce the variables

$$\rho := |\psi|^2, \qquad \mu := \hbar \operatorname{Im}(\psi^* \psi_x), \tag{1.2}$$

in terms of which one finds that (1.1) becomes exactly

$$\left.\begin{aligned} \rho_t + \mu_x &= 0, \\ \mu_t + \left(\frac{\mu^2}{\rho} - \frac{\rho^2}{2}\right)_x &= \frac{\hbar^2}{4}((\log\rho)_{xx}\rho)_x, \end{aligned}\right\} \tag{1.3}$$

with initial data $\rho(x,0) = A(x)^2$ and $\mu(x,0) = A(x)^2 S'(x)$. It now appears more attractive to set $\hbar = 0$ in this equivalent system, since there would remain terms that could balance, and because the initial data are now generally independent of $\hbar$. Thus, one feels justified in proposing the system of equations

$$\left.\begin{aligned} \rho_{\circ t} + \mu_{\circ x} &= 0, \\ \mu_{\circ t} + \left(\frac{{\mu_\circ}^2}{\rho_\circ} - \frac{{\rho_\circ}^2}{2}\right)_x &= 0, \end{aligned}\right\} \tag{1.4}$$

with initial data $\rho_\circ(x,0) = A(x)^2$ and $\mu_\circ(x,0) = A(x)^2 S'(x)$ as a model for the asymptotic dynamics of (1.1), at least for sufficiently small t independent of $\hbar$.

The difficulty with this reasoning is that the quasilinear system of equations (1.4) is elliptic rather than hyperbolic, and consequently the Cauchy problem posed at $t = 0$ is an ill-posed problem. Solutions to the initial-value problem for (1.4) can only exist for analytic initial data. So the semi-classical asymptotics for (1.1) would appear to distinguish between analytic initial data and (merely) infinitely differentiable data. This information alone is somewhat distressing from the point of view of physical applications, where 'noisy' data are typical. More careful analysis is clearly required.

Using the tool of the inverse-scattering transform, which Zakharov & Shabat (1972) showed exactly linearizes (1.1), some recent progress has been made in the semi-classical analysis of (1.1). The first step in the analysis is to obtain the discrete eigenvalues, associated auxiliary spectrum, and reflection coefficient for a certain non-self-adjoint linear operator that explicitly encodes the initial data $A(x)$ and $S(x)$, and depends explicitly on $\hbar$ in a singular way. For semi-classical asymptotics, one needs corresponding asymptotic information about the spectrum. In a careful numerical study, Bronski (1996) has shown that for $S'(x)$ not identically zero the discrete eigenvalues accumulate as $\hbar \downarrow 0$ on curves in the complex plane that depend on the two functions $A(x)$ and $S(x)$ in the initial data. These curves were later explained by Miller (2001) using turning point (WKB) theory. These results concerned the distribution of eigenvalues for fixed initial data, but Bronski (2001) later studied numerically the effect on the spectral curves of small non-analytic perturbations of the initial data, and found surprising sensitivity of the curve shapes for small $\hbar$. This sensitivity of the spectrum would seem to suggest similar sensitivity in the solution of (1.1). This sensitivity has indeed been observed in sufficiently resolved numerical simulations (Bronski & Kutz 1999).

For $S'(x) \equiv 0$ (in which case by a gauge transformation one may simply assume $S(x) \equiv 0$), formal WKB analysis predicts more stable properties of the spectrum. Assuming that $A(x)$ is bell shaped, one finds (Ercolani *et al.* 1993) that the WKB approximations for the eigenvalues all lie on the imaginary axis between $-\mathrm{i}A$ and $\mathrm{i}A$, where A is the maximum value of $A(x)$. On the other hand, the formal model initial-value problem (1.4) with $\mu_\circ(x,0) \equiv 0$ remains ill posed and still requires analyticity of $\rho_\circ(x,0) = A(x)^2$ for the existence of a solution. Therefore one expects the

semi-classical asymptotics for the 'slow phase' case of $S(x) \equiv 0$ to display sensitivity to analytic properties of the function $A(x)$ similar to what is expected for the 'fast-phase' case, but the inverse-scattering theory is somewhat simpler because the eigenvalues are more confined.

In at least one case with $S(x) \equiv 0$, the spectrum of the Zakharov–Shabat operator is known exactly. This is the family $A(x) = A\,\mathrm{sech}(x)$ for which Satsuma & Yajima (1974) have computed the exact spectrum for all $\hbar$. With the use of this explicit spectral data, exact solutions of (1.1) were constructed corresponding to values of $\hbar$ as small as 0.1 by Miller & Kamvissis (1998) (corresponding to 20 soliton eigenvalues) and the resulting pictures gave good evidence that ρ and μ converged strongly as $\hbar \downarrow 0$ to $\rho_\circ$ and $\mu_\circ$ satisfying (1.4) for $|t| < t^{(1)}(x)$, where $t^{(1)}(x)$ is a certain $\hbar$-independent phase-transition boundary, the 'primary caustic'. For times after the primary caustic, the fields ρ and μ become oscillatory, but analysis of the exact solutions provided good evidence that ρ and μ continued to converge in the weak sense, although not to any solution of (1.4). The oscillatory behaviour after the primary caustic displays a regular spatio-temporal pattern of maxima, and this pattern was observed to break down again at a second curve $|t| = t^{(2)}(x)$, a 'secondary caustic'. These reconstructions did not rule out the possibility of the existence of further, 'higher-order' caustics.

Many of these observations for $S(x) \equiv 0$ and $A(x) = A\,\mathrm{sech}(x)$ have been recently made rigorous by Kamvissis *et al.* (2002) by applying careful asymptotic analysis to the matrix Riemann–Hilbert factorization problem, whose solution comprises the inverse-scattering step in the solution of (1.1). The starting point for the analysis in Kamvissis *et al.* (2002) is a set of purely reflectionless scattering data with certain asymptotic properties that are consistent with formal WKB theory in the slow phase case of $S(x) \equiv 0$ with $A(x)$ bell shaped but arbitrary. This ensemble of eigenvalues is the *exact* scattering data for the Satsuma–Yajima case of $A(x) = A\,\mathrm{sech}(x)$, in which case the rigorous asymptotics for the initial-value problem (1.1) follow. In the case of more general $A(x)$, the exact correspondence between the spectrum of the 'true' initial data and the WKB spectrum is lost, and while the asymptotics for the inverse problem are rigorous, there is a gap in the argument that they provide the semi-classical limit for (1.1). Nonetheless, the inverse theory applied to the WKB spectra for $S(x) \equiv 0$ and $A(x)$ with the following properties:† (i) $A(x)$ is bell shaped and even in x, (ii) $A(x)$ is real analytic, (iii) $A''(0)$ is strictly negative, and (iv) $A(x)$ decays exponentially for large $|x|$, does in fact provide an implicit algebraic representation of the solution of the model problem (1.4) corresponding to the initial data $\mu_\circ(x, 0) \equiv 0$

† The technical reason for these conditions is that together they imply that the WKB formula for the density of states in the imaginary interval $[0, \mathrm{i}A(0)]$ is the restriction of a function analytic and bounded in some 'inverted teardrop'-shaped neighbourhood of this interval with the corner of the teardrop at the origin. This property is essential for the analytic deformations of the Riemann–Hilbert problem exploited in Kamvissis *et al.* (2002). For $S(x) \equiv 0$, the formal WKB theory is the same as for the stationary Schrödinger operator with potential $V(x) = -A(x)^2$ and negative energy E equal to the square of the imaginary eigenvalue for the Zakharov–Shabat operator. It follows that the width of the potential well must be an analytic function of the bound state energy for the density of states to be analytic in the interior of the range of energies. This explains the analyticity condition on $A(x)$. Also, if the potential well is too flat at the bottom or top, then the density of states will be infinite at the corresponding extreme values of the bound state energy. The convexity condition taken together with analyticity ensures that the WKB density is analytic at the bottom of the well corresponding to the point $\mathrm{i}A(0)$, and the decay condition on $A(x)$ ensures that the WKB density is finite at the top of the well corresponding to the origin. This explains the convexity and decay conditions on $A(x)$.

and $\rho_\circ(x,0) = A(x)^2$. We will give evidence in this paper that the condition of analyticity is essential, while at least the condition of strict convexity at the maximum is merely technical.

This paper will be concerned wholly with the slow phase case of $S(x) \equiv 0$. We have carried out a large number of numerical simulations of the initial-value problem (1.1) in order to investigate the sensitivity of the semi-classical behaviour with respect to the analyticity properties of the function $A(x)$. These simulations provide evidence that for analytic functions $A(x)$ there exists a finite time-interval, independent of $\hbar$, in which the solution to the system (1.3) equivalent to (1.1) is close to the solution of the ill-posed model problem (1.4). The length of this time-interval depends on 'how analytic' the function $A(x)$ is. If in a certain limit $A(x)$ becomes non-analytic at some (real) point $x = x_0$, the primary caustic curve becomes attached to this point at $t = 0$, i.e. $t^{(1)}(x_0) = 0$, and there is no time-interval at all in which the solution of (1.3) is approximated by any solution of (1.4) uniformly for all x. On the other hand, there are disjoint regions to either side of the primary caustic where ρ and μ appear to be smooth and therefore should be well approximated by *some* solutions of the elliptic system (1.4).

Where possible, we compare our numerical data with exact solutions of the ill-posed initial-value problem for the elliptic system (1.4). These solutions are obtained generally by solving a simultaneous system of algebraic equations in which x and t appear as explicit parameters (Kamvissis *et al.* 2002). Below we will use a simple reduction of the exact solution valid under the constraint $x = 0$ to estimate the breaking time. Not only does this analysis provide evidence that solutions of (1.1) for finite $\hbar$ approach (strongly in the variables ρ and μ in some finite time-interval) corresponding solutions of the elliptic system (1.4) as $\hbar$ tends to zero, but it will also lead us to the conclusion that the semi-classical limit of (1.1) does not behave continuously with respect to the initial data, even if those data are taken to lie in a class of functions uniformly analytic in some strip about the real axis.

2. Numerical integration and data analysis

(*a*) *Numerical integration scheme*

The integrations of (1.1) considered here are all undertaken using a modification of the Fourier split-step method for nonlinear Schrödinger equations. A general description of this method can be found in Newell & Moloney (1992). Since all the initial conditions considered are even and there is no symmetry breaking in the governing equation, the implementation of the Fourier method uses cosine transforms rather than full Fourier transforms. This has two advantages. First, the computational cost is halved, which is a not an insignificant consideration for these extremely stiff computations. Second, the numerical growth of odd modes due to round-off errors is completely eliminated. All the numerical integrations were performed to 32 significant figures of accuracy (quadruple precision on an SGI R10000). This was found to be necessary to ensure accurate computations at $\hbar$ values down to 0.025. For example, for simulations with the previously discussed initial condition $A(x) = 2\,\mathrm{sech}(x)$, it was found that, independent of spatial and temporal resolution, significant errors in the position of the primary caustic occurred when computations were performed in double precision. As $\hbar$ is increased, this loss of accuracy is manifested at larger

times; for $\hbar = 0.05$ it was found that computations in double precision could not be used to accurately determine the position of the secondary caustic.

Two factors determine the number of collocation points or Fourier modes used for these computations. First, the microstructure must be accurately resolved. The size of the microstructure is $O(\hbar)$, and so, as $\hbar$ is decreased, the number of modes must be increased. More important here is the need to accurately resolve the initial conditions in the limit of non-analyticity. Truly non-analytic initial conditions cannot really be considered with any discrete numerical scheme, as any initial condition defined on some discrete grid can be represented by an analytic Fourier series. Rather, we must consider a sequence of initial conditions tending toward the limit of non-analyticity. Thus the determining factor for spatial resolution here is that for the smallest value of $\hbar$ considered the initial condition is sufficiently resolved, such that no distinguishable difference in the numerical solution occurs as the resolution is decreased further. The computations presented were all performed on the region $x \in [0, 4]$ (solutions for negative x were simply obtained by reflection), and for $\hbar = 0.025$ it was found that 4097 cosine modes or collocation points were necessary to accurately resolve the solution. This gives a spatial resolution of $O(10^{-3})$, which is an order of magnitude smaller than any microstructure. To ensure temporal accuracy and stability it was found that time-steps of length $O(10^{-6})$ were necessary for this spatial resolution. The accuracy of the solution was monitored by calculating the energy (square of the $L^2(\mathbb{R})$ norm) and Hamiltonian of the solution, defined, respectively, as

$$E = \int_{-\infty}^{\infty} |\psi|^2 \, \mathrm{d}x, \qquad H = \int_{-\infty}^{\infty} (\hbar^2 |\psi_x|^2 - |\psi|^4) \, \mathrm{d}x, \tag{2.1}$$

both of which are conserved quantities for (1.1). The maximum absolute error over all integrations for either of these quantities was observed to occur in the Hamiltonian near the first breaking time, however this error was found to be less than $O(10^{-6})$. The evolution of the Fourier spectrum was also monitored to ensure exponential decay of the spectrum in the large-wavenumber tails at all times.

(*b*) *Data analysis techniques, qualitative and quantitative*

For the purpose of qualitative analysis, a useful representation of the output of each numerical simulation is as a density plot of $|\psi(x,t)|^2$ over the (x,t)-plane, in which maxima appear lighter and minima darker. From such plots, we are able to identify features of the solution like phase-transition boundaries or caustics, and also to characterize the spatio-temporal patterns appearing on the microscales after wave breaking.

The most important quantity associated with each simulation is a measure of t_{c}, the first time of wave breaking. This is the first time when the field $|\psi(x,t)|^2$ as a function of x first fails to be 'smooth', by which we mean that significant features appear on microscales of size $\hbar$ in x. There is of course some arbitrariness about quantifying the breaking time from the output data for finite non-zero $\hbar$, but because microstructure forms on spatio-temporal scales of size $\hbar$, we know that any two reasonable methods for measuring the breaking time must disagree by at most a small (order $\hbar$) amount. We denote a numerical estimate of the breaking time by t_{num} and reserve the notation t_{c} for the limiting value of the breaking time as $\hbar$ tends to zero. The former can depend on $\hbar$ while the latter cannot.

Specifically, we looked for the first sudden jump in the time evolution of L^2 norms of the *residual* defined as the right-hand side of (1.3), and also in the time evolution of the potential energy functional

$$U[\psi(\cdot,t)] := \int_{-\infty}^{\infty} |\psi(x,t)|^4 \,\mathrm{d}x. \tag{2.2}$$

Both of these quantities suddenly become quite large when narrow or large-amplitude structures develop in the field $\rho = |\psi|^2$, and consequently we can use the corresponding transition time as a definition of t_{num}.

Norms of the residual have the advantage that they are directly tied to the discrepancy between the system (1.3) and the model problem (1.4), but the residual is hard to compute accurately when the field $|\psi(x,t)|^2$ decays to zero rapidly (faster than Gaussian) with large x. On the other hand, from the point of view of fibre-optic experiments, looking for large deviations in the potential energy is attractive because $U[\psi]$ is equal to the *intensity autocorrelation trace*

$$Q(\Delta x,t) := \int_{-\infty}^{\infty} |\psi(x,t)|^2 |\psi(x-\Delta x,t)|^2 \,\mathrm{d}x \tag{2.3}$$

evaluated at zero delay, $\Delta x = 0$. In fibre-optic experiments it is not possible practically to measure the intensity $I := |\psi(x,t)|^2$ directly at any given moment x; *the only observable quantities* are in fact correlation traces like (2.3) and the invariant pulse power corresponding to the L^2 norm of $\psi(\cdot,t)$. Both of these measures of breaking time have the desirable property that they do not presume that the wave breaking first occurs at any one particular value of x. This makes the breaking time estimates as useful in simulations when the primary caustic is first observed to form for non-zero x as in simulations where the caustic first appears at $x = 0$.

3. Analytical method for determining breaking times

Given a real analytic bell-shaped function $A(x)$, even in x and hence with maximum value A achieved at $x = 0$, it is possible to express the exact solution of the Cauchy problem for the elliptic system (1.4) in terms of the simultaneous solution of a system of two algebraic equations, parametrized by x and t, in two unknowns (Kamvissis *et al.* 2002). For t small, the system can be solved uniquely.

It is difficult to obtain explicit information about the solution for general values of x and t. However, when $x = 0$, the system of equations degenerates into a single equation, which expresses t as an explicit function of $\rho_\circ$. This relation is

$$|t| = \frac{2}{\pi\sqrt{\rho_\circ}} \int_0^{-\mathrm{i}A^{-1}(\sqrt{\rho_\circ})} E\left(1 - \frac{A(\mathrm{i}y)^2}{\rho_\circ}\right) \mathrm{d}y, \tag{3.1}$$

for $\rho_\circ > A^2$, where $E(m)$ is the complete elliptic integral of the second kind, and $A^{-1}(\cdot)$ is an inverse function to $A(\cdot)$. It is assumed here that $A''(0)$ is strictly negative. Because $A(x)$ is analytic and real for real x, with a maximum at $x = 0$, $A(\mathrm{i}y)$ is also real for real y sufficiently small, and is increasing in $|y|$ because $A''(0) < 0$. The upper limit of integration refers precisely to the value of $y > 0$ for which $A(\mathrm{i}y) = \sqrt{\rho_\circ} > A = A(0)$. Also, one has $\mu \equiv 0$ for $x = 0$.

If it is known *a priori* that the semi-classical solution of (1.1) breaks down first at $x = 0$, then the breaking time corresponds to the first time when $\rho_\circ$ fails to be analytic as a function of t. In any case, even if the solution of the elliptic system (1.4) first breaks down at a non-zero value of x, this calculation still gives an upper bound for the breaking time that we denote by $t_{\rm ub}$. Thus, one is really interested in $\rho_\circ$ as a function of t, which requires inverting the relation (3.1). Under some conditions on the function $A(x)$, the right-hand side of (3.1) will define an analytic function of $\rho_\circ$ for all $\rho_\circ > A^2$. When this is the case, the earliest possible singularity $t_{\rm ub}$ of $\rho_\circ(0,t)$ corresponds to the first critical point of the right-hand side of (3.1), which may occur for $\rho_\circ = \infty$. Note that t is necessarily increasing with respect to $\rho_\circ$ for $\rho_\circ - A^2$ sufficiently small. If the first critical point of the right-hand side of (3.1) is called $\rho_{\circ_{\rm c}}$, then the $x = 0$ breaking time $t_{\rm ub}$ for $\hbar = 0$ is defined by simply substituting $\rho_{\circ_{\rm c}}$ into the right-hand side of (3.1). Under less favourable circumstances, the function $A(x)$ can have a pair of complex singularities on the imaginary x-axis at which $A(x)^2$ takes a finite value $\rho_{\circ_{\max}}$. When this is the case, the right-hand side of (3.1) will fail to be analytic at $\rho_\circ = \rho_{\circ_{\max}} > A^2$. If there are no critical points of the right-hand side of (3.1) for $\rho_\circ \in (A^2, \rho_{\circ_{\max}})$, then the function $\rho_\circ(t)$ obtained by inverting (3.1) will first fail to be analytic at a time $t_{\rm ub}$ obtained by evaluating (3.1) for $\rho_\circ = \rho_{\circ_{\max}}$. Otherwise, the $x = 0$ breaking time $t_{\rm ub}$ is determined as in the case above, which essentially corresponds to $\rho_{\circ_{\max}} = \infty$. One deduces whether $\rho_{\circ_{\max}} < \infty$ simply by looking at the function $A(\mathrm{i}y)$ for positive real y in the range where $A(\mathrm{i}y)$ is analytic and increasing and setting $\rho_{\circ_{\max}}$ equal to the maximum value of $A(\mathrm{i}y)^2$ in this range.

We now examine several particular cases of this analysis that are relevant to the numerical simulations to be presented later.

Case 1: $A(x) = 2\exp(-x^2)$. In this case, the upper limit of integration in (3.1) is

$$-\mathrm{i}A^{-1}(\sqrt{\rho_\circ}) = \sqrt{\log(\sqrt{\rho_\circ}/2)}, \tag{3.2}$$

and $A(\mathrm{i}y) = 2\exp(y^2)$. The right-hand side of (3.1) is analytic for all $\rho_\circ > A^2 = 4$. The first critical point of $t(\rho_\circ)$ corresponds to a value of $t_{\rm ub} = 0.188709$.

Case 2: $A(x) = 2\exp(-x^4)$. In this case, we have $A''(0) = 0$, and formula (3.1) is not valid.

Case 3: $A(x) = 2\nu\exp(-x^2) + 2\varepsilon\exp(-(x^2+\sigma^2)^{3/2})$. We consider $\sigma > 0$. Unlike the other two cases, here $A(x)$ is not entire unless $\varepsilon = 0$. Thus, we also consider ε to be strictly positive here. The upper limit of integration in (3.1) cannot generally be found explicitly and the implicit representation of $-\mathrm{i}A^{-1}(\sqrt{\rho_\circ})$ must suffice for an application of numerical root finding. For all non-zero ε, the right-hand side of (3.1) is analytic only for $\rho_\circ \in (A^2, \rho_{\circ_{\max}}) = ((2\nu + 2\varepsilon\exp(-\sigma^3))^2, (2\nu\exp(\sigma^2) + 2\varepsilon)^2)$.

Passing to the subcase of $\nu = 0$ and $\varepsilon = 1$, it turns out that for σ sufficiently small, the right-hand side of (3.1) defines a strictly increasing function of $\rho_\circ$ in its interval of analyticity, and therefore the $x = 0$ critical time $t_{\rm ub}$ is defined by the time t for $\rho_\circ = \rho_{\circ_{\max}}$. On the other hand, there exists a bifurcation value $\sigma = \sigma_{\rm bif}$ such that for $\sigma > \sigma_{\rm bif}$, there is a first critical point $\rho_{\circ_{\rm c}}$ of the right-hand side of (3.1) in the interval of analyticity $(A^2, \rho_{\circ_{\max}})$ and thus, for $\sigma > \sigma_{\rm bif}$, the $x = 0$ critical time $t_{\rm ub}$ is defined by evaluating the right-hand side of (3.1) for $\rho_\circ = \rho_{\circ_{\rm c}}$. On the other hand, for $\sigma < \sigma_{\rm bif}$, the $x = 0$ critical time is given by $t_{\rm ub} = t(\rho_{\circ_{\max}})$. This information is displayed in the solid curve in figure 5*a*.

Table 1. *Initial data for the three series of numerical experiments*
(In series C we fixed $\hbar = 0.05$.)

series	initial data $\psi(x,0) = A(x)$	parameters varied
A	$A(x) = 2\mathrm{e}^{-(x^2+\sigma^2)^{3/2}}$	σ and $\hbar$
B	$A(x) = 2\mathrm{e}^{-x^2} + 2\varepsilon\mathrm{e}^{-\lvert x\rvert^3}$	ε and $\hbar$
C	$A(x) = 2\mathrm{e}^{-x^2} + 2\varepsilon\mathrm{e}^{-(x^2+\sigma^2)^{3/2}}$	ε and σ

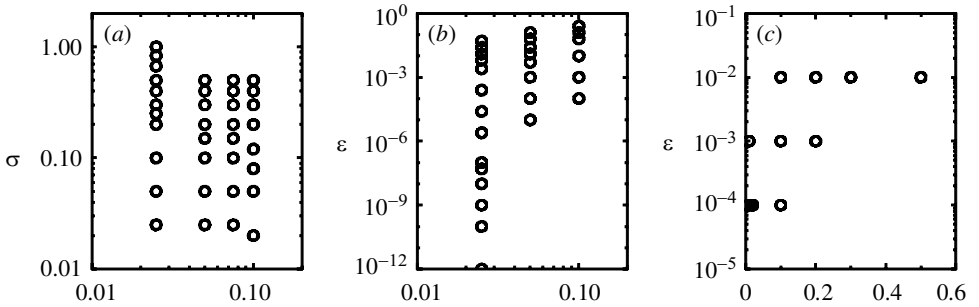

Figure 1. Parameter values for the three series of numerical experiments.
(*a*) Series A; (*b*) series B; (*c*) series C.

A second subcase of interest is that of $\nu = 1$. Here, one again finds that for sufficiently small σ there are no critical points of the right-hand side of (3.1) in the interval of analyticity, and thus the endpoint determines the time t_{ub}. An interesting phenomenon is that for this range of values of σ, the limit $\lim_{\varepsilon\downarrow 0} t(\rho_{\circ_{\max}})$ exists and displays non-trivial structure as a function of σ, but this limit is *not* the same as the clearly σ-independent value of $t_{\mathrm{ub}} = 0.188\,709$ for $\varepsilon = 0$ obtained above in § 3 for the Gaussian initial data. This phenomenon is displayed in figure 10*b*.

4. The numerical experiments

We carried out three series of numerical experiments. In each series, the form of the initial condition was fixed, and two parameters were varied. In series A and series B, one of the parameters varied was $\hbar$, while in series C we fixed $\hbar = 0.05$ and varied two independent parameters in the initial condition. The three series of experiments are summarized in table 1 and figure 1.

(*a*) *Series A: effects of complex singularities in the initial data*

In series A we studied the effect of the presence of a singularity of the initial condition $A(x)$ on the imaginary x-axis a distance σ from the real axis. As shown in table 1, the initial condition had the form $A(x) = 2\exp(-(x^2+\sigma^2)^{3/2})$, which has a branch point for $x = \pm\mathrm{i}\sigma$ but is real analytic for real x as long as $\sigma > 0$. In this series of experiments, we varied both σ and $\hbar$. A representative family of plots corresponding to the smallest value of $\hbar$ in our experiments is shown in figure 2.

When σ is large enough and fixed as $\hbar$ tends to zero, the pattern observed is like that shown in figure 2*d*. In this plot, we see that the oscillations first appear at two

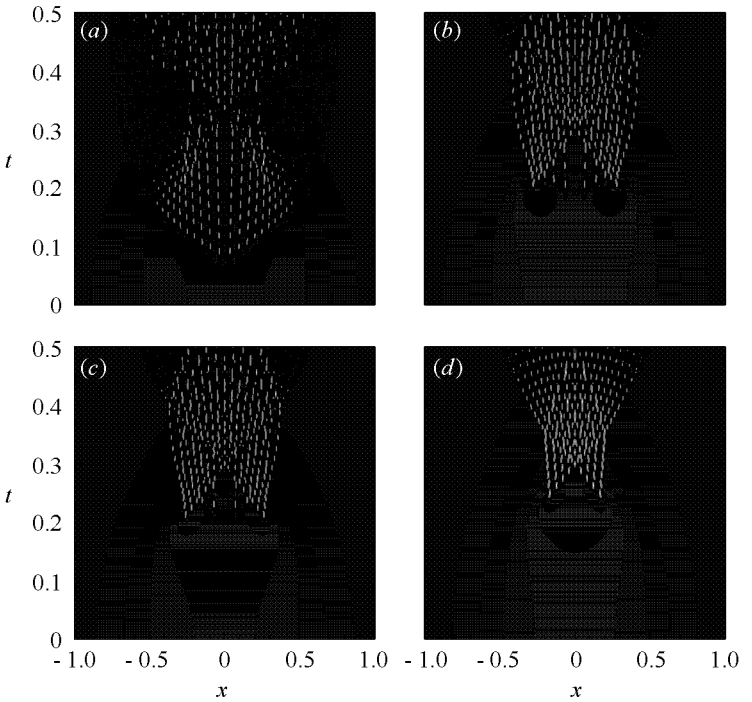

Figure 2. Sample output from series A with $n = 0.025$.
(*a*) $\sigma = 0$, (*b*) $\sigma = 0.25$, (*c*) $\sigma = 0.3$, (*d*) $\sigma = 0.5$.

opposite points $x = \pm x_0$, rather than at $x = 0$. The t values of these points define the breaking time t_c. Now, to our knowledge, previous experiments had observed the formation of higher-order caustics, but these phase-transition boundaries always seemed to spontaneously appear due to a self-interaction of the wave pattern inside the primary caustic. But figure 2*d* shows that higher-order caustics can also arise due to the nonlinear interference of wave patterns contained in lower-order caustics. In any case, as σ is decreased for fixed $\hbar$, a new pattern emerges between the two symmetrical cones of the primary caustic, as seen in figure 2*c*. We refer to this new pattern as a 'beard'. For yet smaller values of σ, the beard extends to shorter times, until it overtakes the two points $x = \pm x_0$ of wave breaking; this is shown in figure 2*b*. Once σ has become small enough for this to occur, the breaking time t_c is defined by the earliest point of the beard, which occurs at $x = 0$. Finally, for σ small enough compared with $\hbar$, the beard dominates the entire pattern as shown in figure 2*a*, and there is evidence that the breaking time defined by the beard location at $x = 0$ becomes small as $\hbar \downarrow 0$. That is, the primary caustic $t = t^{(1)}(x)$ is connected to $t = 0$ at the origin and in the limit there is no open time-interval containing $t = 0$ in which the elliptic model problem (1.4) provides a good description of the dynamics uniformly in x. On the other hand, it appears that the caustic $t = t^{(1)}(x)$ does not contract to $t = 0$ for all x, so there would seem to be regions to the left and right of the oscillatory region where *some* solutions of (1.4) describe the asymptotic behaviour. These two solutions do not need to be matched together in the centre. Incidentally, we also found that, for values of σ much larger than the value corresponding to

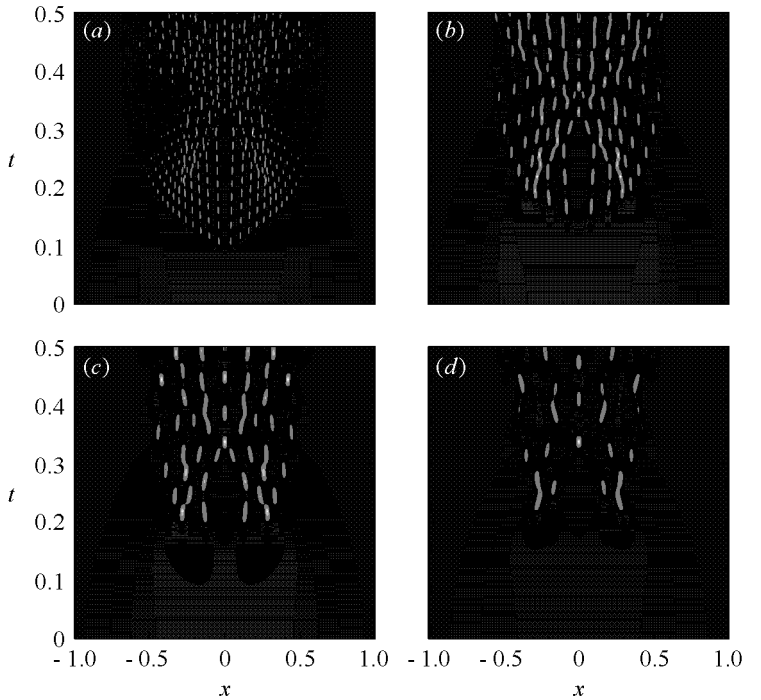

Figure 3. Sample output from series A with $\sigma = 0.05$.
(*a*) $\hbar = 0.025$, (*b*) $\hbar = 0.05$, (*c*) $\hbar = 0.075$, (*d*) $\hbar = 0.1$.

figure 2*d*, the two foci $x = \pm x_0$, well defined in the limit $\hbar \downarrow 0$, ultimately merge into one at $x = 0$, and there is no discernible beard.

We also want to illustrate the effect of tuning down $\hbar$ for a particular fixed initial condition. A collection of corresponding plots from the numerical experiments of series A is given in figure 3. These plots indicate that the formation of a beard is not only related to σ, the distance of the nearest singularity of the initial data from the real x-axis. The key point is that a beard forms when σ is sufficiently small *compared with* $\hbar$. Thus, whereas figure 3*a* is qualitatively quite similar to the plot for the non-analytic case of $\sigma = 0$ in figure 2, we see that as $\hbar$ increases for fixed σ the beard contracts upward until at some value of $\hbar$ it ceases to define the breaking time t_{num} (as in figure 3*c*). Ultimately the wave breaking occurs at two distinct x values as in figure 3*d*. Although less well resolved than the plots in the previous figure,† this plot qualitatively resembles the 'beardless' figure 2*d*.

We want to draw the reader's attention to the structure of the 'beard' in figure 2*a*. The oscillations of $|\psi(x,t)|^2$ that make up the microstructure are qualitatively different from the hexagonal lattice structures present in the 'beardless' plots. But nonetheless they are regular. Using the language of integrable systems, we would classify these oscillations in the 'beard' as being a modulated multiphase wave of genus two. One of the nonlinear phases present in the representation of the waveform in terms of Riemann theta functions has a small amplitude compared with

† Less well resolved only from the point of view of the semi-classical limit, i.e. in the sense that ¯ is bigger. As described in § 2 *a*, the numerical scheme we used to integrate (1.1) was always sufficiently accurate for the range of values of ¯ studied in this paper.

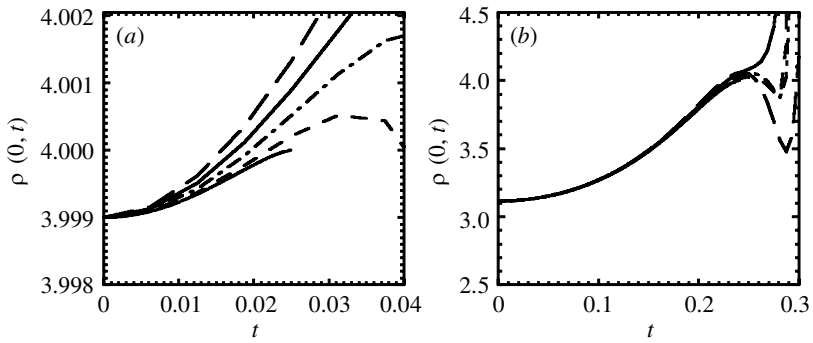

Figure 4. Numerical data for $\rho(0,t) := |\psi(0,t)|^2$ as a function of t for (a) $\sigma = 0.05$ and (b) $\sigma = 0.5$, and several different values of $\hbar$ shown with dashed curves. These are compared with the exact solution $\rho_\circ(0,t)$ of the model ($\hbar = 0$) problem (1.4) computed using formula (3.1) and shown with solid curves. At the t value where the solid curves stop, the solution of (1.4) fails to be analytic and further continuation is not possible.

the others (i.e. one of the spectral gaps is small), so the wave pattern looks like a cnoidal wavetrain (genus one) with a superimposed time-periodic modulation of the peak amplitude. This would explain the 'dashed-line' structure within the beard in part (a). It is intriguing that this plot is for $\sigma = 0$, where the initial data have a real point of non-analyticity, and we expect immediate wave breaking at $t = 0$ (more concerning this below). So we are predicting that even if the initial data are non-analytic at isolated points, the semi-classical behaviour for $t > 0$ close to the singularity will be described by genus two theta functions, as when analytic initial data break spontaneously at a strictly positive time.

Using the numerical data for any strictly positive σ, we can confirm semi-classical convergence to the solution of the ill-posed model problem (1.4) by comparing $|\psi(0,t)|^2$ with the function $\rho_\circ(0,t)$ obtained by inverting the relation (3.1). The results are shown in figure 4 for $\sigma = 0.05$ (in figure 4a) and $\sigma = 0.5$ (in figure 4b) for several different values of $\hbar$. The numerical data display strong convergence to $\rho_\circ(0,t)$ as $\hbar \downarrow 0$ for those times until (3.1) can no longer be inverted, after which time a single-valued solution of the elliptic model problem (1.4) no longer exists.

From each of the simulations in series A, we extracted an estimate $t_{\rm num}$ of the breaking time $t_{\rm c}$ as described in § 2 b. The results are given in figure 5. In figure 5a we give the estimated breaking time $t_{\rm num}$ as a function of σ for several different values of $\hbar$, as determined from the numerical data. However, also included in this plot is a graph of the curve $t_{\rm ub}$ for $\hbar = 0$ as a function of σ, as calculated from the first singularity of the function $\rho_\circ(0,t)$ obtained by inverting (3.1). This latter quantity does not always agree with the breaking time $t_{\rm c}$, since oscillations may appear earlier at non-zero values of x, but it provides an upper bound that is sharp when breaking first occurs at $x = 0$. The derivative discontinuity and sharp upturn in this curve near $\sigma = 0.9$ correspond to a true bifurcation at some definite point $\sigma = \sigma_{\rm bif}$. This bifurcation occurs because for $\sigma < \sigma_{\rm bif}$ the function $t(\rho_\circ)$ is smooth and monotonically increasing for $A(0)^2 < \rho_\circ < \rho_{\circ_{\max}} = 4$, so the first singularity always occurs for $\rho_\circ = 4$, while for $\sigma > \sigma_{\rm bif}$ there is a local maximum of $t(\rho_\circ)$ in this interval that causes the breakdown. The derivative of the function $t(\rho_\circ)$ vanishes at $\rho_\circ = 4$ for $\sigma = \sigma_{\rm bif}$, which explains the continuity at the bifurcation point. We note

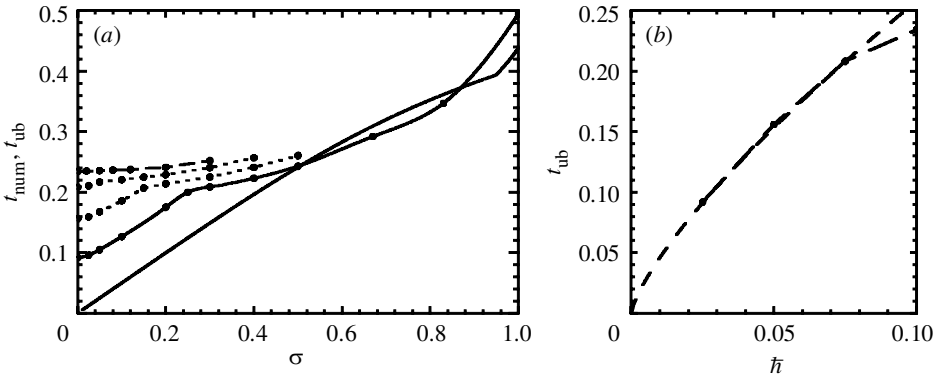

Figure 5. (*a*) The estimated breaking time for series A as a function of the analyticity parameter σ for several different values of $\hbar$. From top to bottom, the data curves correspond to $\hbar = 0.1$, 0.075, 0.05, 0.025. The solid curve is the breaking time $t_{\rm ub}$ at $x = 0$ obtained for the limiting case $\hbar = 0$ from the analytical method described in the text. (*b*) Estimated breaking time for series A with $\sigma = 0$ as a function of $\hbar$, and a fit to $t_{\rm num} = 1.45\hbar^{3/4}$.

that for small $\hbar$ the data curves follow the analytical prediction for the break time both for sufficiently small σ and for sufficiently large σ. In both of these regimes, the wave breaking actually occurs first at $x = 0$, whereas in the intermediate regime the wave breaking occurs first for non-zero values of x and therefore the exact solution of (1.4) that is encoded in formula (3.1) for $x = 0$ cannot capture the earliest interesting events.

In figure 5*b*, we compare the breaking times $t_{\rm num}$ obtained for $\sigma = 0$ (for which the theoretical breaking time at $\hbar = 0$ is for $x = 0$ is $t_{\rm ub} = 0$) with a fit to the function $t_{\rm num} = 1.45\hbar^{3/4}$, thus confirming the contraction of the caustic curve $t^{(1)}(x)$ to $t = 0$ with decreasing $\hbar$. We have remarked in § 2 *b* that any method of estimating breaking times for finite non-zero $\hbar$ can only be accurate up to an error the size of the typical scales of the spatio-temporal microstructure that is present. As these are order $\hbar$, our estimates of breaking times have errors of a similar size. But the fact that the data shown in figure 5*b* indicate a breaking time $t_{\rm num}$ large compared with $\hbar$ while still going to zero with $\hbar$, suggests that the fit to the data is sensible (we would have to admit less confidence if the fit suggested an exponent greater than or equal to 1).

Now, it has been remarked that the analytical method given in Kamvissis *et al.* (2002) for solving the elliptic system (1.4), and which for $x = 0$ degenerates into inverting the explicit relation (3.1), fails when $A''(0) = 0$. We have also observed that when $\sigma = 0$, there is no time-interval in which a solution of (1.4) exists, which is indicated in the numerical simulations by the pulling back of the primary caustic to $t = 0$. But for the $\sigma = 0$ initial data we have $A(x) = 2\exp(-|x|^3)$, which not only fails to be analytic at $x = 0$ but also when considered as a real function of real x satisfies $A''(0) = 0$. Thus it is of some interest to determine whether the pulling back of the primary caustic as $\hbar \downarrow 0$ is connected with the non-analyticity or with the non-convexity. To examine this question, we also carried out numerical simulations of (1.1) for several values of $\hbar$ with the initial condition $\psi(x, 0) = A(x) = 2\exp(-x^4)$. These data are analytic but have an inflection point at the peak ($x = 0$). The output of two of our simulations is given in figure 6. Here, we see that as $\hbar$ is halved, the breaking time $t_{\rm num}$ hardly changes at all, in contrast with the data given in figure 5*b*

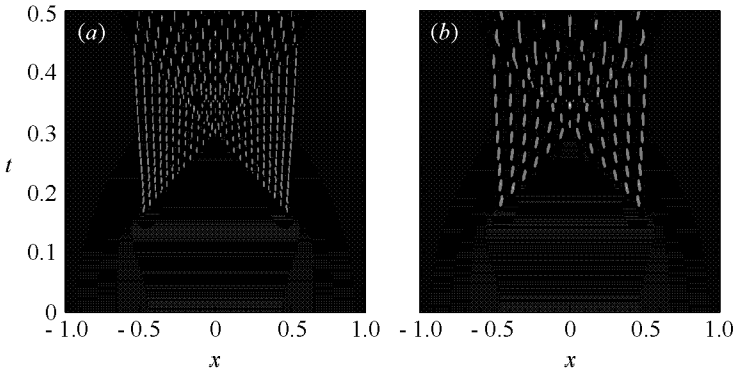

Figure 6. For comparison, two simulations with initial data $A(x) = 2\exp(-x^4)$. (*a*) $\hbar = 0.025$, (*b*) $\hbar = 0.05$.

for the non-analytic case of $A(x) = 2\exp(-|x|^3)$. This suggests strongly that the lack of convexity at the peak is only a technical obstruction to the construction given in Kamvissis *et al.* (2002) of solutions of the elliptic system (1.4), whereas lack of analyticity is an essential obstruction.

This indicates that the relation between the behaviour of even, bell-shaped initial data exactly at the centre point $x = 0$ and the behaviour of the corresponding solutions of (1.1) are rather subtle. For even data, one could consider the equivalent initial/boundary-value problem for (1.1) on the half-line $x \geqslant 0$, where the solution $\psi(x,t)$ satisfies Neumann boundary conditions at $x = 0$, i.e. $\psi_x(0,t) = 0$. In this interpretation, the above conclusions indicate that the initial data for the problem on the half-line must not only be analytic for $0 < x < \infty$, but must also be analytic *at the boundary point* $x = 0$ in the sense that the symmetric extension of the data has a power series expansion at $x = 0$ for a solution of the elliptic model problem (1.4) to exist uniformly for any positive time at all. If for the half-line problem, we give the initial data $\psi(x,0) = A(x) = 2\exp(-x^3)$ for $x \geqslant 0$, then there exists an analytic extension to all real x, but it is not even in x and the primary caustic is attached to $t = 0$ at $x = 0$.

(*b*) *Series B: effects of small non-analytic perturbations*

In series B, we investigated the effect of a non-analytic perturbation added to pure Gaussian (and hence analytic) initial data, as the size of the perturbation is varied with $\hbar$. As shown in table 1, the initial data for (1.1) in series B were of the form $\psi(x,0) = A(x) = 2\exp(-x^2) + 2\varepsilon\exp(-|x|^3)$. Here, ε parametrizes the size of the perturbation. Note that the perturbation term is not only small with ε in the absolute sense, but also in the relative sense when compared with $A(x)$ for $\varepsilon = 0$.

We chose a perturbation that, while not analytic at $x = 0$, was still of class $C^2(\mathbb{R})$. Indeed, it has already been shown by Bronski & Kutz (1999) that non-smooth (in particular, 'tent-shaped') perturbations can considerably degrade the structures present in the evolution of an unperturbed analytic initial condition. We wanted to make sure that the initial data were smooth enough that the classical nature of solutions of the initial-value problem (1.1) for fixed $\hbar$ was not an issue.

A representative sequence of simulations from series B for $\hbar = 0.025$ is shown in figure 7. These four experiments display the remarkable sensitivity of the semi-

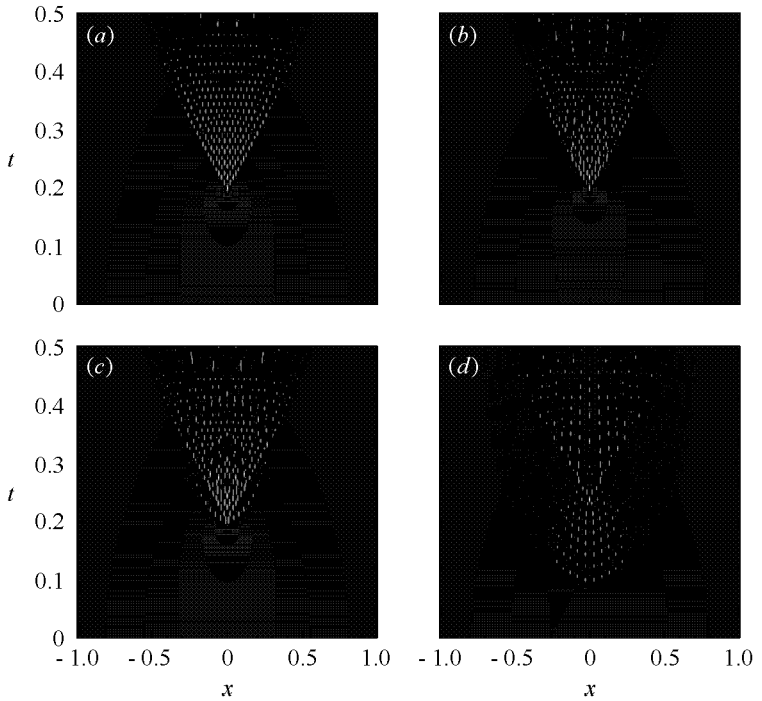

Figure 7. Sample output from series B with $\hbar = 0.025$.
(*a*) $\varepsilon = 0$, (*b*) $\varepsilon = 10^{-12}$, (*c*) $\varepsilon = 10^{-10}$, (*d*) $\varepsilon = 0.05$.

classical limit for (1.1) to the presence of non-analyticity in the initial data. Figure 7*a* shows the evolution for $\hbar = 0.025$ of the unperturbed Gaussian initial condition $\psi(x,0) = A(x) = 2\exp(-x^2)$. This plot shows qualitatively similar features to those obtained from exact solutions of (1.1) for $\psi(x,0) = A(x) = 2\,\mathrm{sech}(x)$ in Miller & Kamvissis (1998). Namely, there is a smooth field outside of a primary caustic curve $t^{(1)}(x)$ anchored at $t = t_c$ and $x = 0$. Inside the primary caustic curve there is a regular hexagonal lattice pattern of maxima indicating microstructure that is well approximated (for small $\hbar$) by a modulated genus two hyperelliptic solution of (1.1). At the top of the figure just before $t = 0.5$, one begins to see the formation of a secondary caustic curve $t = t^{(2)}(x)$ separating the hexagonal pattern from a region of more complicated microstructure. Now, in passing to figure 7*b* corresponding to $\varepsilon = 10^{-12}$, there already appear qualitative differences in the evolution, not so much in the location of the primary caustic, but more in the details of the pattern of maxima inside the caustic. However, by the time ε is increased only to a value of $\varepsilon = 10^{-10}$, a beard has formed that changes the shape of the primary caustic. This caustic becomes larger and gets closer to $t = 0$ as ε is increased (see figure 7*d*).

In fact, the contraction of the primary caustic to $t = 0$ is due not only to the size of ε, but more accurately to the size of ε compared with $\hbar$. This is illustrated in figure 8, which indicates the region of ε values where the transition in behaviour occurs for a different larger value of $\hbar$. Here we see that the regular pattern within the primary caustic survives for $\hbar = 0.05$ until $\varepsilon \sim 10^{-4}$, a much larger value (about a hundred million times larger!) than in the experiments shown in figure 7 for $\hbar$ only half the size.

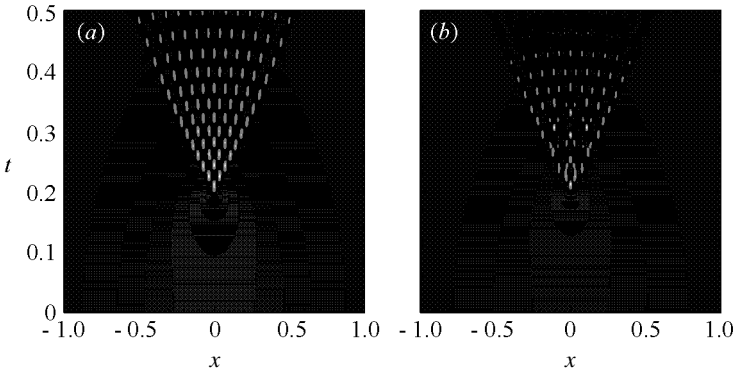

Figure 8. Sample output from series B with $\hbar = 0.05$. (*a*) $\varepsilon = 10^{-5}$, (*b*) $\varepsilon = 10^{-4}$.

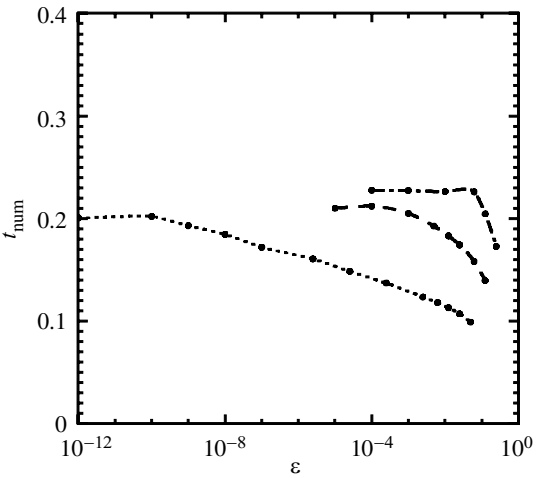

Figure 9. The estimated breaking time for series B as a function of the perturbation parameter ε for several different values of $\hbar$. Top curve, $\hbar = 0.1$; middle curve, $\hbar = 0.05$; bottom curve, $\hbar = 0.025$.

For each of the simulations in series B, we computed the estimated breaking time t_{num} as described in § 2 *b*. The results are shown in figure 9. In this figure, the estimated breaking time t_{num} is plotted as a function of ε for several different values of $\hbar$. Note that the analytical method for predicting the breaking time in series B always gives the result $t_{\mathrm{ub}} = 0$ for all $\varepsilon > 0$ because the data are non-analytic at $x = 0$. We see clearly that for any given value of $\hbar$ the breaking time is insensitive to ε for ε sufficiently small, but that transition region of ε is not uniform as $\hbar \downarrow 0$. We conclude that the effect of class $C^2(\mathbb{R})$ non-analytic perturbations is always that the primary caustic is drawn back to $t = 0$ for any ε in the limit $\hbar \downarrow 0$, but that for any given $\hbar$ there exist ε sufficiently small that the effect goes unnoticed.

(*c*) *Series C: relative effects of size and proximity of complex singularities of perturbations*

In series C we fixed $\hbar = 0.05$, a value that we want to suppose is small enough to be 'in' the semi-classical limit. We then computed estimated breaking times t_{num} for

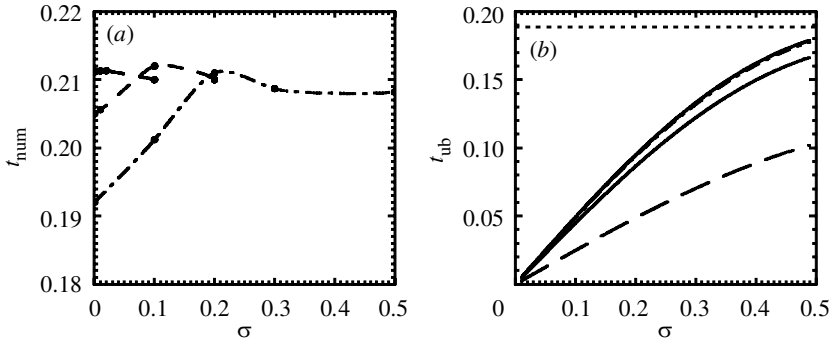

Figure 10. The estimated breaking time $t_{\rm num}$ for series C as a function of the analyticity parameter σ for several different values of the perturbation parameter ε. (*a*) Numerical data for $\hbar = 0.05$. From top to bottom, the three spline-interpolated datasets correspond to $\varepsilon = 10^{-4}$, 10^{-3}, 10^{-2}. (*b*) Analytical breaking time $t_{\rm ub}$ at $x = 0$ for the theoretical limit $\hbar = 0$. The horizontal line corresponds to the theoretical breaking time for $\varepsilon = 0$, which agrees well with the breaking time observed in the unperturbed Gaussian simulation shown in figure 7*a*. There are also five curves shown in the plot corresponding to (from top to bottom), $\varepsilon = 10^{-4}$, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}. The top two curves are so close as to be indistinguishable to the eye, and they thus represent the limit of ε tending to zero from above.

perturbed Gaussian initial conditions (cf. table 1) given by

$$\psi(x,0) = A(x) = 2\exp(-x^2) + 2\varepsilon\exp(-(x^2+\sigma^2)^{3/2})$$

over a collection of parameter pairs (ε, σ). The plots of the particular simulations in the series are qualitatively similar to those given in § 4 *b* for series B, so we omit these here. However, for each experiment in series C we again computed the estimated breaking time $t_{\rm num}$ using the data-analysis methods described in § 2 *b*. These results are shown in figure 10. Figure 10*a* shows the estimated breaking times plotted against σ for $\varepsilon = 10^{-4}$, $\varepsilon = 10^{-3}$ and $\varepsilon = 10^{-2}$ calculated from the numerical data for $\hbar = 0.05$. For comparison, figure 10*b* shows the analytical predictions for the breaking times $t_{\rm ub}$ at $x = 0$ obtained by inverting the relation (3.1) and looking for the earliest singular point. From top to bottom, the curves in this plot correspond to $\varepsilon = 10^{-4}$, $\varepsilon = 10^{-3}$, $\varepsilon = 10^{-2}$, $\varepsilon = 10^{-1}$ and $\varepsilon = 1$, and the horizontal line corresponds to $\varepsilon = 0$. The top two curves corresponding to $\varepsilon = 10^{-4}$ and $\varepsilon = 10^{-3}$ are so close to each other that they are indistinguishable to the eye. We interpret this as the limit of the theoretical breaking time as ε tends to zero from positive values. Note that this limit does *not* agree with the theoretical breaking time for $\varepsilon = 0$, which is naturally independent of σ. From the numerical data in figure 10*a*, we see that for any given ε, there exists a range of σ sufficiently large that the breaking time is relatively insensitive to σ, while for σ smaller than some transition value the breaking time decreases sharply as σ decreases. Again, this effect is not uniform in ε, the perturbation size. If ε is made smaller than before, then the transition value of σ is correspondingly smaller. We see from figure 10*a* that when ε is sufficiently small (for the fixed value of $\hbar$ for which the simulations in series C were carried out) the transition point of the curve has disappeared to the left of $\sigma = 0$, and consequently for such ε the breaking time becomes more or less independent of σ all the way down to $\sigma = 0$, the non-analytic case. This is the effect of the finiteness of $\hbar$ in this series of simulations.

5. Banach spaces of analytic functions

The numerical simulations presented above seem to suggest a limited kind of stability for small $\hbar$ as long as sufficient control is placed upon the admissible perturbations. But one wants to know what to expect as $\hbar$ is made smaller than any other scales in the problem, and we truly are in the semi-classical limit. Clearly, to have any hope of stability for the limit in any sense, one must have both analytic unperturbed initial data and also a class of admissible perturbations consisting of analytic functions. And to measure the effects of perturbations of a small 'size' one must work with a set of analytic functions that make up a Banach space. This requires making some choices because, for example, the subspace of all real analytic functions in $L^2(\mathbb{R})$ is not closed with respect to the natural $L^2(\mathbb{R})$ norm. So what are the right function spaces, and what are the right norms?

One general way to equip spaces of analytic functions with a norm is consistent with all of the experiments we have presented in this paper for which the initial data for (1.1) were analytic. Let A_δ denote the 'strip space' of real analytic functions $\psi(x)$ of real x that can be continued into the complex x-plane throughout the open symmetric strip surrounding the real axis of uniform width 2δ, and that decay to zero uniformly as x tends to infinity within the strip. Introduce the norm

$$\|\psi\|_\delta^\infty := \sup_{x\in\mathbb{R},\ |y|<\delta} |\psi(x+\mathrm{i}y)|. \tag{5.1}$$

Since uniform limits of analytic functions are themselves analytic, this norm makes A_δ^∞, the subset of functions in A_δ that are uniformly bounded as $|y| \to \delta$, into a Banach space. So for example, in the numerical experiments of series A or series C with the parameter σ held fixed, all of the initial data are contained in A_δ^∞ for all $\delta < \sigma$, and the terms proportional to ε in series C can be considered to be small with ε in the $\|\cdot\|_\delta^\infty$ norm. The strip spaces may be equipped with other norms as well. For example replacing the L^∞ norm over lines parallel to the real axis in our definition by the L^2 norm gives the structure of a Banach space to the subset A_δ^2 of A_δ containing functions for which the norm is finite. The spaces A_δ^2 are the Gevray spaces. They differ essentially from the A_δ^∞ spaces only in that they admit more singular behaviour (e.g. negative one fourth root singularities) at the boundary of the strip, so $A_\delta^\infty \subset A_\delta^2$. Since our immediate usage of these spaces does not involve behaviour on the boundary (if singularities were to appear on the boundary, we would just make δ smaller), either of these Banach spaces, or indeed other ones similarly defined as subsets of the basic strip space A_δ, will suffice.

Another way to give analytic functions a Banach space structure is to work in band-limited function spaces. Let F_K denote the linear subspace of $L^2(\mathbb{R})$ consisting of functions $\psi(x)$ whose Fourier transforms $\hat{\psi}(k)$ vanish identically for almost all k with $|k| > K$. We equip this space with the L^2 norm, and then by the Plancherel theorem we see that this makes F_K into a Banach space (it is closed). The functions in F_K are necessarily analytic in x; in fact it is a consequence of Paley–Wiener theory that the union of the F_K spaces consists of those functions in $L^2(\mathbb{R})$ that continue for complex x into entire functions of exponential type. Therefore if $\psi(x) \in F_K$, we also have $\psi(x) \in A_\delta^\infty$ for all $\delta > 0$, so these spaces are more restrictive than the strip spaces defined above. In fact, none of the initial data for our numerical experiments fit into any of the F_K spaces; even the entire cases $A(x) = 2\exp(-x^2)$ and $A(x) = 2\exp(-x^4)$ have Fourier transforms that are not compactly supported.

In spite of the fact that it is hard to write down a function $\psi(x)$ that is in an F_K space without Fourier theory, the Banach spaces F_K are particularly attractive from the point of view of physical applications, since the norm is just the usual $L^2(\mathbb{R})$ energy norm. Moreover, it is quite reasonable to suppose that in applications, initial data may be considered to be smoothed by a band-limiting process. For example, in optical experiments like those reported in Sudo *et al.* (1989), noisy pulses generated by a laser can easily be routed through a band-pass Fabry–Perot filter prior to launching into the optical fibre. This process effectively yields an ensemble of initial data for (1.1) that is uniformly in some F_K space, where K is the half-width of the filter transmission peak.

One disadvantage of the F_K spaces is that due to the nonlinearity they are not preserved under the evolution of the focusing nonlinear Schrödinger equation (1.1) for any $\hbar > 0$. A space preserved under the evolution might appear more useful, since the evolution under (1.1) to some $t > 0$ of one initial condition $\psi(x)$ could be interpreted as another initial condition, just as good as the first. But we need to emphasize that we are interested in taking the limit $\hbar \downarrow 0$ before examining the behaviour of the solution at different values of $t > 0$. Since $\hbar$-independent analytic initial data of the form $\psi(x) = A(x)$ will evolve semi-classically for $t > 0$ sufficiently small into the WKB form $\psi(x,t) \approx A(x,t)\exp(\mathrm{i}S(x,t)/\hbar)$ with $S(x,t) \not\equiv 0$, the asymptotic solution $\psi(x,t)$ of (1.1) fails to exist as an ordinary function for any $t > 0$. Thus, the strip spaces also fail to be invariant under the *semi-classical* evolution of (1.1). So we take the point of view in this paper that the 'semi-classical nonlinear Schrödinger equation' is a map taking some class of smooth, $\hbar$-independent initial data (e.g. A_δ^∞ or F_K) to certain output data (e.g. coordinates of caustic curves), that we would like, ideally, to be continuous functionals.

In fact, the calculations presented in figure 10*b* show that the control of being in the strip space A_δ^∞ and working with perturbations that are uniformly small in the corresponding norm is *not* sufficient to yield stability of the asymptotic $\hbar \downarrow 0$ behaviour. This is because one sees that for fixed $\sigma > 0$ and for any positive $\delta < \sigma$, the perturbation of the Gaussian under consideration has a $\|\cdot\|_\delta^\infty$ norm that is arbitrarily small as ε tends to zero from above. But the theoretical calculations based on the exact solution of the elliptic problem (1.4) at $x = 0$ show clearly that for any $\sigma > 0$

$$\lim_{\varepsilon \downarrow 0} t_{\mathrm{ub}} \neq t_{\mathrm{ub}}|_{\varepsilon=0}, \tag{5.2}$$

i.e. the time of first wave breaking at $x = 0$ in the semi-classical limit is not a continuous functional on the Banach space A_δ^∞ for any $\delta < \sigma$.

The calculations and numerical simulations presented here do not allow us to make any definitive statements regarding stability of such functionals on the band-limited function spaces F_K. However, since all functions in F_K are necessarily entire, there can be no complex singularities on the imaginary x-axis of the type that leads to the discontinuity (5.2). We feel compelled to conjecture that all reasonable features of the semi-classical behaviour of solutions of (1.1)—features like weak limits of ρ and μ in the vicinity of fixed x and t not on any caustic curve, or the t coordinate of a particular caustic curve at a given fixed value of x—are stable with respect to perturbations in any given F_K space (i.e. are continuous functionals on F_K). One would expect the modulus of continuity to not be uniform with respect to K.

6. Conclusion

Whereas semi-classical or dispersionless limits for 'stable' integrable partial differential equations are fairly well understood (Jin *et al.* 1999; Lax & Levermore 1983*a–c*), we are only beginning to explore the corresponding questions for 'unstable' problems like the focusing nonlinear Schrödinger equation (1.1). The formal semi-classical limit of (1.1) meant to be valid for sufficiently small times independent of $\hbar$ turns out to be an ill-posed initial-value problem (1.4) that can only be solved for analytic initial data. It is therefore clear that for small $\hbar$ some kind of extreme sensitivity to the presence in the initial data of points of non-analyticity either on or possibly just sufficiently close to the real x-axis should be expected. The three series of numerical experiments presented in this paper were undertaken to investigate this sensitivity. The simple formula (3.1) giving the exact solution to the model problem (1.4) for appropriate data was also very useful in this analysis.

A quantitative observation based on images like those shown in figures 2, 3, 7 and 8 is that the presence of non-analytic points of the initial data near the real x-axis leads to the formation of 'beards' on a regular background wave pattern. When these non-analytic points are far from the real axis, or when the perturbation is not too large, the beards do not change the caustic curves much, and tend to influence the patterns of maxima within the caustics. In some of the plots, we see the hexagonal pattern characteristic of modulated genus two wavetrains go over into a 'chain-link' pattern possibly characteristic of higher genus. But when the analytic points become close to the real axis, or if the perturbation is sufficiently large, the beard dominates the whole wave pattern and completely determines the primary caustic. In our experiments, the wave pattern within the beard attached to the non-analytic point $x = 0$ in the semi-classical limit does not appear to be too irregular. This suggests the intriguing possibility that in the semi-classical limit, the nonlinear Schrödinger equation (1.1) automatically and immediately (i.e. on time-scales much shorter than $\hbar$) 'repairs' isolated non-analytic points in the initial data by passing locally to a modulated multiphase waveform of sufficiently large but finite genus.

A second qualitative observation is that the first wave breaking does not always occur at $x = 0$, although the initial data are even and bell shaped as a function of real x. The clearest example of breaking first occurring for a symmetric pair of non-zero x values $\pm x_0$ is the evolution of the non-convex initial data $\psi(x, 0) = A(x) = 2\exp(-x^4)$ shown in figure 6, but it is also seen in the plot corresponding to $\sigma = 0.5$ in figure 2. For this latter simulation, a careful look at figure 4*b* gives circumstantial evidence that the exact solution of the elliptic model problem (1.4) evaluated at $x = 0$ fails to be analytic at exactly the time that the two wave envelopes born at $x = \pm x_0$ at the breaking time merge at $x = 0$. Now it is familiar from weak dispersion theory that the cusp points of caustic curves are typically encoded in the model problem analogous to (1.4) as singular points in the solutions. On the other hand, one does not ordinarily expect the caustic curves themselves to be obtainable directly from this leading-order model; they arise either from a Gurevich–Pitaevskii type matching procedure (Gurevich & Pitaevskii 1974) for connecting to a solution of higher-order Whitham equations supposed to be valid within the caustic, or from a global theory like the variational theory of Lax & Levermore (1983*a*). However, our numerical experiments in conjunction with the exact solution of the elliptic

problem (1.4) afforded by the relation (3.1) suggest that the point where the two caustic envelopes overlap is also encoded in the quasilinear leading-order problem.

This last point underscores the utility of the simple formula (3.1) for determining semi-classical asymptotic properties of the solution of the nonlinear Schrödinger equation for absolutely arbitrary initial data of the type we have considered in this paper. For such data when $\hbar$ is small, inversion of this explicit formula (i.e. merely turning the graph of the explicit function $t(\rho_\circ)$ on its side) provides a very accurate prediction $\rho_\circ(0,t)$ for the square modulus $|\psi(0,t)|^2$ of the solution of the full nonlinear Schrödinger equation (1.1) at $x=0$ until wave breaking occurs. Although there is no general proof at the moment (see Kamvissis *et al.* (2002) for a special case), all evidence to date points to the conclusion that the solution of the equivalent problem (1.3) is strongly approximated by the solution of the elliptic model problem (1.4), uniformly in x for all t until the first singularity forms in the latter solution. At the worst, a simple formula like (3.1) for the solution of the model problem can assist expensive stiff numerical simulations of the full nonlinear Schrödinger equation close to the semi-classical limit. And at best, it can provide an alternative to them. The time of first appearance of a singularity in the function $\rho_\circ(0,t)$ is expected to provide an upper bound on the breaking time, a bound that is sharp in cases when the waves break at $x=0$.

One of the predictions of the exact solution to the elliptic model problem afforded by the relation (3.1) is that certain properties of the semi-classical limit for the focusing nonlinear Schrödinger equation do not necessarily depend continuously on the size of a real analytic perturbation of the initial data, even if the size of the perturbation is controlled uniformly in a strip about the real axis with non-trivial width in the imaginary direction. In making this statement, we are thinking of $\hbar$ going to zero before any other perturbation parameters. As shown in figure 10, it is not the size of the perturbation that matters in determining the time of the first singularity of $\rho_\circ$ at $x=0$, but rather whether it is present at all and whether it has complex singularities.

This sort of instability of semi-classical asymptotics is not unique to the focusing nonlinear Schrödinger problem. Indeed it has been observed in the defocusing case as well by Forest & McLaughlin (1998). These authors show that certain piecewise-constant initial data evolve in such a way as to never 'break', while any C^∞ smoothing of these data, however fine, leads to wave breaking and the subsequent onset of $O(\hbar)$ wavelength oscillations in the amplitude of the solution in finite time. This is especially interesting because the defocusing nonlinear Schrödinger equation, for which the semi-classical asymptotics were expounded by Jin *et al.* (1999), is one of the well-understood, 'stable' problems.

For the *focusing* nonlinear Schrödinger equation it appears that complex singularities in the initial data can lead to the discontinuity of certain functionals associated with the semi-classical limit of (1.1) even for data analytic on the real x-axis. This fact leads us to suspect that the right spaces to consider in order to obtain stability of the limit are spaces of entire functions, like the band-limited spaces F_K discussed in § 5. These spaces are especially interesting from the point of view of physical applications of the focusing nonlinear Schrödinger equation, because filtering processes that can lead to effectively band-limited initial data are quite natural and familiar in the laboratory setting. For appropriately 'pre-filtered' initial data, the semi-classical limit for (1.1) may be quite stable indeed.

This work was supported by the Australian Research Council and Monash University. S.R.C. was supported through ARC grant no. A89899458, while P.D.M was supported by a Monash University Logan Fellowship. The authors thank L. K. Antanovskii for stimulating discussions.

References

Antanovskii, L. K. 1996 Microscale theory of surface tension. *Phys. Rev.* E **54**, 6285–6290.

Antanovskii, L. K., Rogers, C. & Schief, W. K. 1997 A note on a capillarity model and the nonlinear Schrödinger equation. *J. Phys.* A **30**, L555–L557.

Bronski, J. C. 1996 Semiclassical eigenvalue distribution of the Zakharov–Shabat eigenvalue problem. *Physica* D **97**, 376–397.

Bronski, J. C. 2001 Spectral instability of the semiclassical Zakharov–Shabat eigenvalue problem. *Physica* D **152–153**, 163–170.

Bronski, J. C. & Kutz, J. N. 1999 Numerical simulation of the semi-classical limit of the focusing nonlinear Schrödinger equation. *Phys. Lett.* A **254**, 325–336.

Ercolani, N. M., Jin, S., Levermore, C. D. & MacEvoy Jr, W. 1993 The zero dispersion limit of the NLS/mKdV hierarchy for the nonselfadjoint ZS operator. Preprint.

Forest, M. G., & McLaughlin, K. T.-R. 1998 Onset of oscillations in nonsoliton pulses in nonlinear dispersive fibers. *J. Nonlin. Sci.* **7**, 43–62.

Gurevich, A. V. & Pitaevskii, L. P. 1974 Nonstationary structure of a collisionless shock wave. *Sov. Phys. JETP* **38**, 291–297. (In Russian as *Z. Eksperiment. Teoret. Fiz.* **65**, 590–604 (August 1973).)

Jin, S., Levermore, C. D. & McLaughlin, D. W. 1999 The semiclassical limit of the defocusing NLS hierarchy. *Commun. Pure Appl. Math.* **52**, 613–654.

Kamvissis, S., McLaughlin, K. T.-R. & Miller, P. D. 2002 Semiclassical soliton ensembles for the focusing nonlinear Schrödinger equation. (Submitted.)

Lax, P. D. & Levermore, C. D. 1983*a* The small dispersion limit of the Korteweg–de Vries equation. I. *Commun. Pure Appl. Math.* **36**, 253–290.

Lax, P. D. & Levermore, C. D. 1983*b* The small dispersion limit of the Korteweg–de Vries equation. II. *Commun. Pure Appl. Math.* **36**, 571–593.

Lax, P. D. & Levermore, C. D. 1983*c* The small dispersion limit of the Korteweg–de Vries equation. III. *Commun. Pure Appl. Math.* **36**, 809–830.

Miller, P. D. 2001 Some remarks on a WKB method for the nonselfadjoint Zakharov–Shabat eigenvalue problem with analytic potentials and fast phase. *Physica* D **152–153**, 145–162.

Miller, P. D. & Kamvissis, S. 1998 On the semiclassical limit of the focusing nonlinear Schrödinger equation. *Phys. Lett.* A **247**, 75–86.

Newell, A. C. & Moloney, J. V. 1992 *Nonlinear optics.* Addison Wesley.

Satsuma, J. & Yajima, N. 1974 Initial value problems of one-dimensional self-modulation of nonlinear waves in dispersive media. *Supp. Prog. Theor. Phys.* **55**, 284–306.

Sudo, S., Itoh, H., Okamoto, K. & Kubodera, K. 1989 Generation of 5 THz repetition optical pulses by modulation instability in optical fibers. *Appl. Phys. Lett.* **54**, 993–994.

Zakharov, V. E. & Shabat, A. B. 1972 Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Sov. Phys. JETP* **34**, 62–69. (In Russian as *Z. Eksperiment. Teoret. Fiz.* **61**, 118–134 (July 1971).)