154

Wavefront dislocations in the Aharonov–Bohm effect and its water wave analogue

M V Berry, R G Chambers, M D Large, C Upstill and J C Walmsley

H H Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, England

Abstract We study the wavefronts (i.e. the surfaces of constant phase) of the wave discussed by Aharonov and Bohm, representing a beam of particles with charge q scattered by an impenetrable cylinder of radius R containing magnetic flux Φ. Defining the quantum flux parameter by $\alpha = q\Phi/h$, we show that for the case $R = 0$ the wave ψ_{AB} possesses a wavefront dislocation on the flux line, whose strength (i.e. the number of wave crests ending on the dislocation) equals the nearest integer to α. When α passes through half-integer values, the strength changes, by wavefronts unlinking and reconnecting along a nodal surface. In quantum mechanics this phase structure is unobservable, but we devise an analogue where surface waves on water encounter an irrotational 'bathtub' vortex; in this case α depends on the frequency of the waves and the circulation of the vortex. Experiments show dislocation structures agreeing with those predicted. ψ_{AB} is an unusual function, in which incident and scattered waves cannot be clearly separated in all asymptotic directions; we discuss its properties using a new asymptotic method.

Résumé L'article est consacré aux surfaces d'onde (surfaces de phase constante) de l'onde, introduite par Aharonov et Bohm, qui décrit un faisceau de particules de charge q diffusé par un cylindre impénétrable, de rayon R, traversé par un flux magnétique Φ. Soit $\alpha = q\Phi/h$ le paramètre adapté à une description quantique du flux. On montre que, pour $R = 0$, l'onde ψ_{AB} de Aharonov et Bohm présente une dislocation de la surface d'onde sur la ligne de flux; la 'force' de cette dislocation, c'est-à-dire le nombre de maxima de vibration se terminant sur la dislocation, est égale au nombre entier le plus proche de α. Quand α passe par une

valeur demi-entière, la 'force' change, les surfaces d'onde se réarrangeant le long d'une surface nodale. Cette structure des surfaces d'onde n'est pas observable en Mécanique Quantique, mais l'on peut proposer une analogie hydrodynamique, où des ondes de surface sur l'eau rencontrent un tourbillon irrotationnel (du type 'vidange de baignoire'); dans ce cas, α dépend de la fréquence des ondes et de la circulation du tourbillon. L'expérience met en évidence des structures de dislocations en accord avec les prédictions théoriques. ψ_{AB} est une fonction de comportement inhabituel, pour laquelle les ondes incidente et diffusée ne peuvent être clairement distinguées dans toutes les directions asymptotiques; ses propriétés sont analysées à l'aide d'une méthode asymptotique originale.

1 Introduction

In an influential paper, Aharonov and Bohm (1959) studied the quantum mechanics of a beam of particles (with charge q and mass m) incident normally on a long thin cylinder containing a magnetic field $\boldsymbol{B}(\boldsymbol{r})$ parallel to its length. They supposed that the electrons could not penetrate into the cylinder and that the magnetic field could not leak out. This mutual inaccessibility of particles and field ensures that in classical mechanics the scattering pattern of particles beyond the cylinder cannot depend on the field inside. But in Schrödinger's equation it is the magnetic vector potential $\boldsymbol{A}(\boldsymbol{r})$, and not the field, that determines the wavefunction, and $\boldsymbol{A}(\boldsymbol{r})$ outside the cylinder contains the imprint of the field inside via the magnetic flux Φ, given by the Stokes relation

$$\Phi = \oint \boldsymbol{A}(\boldsymbol{r}) \cdot d\boldsymbol{r} = \oiint \boldsymbol{B}(\boldsymbol{r}) \cdot d\boldsymbol{S}, \qquad (1)$$

where the integration path encloses the cylinder. Aharonov and Bohm (1959) showed that the quantum mechanical scattering pattern does indeed depend on Φ, in a manner confirmed experimentally by Chambers (1960) and Möllenstedt and Bayh (1962). This result, and its implication that in quantum mechanics the vector potential has a direct physical significance (as opposed to classical mechanics where it is a mathematical device), have become known as the Aharonov–Bohm (AB) effect.

The AB effect has given rise to controversy, to which we shall devote a few remarks at the end of the article. Our main purpose, however, is to describe and illustrate some little-known properties of the AB wavefunction. In §2 we explain that, far from the cylinder, the incident and scattered waves can be distinguished everywhere except near the forward direction, where they are inextricably connected; without this connection both incident and scattered waves would be multivalued. The results are most transparently obtained by some unusual asymptotics described in the Appendix.

Section 3 is the heart of the paper, and concerns the wavefronts, i.e. the lines of constant phase of

the wavefunction. Their behaviour depends on the quantum flux parameter α, defined by

$$\alpha = q\Phi/h, \qquad (2)$$

where h is Planck's constant. We show that a number of wave crests, equal to the nearest integer to α, end on the cylinder containing the flux, which is therefore a 'wavefront dislocation' in the sense of Nye and Berry (1974). In quantum mechanics these phase singularities are unobservable, since as shown by Wu and Yang (1975) all observables (such as the AB scattering cross section) depend not directly on α but on $\exp(2\pi i\alpha)$, and so $\alpha + 1$ and α are indistinguishable.

The presence of interesting but unobservable topology in the wavefunction prompted us to devise an analogue system in which the wavefronts can be seen directly. We show in §4 that surface waves on water crossing an irrotational (bathtub) vortex constitute such a system. Qualitative and quantitative experimental confirmation of the predicted properties of the wavefront dislocations are presented in §5.

2 Aharonov–Bohm wavefunction

Let the incident particles be represented by a plane wave with wavenumber k incident from $x = +\infty$, $\theta = 0$ (figure 1) on an impenetrable cylinder of radius R centred on the z axis and containing the flux Φ. A suitable vector potential yielding zero field outside the cylinder and satisfying (1) is

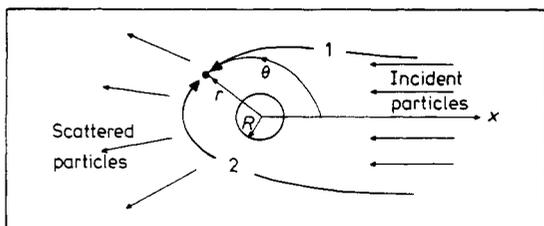$$\mathbf{A}(\mathbf{r}) = (\Phi/2\pi r)\hat{\boldsymbol{\theta}}, \qquad (3)$$

where $\hat{\boldsymbol{\theta}}$ is the azimuthal unit vector. Of course infinitely many other vector potentials are possible, related to (3) by gauge transformations; a careful discussion of these is given by Ingraham (1972).

The wavefunction $\psi(\mathbf{r})$ must satisfy the following conditions:
(i) Schrödinger's equation:

$$\frac{1}{2m}(-i\hbar\nabla - q\mathbf{A}(\mathbf{r}))^2\psi(\mathbf{r}) = \frac{\hbar^2 k^2}{2m}\psi(\mathbf{r}), \qquad (4)$$

with $\mathbf{A}$ given by (3).

Figure 1 Scattering geometry and coordinates for the Aharonov–Bohm effect, showing two paths reaching the same point by topologically different routes.

(ii) Single-valuedness (for detailed discussions of this requirement see Merzbacher (1962), Tassie and Peshkin (1961) and Kretschmar (1965a)).
(iii) Impenetrability:

$$\psi = 0 \qquad \text{when } r = R. \qquad (5)$$

(iv) Asymptotics: as $r \to \infty$, ψ must be the sum of the incident wave plus a purely outgoing wave; the mathematical expression of this condition will be discussed later.

To see how subtle the problem is we first describe a tempting but unprofitable approach. Consider the partial-wave expansion of the wave $\psi_0(\mathbf{r})$ corresponding to scattering by the cylinder containing zero flux, namely

$$\psi_0(\mathbf{r}) = \sum_{l=-\infty}^{\infty} \frac{(-i)^{|l|}\exp(il\theta)}{H_{|l|}^{(1)}(kR)}$$
$$\times \{J_{|l|}(kr)H_{|l|}^{(1)}(kR) - J_{|l|}(kR)H_{|l|}^{(1)}(kr)\}, \qquad (6)$$

where J and H denote the usual Bessel functions (Abramowitz and Stegun 1964). This satisfies conditions (i)–(iv) with $\Phi = 0$. When $\Phi \neq 0$, it is easily shown by direct substitution that Schrödinger's equation (4) is satisfied by

$$\psi(\mathbf{r}) = \psi_0(\mathbf{r})\exp\left\{iq\int^r \mathbf{A}(\mathbf{r}')\cdot d\mathbf{r}'/\hbar\right\} = \psi_0(\mathbf{r})\exp(i\alpha\theta), \qquad (7)$$

where α is the flux parameter (equation (2)). This also satisfies conditions (iii) and (iv) but violates (ii) because it is not single-valued unless α is an integer. Therefore it is not the correct solution.

The failure of single-valuedness should not be interpreted as a requirement that α must be quantised (as in the case of superconductivity, for example, where different physical principles operate—see Merzbacher (1962)). Instead, the multivalued elementary solutions

$$\exp\{i(l+\alpha)\theta\}J_{|l|}(kr) \quad \text{and} \quad \exp\{i(l+\alpha)\theta\}H_{|l|}^{(1)}(kr) \qquad (8)$$

must be replaced by the single-valued solutions

$$\exp(il\theta)J_{|l-\alpha|}(kr) \quad \text{and} \quad \exp(il\theta)H_{|l-\alpha|}^{(1)}(kr). \qquad (9)$$

Equation (6) can now be modified to give the correct solution, which we write as

$$\psi(\mathbf{r}) = \psi_{AB}(\mathbf{r}) - \psi_R(\mathbf{r}) \qquad (10)$$

where

$$\psi_{AB}(\mathbf{r}) = \sum_{l=-\infty}^{\infty} (-i)^{|l-\alpha|}\exp(il\theta)J_{|l-\alpha|}(kr) \qquad (11)$$

and

$$\psi_R(\mathbf{r}) = \sum_{l=-\infty}^{\infty} (-i)^{|l-\alpha|}\exp(il\theta)\frac{J_{|l-\alpha|}(kR)}{H_{|l-\alpha|}^{(1)}(kR)}H_{|l-\alpha|}^{(1)}(kr). \qquad (12)$$

Since this solution is a superposition of functions of the form of solution (9), it obviously satisfies conditions (i) and (ii); moreover it also satisfies (iii). As the cylinder radius R tends to zero, ψ_R vanishes, leaving ψ_{AB} as the wave in the presence of a single impenetrable flux line. But ψ_{AB} is the wavefunction derived by Aharonov and Bohm (1959), so this argument justifies their claim to have found the correct limiting form as $R \to 0$. A solution more general than equation (10), in that it allows the cylinder to be penetrable, can be found in the appendix of a paper by Kretschmar (1965b).

Now we must show that the wave ψ satisfies the asymptotic condition (iv). Consider first ψ_R: because of the Bessel functions with argument kR, the series converges rapidly for $|l - \alpha| > kR$, and its behaviour as $r \to \infty$ can be found using standard asymptotic forms for $H^{(1)}_{|l-\alpha|}(kr)$, giving

$$\psi_R(r) \xrightarrow{r \to \infty} \left(\frac{2}{\pi i k r}\right)^{1/2} \exp(ikr)$$

$$\times \sum_{l=-\infty}^{\infty} \exp\{i(l\theta - \pi|l - \alpha|)\} \frac{J_{|l-\alpha|}(kR)}{H^{(1)}_{|l-\alpha|}(kR)}. \quad (13)$$

This represents a purely outgoing wave, consistent with condition (iv).

The asymptotics of ψ_{AB} are more subtle. In view of the absence of 'outgoing' Bessel functions $H^{(1)}$ from equation (11), it is far from obvious that ψ_{AB} can represent just the wave scattered by the flux line in addition to the incident wave. That this is indeed the case is shown in the Appendix using arguments simpler than those in the original AB paper. The result is

$$\psi_{AB}(r) \xrightarrow[r \to \infty]{|\theta| < \pi - O((kr)^{-1/2})} \exp(-ikr \cos\theta + i\alpha\theta)$$

$$+ \frac{\exp(ikr) \sin(\pi\alpha)}{(2\pi i k r)^{1/2} \cos(\theta/2)} (-1)^{[\alpha]} \exp\{i([\alpha] + \tfrac{1}{2})\theta\}, \quad (14)$$

where $[\alpha]$ denotes the integer part of α.

To see that the first term of equation (14) correctly represents the incident wave, recall that this must correspond to a probability current of particles directed along $-x$. The current is

$$j(r) = \frac{\hbar}{m} \text{Im}(\psi^* \nabla \psi) - \frac{q}{m} A(r) |\psi|^2, \quad (15)$$

which is the expectation value of the velocity density operator

$$v_{op}(r) \equiv \frac{1}{2m} \{(p_{op} - qA_{op})\delta(r - r_{op})$$

$$+ \delta(r - r_{op})(p_{op} - qA_{op})\}. \quad (16)$$

When applied to equation (14), equation (15) yields the correct result

$$j_{AB}(r) \xrightarrow{r \to \infty} -\frac{\hbar k}{m} \hat{x}. \quad (17)$$

The second term of equation (14) describes a purely outgoing wave. Therefore the solution (10) does indeed satisfy all the conditions (i)–(iv).

It appears from equation (14) that the incident and scattered waves are multivalued functions of θ, but this is not the case because a narrow sector near the forward direction $|\theta| = \pi$ is excluded. We show in the Appendix that within this sector ψ_{AB} cannot be separated into incident and scattered parts, and that in the forward direction itself

$$\psi_{AB}(r, \pm\pi) \xrightarrow{r \to \infty} \exp\{i(kr + [\alpha]\pi)\} \cos\{\pi(\alpha - [\alpha])\}. \quad (18)$$

Thus when $\alpha = N + \tfrac{1}{2}$, where N is an integer, ψ_{AB} vanishes at $|\theta| = \pi$. This is actually a general result, valid for all r and for both ψ_{AB} and ψ_R, i.e.

$$\psi(r, \pm\pi) = 0 \qquad \text{when } \alpha = N + \tfrac{1}{2}. \quad (19)$$

An exact representation of ψ_{AB} when $\alpha = N + \tfrac{1}{2}$, valid for all r and θ, is given by equation (A9) of the Appendix. The existence of a *nodal line* stretching from 0 to ∞ when $\alpha = N + \tfrac{1}{2}$, and the stitching together of incident and scattered waves near $|\theta| = \pi$ for all α, will play an important part in §3.

It is easily verified from solutions (10), (11) and (12) that ψ has the following symmetry properties:

$$\psi(r, \theta; \alpha + N) = \exp(iN\theta)\psi(r, \theta; \alpha), \quad (20)$$

$$\psi(r, \theta; -\alpha) = \psi(r, -\theta; \alpha). \quad (21)$$

These imply that all measurements of the intensity $|\psi|^2$ are periodic in α, that what is observed at θ when $\alpha = \tfrac{1}{2} + \delta$ will be observed at $-\theta$ when $\alpha = \tfrac{1}{2} - \delta$, and that

$$|\psi(r, \theta)|^2 = |\psi(r, -\theta)|^2 \qquad \text{when } 2\alpha = N. \quad (22)$$

Therefore when studying $|\psi|^2$ it is sufficient to consider $0 \leq \alpha \leq \tfrac{1}{2}$.

As α varies, the interference pattern of waves scattered by the cylinder changes. When $kR \to 0$ this pattern is simply the AB scattering cross section obtained from equation (14), namely

$$\sigma_{AB}(\theta) = \frac{\sin^2(\pi\alpha)}{2\pi k \cos^2(\theta/2)}. \quad (23)$$

When $kR \gg 1$, however, the interference pattern is a complicated superposition of contributions from direct, reflected and creeping rays (Keller 1958). In the general case, the primitive notion of 'interference fringes' does not apply. But sometimes, if the pattern is locally the resultant of two contributions of approximately equal strength, it is legitimate to speak about fringes, as in elementary discussions of the AB effect. Then the naive superposition of the contributions, with the magnetic phase difference $\exp\{iq \int^r A(r') \cdot dr'/\hbar\}$ as in equation (7), can yield the correct result. For example, consider angular fringes whose contributions have wave number m

and have gone round opposite sides of the cylinder (paths 1 and 2 on figure 1). The wave beyond the cylinder is then

$$\psi \propto \exp\{i(m\theta + \alpha\theta)\}$$
$$+ \exp\{i(-m(-2\pi + \theta) + \alpha(-2\pi + \theta))\} \quad (24)$$

so that

$$|\psi|^2 \propto 4\cos^2\{m(\theta - \pi) + \alpha\pi\}. \quad (25)$$

This angular fringe pattern shifts periodically with α in precisely the manner observed in experiments.

3 Wavefront dislocations

From now on we ignore ψ_R and consider only ψ_{AB}, in particular its wavefronts. These are defined as lines of constant phase χ of the complex wave

$$\psi_{AB}(r) \equiv |\psi_{AB}(r)| \exp(i\chi(r)) \quad (26)$$

in the plane r, θ. Wave crests are particular wavefronts, defined by

$$\chi(r) = 2M\pi, \quad (27)$$

where M is an integer. Even though $\psi_{AB}(r)$ is a single-valued function, $\chi(r)$ may be multivalued, in the following sense: during a circuit C in the r plane, χ may change by an integer multiple S_C of 2π, i.e.

$$S_C = \frac{1}{2\pi} \oint_C d\chi = \frac{1}{2\pi} \oint_C \nabla\chi \cdot dr. \quad (28)$$

Within C, S_C wave crests must come to an end. Points at which this happens are singularities of $\chi(r)$, called wavefront dislocations by Nye and Berry (1974) by analogy with dislocations of atomic planes in crystals. At a wavefront dislocation, the modulus $|\psi_{AB}|$ must vanish. S_C will be called the dislocation strength within C.

We now study the dislocation structure of ψ_{AB}, beginning with a calculation of S_C for an anticlockwise circuit C consisting of a very large circle
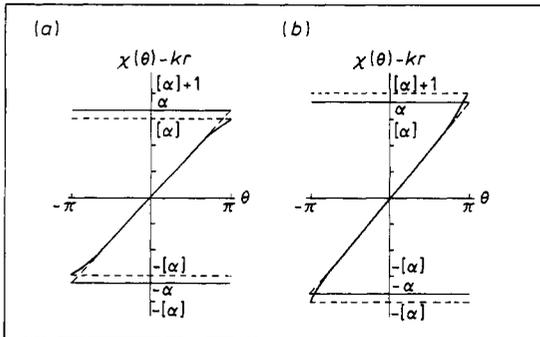
Figure 2 Asymptotic phase $\chi - kr$ for ψ_{AB} as a function of θ. (a) $\alpha - [\alpha] < \frac{1}{2}$. (b) $\alpha - [\alpha] > \frac{1}{2}$.

surrounding the flux line. When the flux parameter is an integer, that is $\alpha = N$, there is no scattered wave and ψ_{AB} is given exactly for all θ (and also in this case for all r) by the first term of equation (14). The phase is then

$$\chi = -kr\cos\theta + N\theta \quad (\alpha = N) \quad (29)$$

whose total change round C is $2\pi N$, so that from equation (28) $S_C = N$. When α is not an integer, the first term of equation (14) dominates everywhere except a narrow angular sector centred on $\theta = \pi$ (see the Appendix, equation (A8)), with width $2\Delta\theta$. Between $-\pi + \Delta\theta$ and $\pi - \Delta\theta$ the phase accumulation is $2\pi\alpha$, which is not an integer. Within this sector, however, χ is not given by equation (14), and when $\theta = \pi$ equation (18) shows that

$$\left.\begin{aligned}
\chi - kr &= [\alpha]\pi \bmod 2\pi \\
&\quad \text{when } \theta = \pi, \alpha - [\alpha] < \tfrac{1}{2}, \\
&= ([\alpha] + 1)\pi \bmod 2\pi \\
&\quad \text{when } \theta = \pi, \alpha - [\alpha] > \tfrac{1}{2}.
\end{aligned}\right\} \quad (30)$$

Figure 2 shows the phase functions implied by this result, defining $\chi + kr$ as zero when $\theta = 0$ and interpreting the 'mod 2π' additions to give continuity with equation (29). It follows that the total phase change round C is $2\pi l_\alpha$, where l_α is the integer closest to α, so that

$$S_C = l_\alpha. \quad (31)$$

Now let C be a very small circle surrounding the flux line. Then $r \to 0$ and the series (11) for ψ_{AB} is dominated by the Bessel function of lowest order, for which l is the integer closest to α. All the angular dependence is in the term $\exp(il\theta)$, so that S_C is again given by equation (31).

The simplest picture consistent with these results is of a wavefront dislocation at the flux line, where S_C wave crests emerge and extend out to the far field. All other wave crests are continuous (from $\theta = -\pi/2$ to $\theta = +\pi/2$). As α varies through $N + \frac{1}{2}$, l_α and hence S_C jump by unity, so that a new wave crest appears or disappears at $r = 0$. The mechanism of this change in S_C when $\alpha = N + \frac{1}{2}$ is a disconnection and reconnection of wavefronts on a nodal line along the negative x axis (equation (19)).

Wave crest patterns are illustrated in figure 3(a) for some integer values of α, and in figure 3(b) for a sequence of values of α between 1 and 2. The pattern for $\alpha = 1\frac{1}{2}$ was drawn with the aid of the exact representation (A9) in the Appendix.

It is interesting to study the probability current j close to the flux line. According to equations (15), (26) and (3),

$$j = \frac{\hbar}{m} |\psi_{AB}|^2 \left(\nabla\chi - \frac{\alpha\hat{\theta}}{r}\right), \quad (32)$$

so that the current is not perpendicular to the wavefronts. Calculations based on equation (11) give
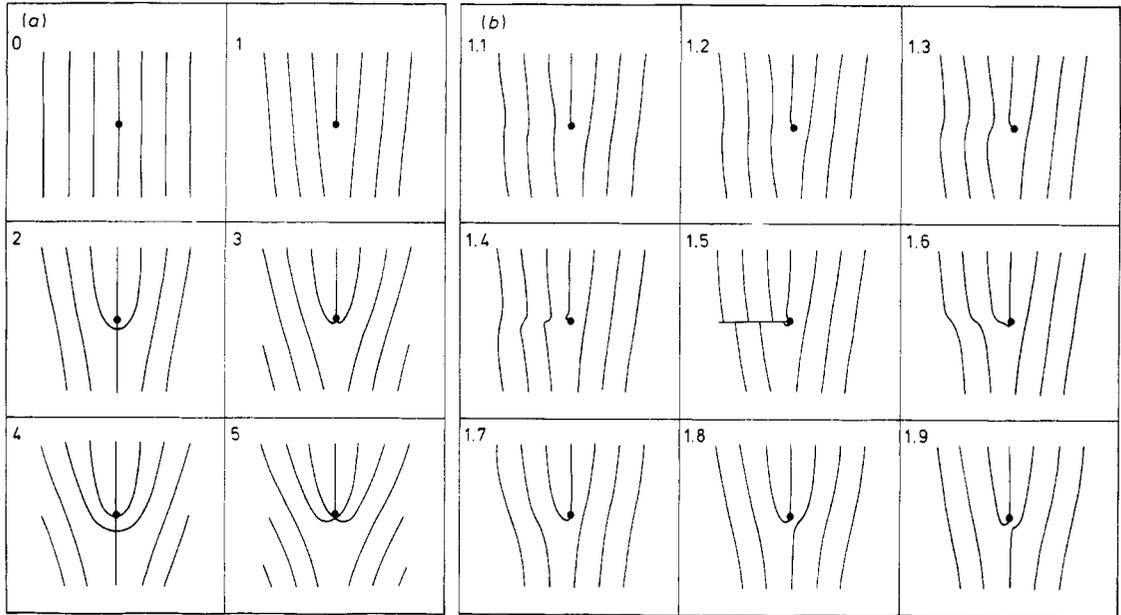
Figure 3 Wave crests of $\exp\{-i\alpha/2\}\psi_{AB}$. (a) Integer values of α from 0 to 5. (b) Values of α between 1 and 2, showing unlinking and reconnection as α passes through $1\frac{1}{2}$. In all cases the waves are incident from the right, and the flux parameter corresponds to a magnetic field at the origin and pointing out of the paper. The wave crests were drawn by interpolating between calculated positions for ($r \to \infty$, $\theta = \pm\pi/2$) and on the x axis, taking account of the calculated directions of the wave crests at $r = 0$.

$$j \to \left. \begin{array}{l} \dfrac{K(l_\alpha - \alpha)\hat{\boldsymbol{\theta}}}{r^{1-2|l_\alpha-\alpha|}} \\[2mm] \text{when } r \to 0, \ \alpha \text{ not near } N+\tfrac{1}{2}; \\[2mm] \to 2K(\alpha - [\alpha] - \tfrac{1}{2})\hat{\boldsymbol{\theta}} \ \ln(1/kr) \\[2mm] \text{when } r \to 0, \ \alpha \text{ near } N+\tfrac{1}{2}, \end{array} \right\} \quad (33)$$

where K is a constant. Therefore the particle current circulates around the flux line, clockwise when $[\alpha] < \alpha < [\alpha] + \tfrac{1}{2}$ and anticlockwise when $[\alpha] + \tfrac{1}{2} < \alpha < [\alpha] + 1$; j vanishes for integral or half-integral values of the flux parameter.

The currents just described are invariant under gauge transformations to a different vector potential satisfying equation (1). The wavefront patterns (figure 3) change, however, but their topology is invariant: gauge transformations can cause arbitrary deformations of the wavefronts, but the strength S_C of the dislocation on the flux line does not change. S_C can be regarded as a 'topological quantum number', conserved under any classically permissible gauge transformation. The disappointing fact, however, is that this striking feature of ψ_{AB} is *unobservable*. This conclusion is not a consequence of the dislocation being a phase property, because phase in the position representation can affect amplitude in the momentum representation. Rather, it is a consequence of the following version of an argument given by Wu and Yang (1975).

Experiments can measure only expectation values of operators corresponding to observables. The canonical momentum is not an observable, but the velocity

$$\boldsymbol{v} = \boldsymbol{p} - q\boldsymbol{A} \quad (34)$$

is. In the state $|\psi\rangle$, any function $f(\boldsymbol{v}_{\mathrm{op}})$ has expectation value

$$\langle\psi| f(\boldsymbol{v}_{\mathrm{op}}) |\psi\rangle = \int d\boldsymbol{r}\psi^*(\boldsymbol{r})f(-i\hbar\nabla - q\boldsymbol{A}(\boldsymbol{r}))\psi(\boldsymbol{r}). \quad (35)$$

In the case we have been considering, $\boldsymbol{A}$ is given by equation (3). Now let the flux parameter change from α to $\alpha' = \alpha + 1$. According to the solutions (10), (11) and (12) and equation (34), ψ and $\boldsymbol{v}_{\mathrm{op}}$ change to

$$\left. \begin{array}{l} \psi' = \exp(i\theta)\psi \\[1mm] \boldsymbol{v}' = \boldsymbol{p} - q\boldsymbol{A}' = \boldsymbol{p} - q\boldsymbol{A} - \hbar\hat{\boldsymbol{\theta}}/r \end{array} \right\} \quad (36)$$

The expectation value of $f(\boldsymbol{v}'_{\mathrm{op}})$ is now

$$\langle\psi'| f(\boldsymbol{v}'_{\mathrm{op}}) |\psi'\rangle = \int d\boldsymbol{r}\psi^*(\boldsymbol{r}) \exp(-i\theta)$$
$$\times f(-i\hbar\nabla - q\boldsymbol{A}(\boldsymbol{r}) - \hbar\hat{\boldsymbol{\theta}}/r) \exp(i\theta)\psi(\boldsymbol{r}), \quad (37)$$

which is easily shown to equal equation (35).

This means that not only the intensity $|\psi|^2$ but all observable quantities are unaffected by changing α by an integer. All that can be measured is the deviation of the flux parameter from the nearest integer. The integer itself, which is precisely the dislocation strength S_C, cannot be observed: there is no 'dislocation strength operator' whose eigenvalues are S_C.

4 The Aharonov–Bohm effect for water waves

In seeking a system where the dislocation in ψ_{AB} can be observed, we first note an analogy between waves in the presence of a vector potential and waves in a moving medium. Such an analogy (albeit somewhat different from the one we shall present) was originally suggested to us by J H Hannay.

Let waves with frequency Ω and wavevector $\boldsymbol{k}$ propagate in a stationary isotropic medium (which may be inhomogeneous). The dispersion relation (i.e. the Hamiltonian) is

$$\Omega = \omega(k, \boldsymbol{r}), \qquad (38)$$

and depends only on the length k of $\boldsymbol{k}$. In a field with vector potential $\boldsymbol{A}$, the dispersion relation becomes

$$\Omega = \omega(|\boldsymbol{k} - q\boldsymbol{A}(\boldsymbol{r})/\hbar|, \boldsymbol{r}). \qquad (39)$$

If instead the medium is moving, with flow velocity $\boldsymbol{U}(\boldsymbol{r})$, then a plane wave is described locally not by

$$\psi = \exp\{i(\boldsymbol{k} \cdot \boldsymbol{r} - \omega(k, \boldsymbol{r})t)\} \qquad (40)$$

but by

$$\psi = \exp\{i(\boldsymbol{k} \cdot (\boldsymbol{r} - \boldsymbol{U}(\boldsymbol{r})t) - \omega(k, \boldsymbol{r})t)\}$$
$$= \exp\{i(\boldsymbol{k} \cdot \boldsymbol{r} - (\omega(k, \boldsymbol{r}) + \boldsymbol{k} \cdot \boldsymbol{U}(\boldsymbol{r}))t)\}, \quad (41)$$

so that the dispersion relation in the moving medium is

$$\Omega = \omega(k, \boldsymbol{r}) + \boldsymbol{k} \cdot \boldsymbol{U}(\boldsymbol{r}). \qquad (42)$$

To lowest order, this can be written as

$$\Omega \approx \omega\left(\left|\boldsymbol{k} + \frac{k\boldsymbol{U}(\boldsymbol{r})}{v_g(k, \boldsymbol{r})}\right|, \boldsymbol{r}\right) \qquad (43)$$

where v_g is the group velocity

$$v_g = \partial\omega(k, \boldsymbol{r})/\partial k. \qquad (44)$$

The step from equation (42) to equation (43) is valid provided

$$|\boldsymbol{U}| \ll v_g. \qquad (45)$$

Comparing equations (43) and (39), we see that the effect of a slowly moving medium is the same as that of a vector potential, the precise analogy being

$$\frac{q\boldsymbol{A}(\boldsymbol{r})}{\hbar} \leftrightarrow \frac{-k\boldsymbol{U}(\boldsymbol{r})}{v_g(k, \boldsymbol{r})}. \qquad (46)$$

From equations (1) and (2), the analogue of the quantum flux parameter, in the case where the stationary medium is homogeneous, is

$$\alpha \leftrightarrow \frac{-\oint \boldsymbol{U} \cdot d\boldsymbol{r}}{\lambda v_g} = \frac{-\omega \oint \boldsymbol{U} \cdot d\boldsymbol{r}}{v_p v_g} \qquad (47)$$

where λ is the wavelength and v_p the phase velocity ω/k.

In the AB effect, the magnetic field vanishes outside the central flux line. Therefore the velocity field $\boldsymbol{U}$ analogous to $\boldsymbol{A}$ must be curl-free, i.e. the waves must propagate on a medium that is flowing irrotationally but with non-zero circulation. This can be achieved with surface waves on swirling water, if the swirling takes the form of an irrotational ('bathtub') vortex. If the density, surface tension and depth of the water are denoted by ρ, γ and d respectively, and if g denotes the acceleration due to gravity, the dispersion law is (Lamb 1945)

$$\omega = \left\{\left(gk + \frac{\gamma k^3}{\rho}\right)\tanh(kd)\right\}^{1/2}. \qquad (48)$$

This analogy provides a means of testing the predictions of §3 concerning wavefront dislocations, because for water waves, as opposed to quantum mechanical waves, the crests can be observed. For the analogy to hold good, the condition (45) must be satisfied, but this will always be the case far from the vortex because $|\boldsymbol{U}| \to 0$ there; the dislocation at $r = 0$ can be identified by counting wave crests at large r. According to equation (47), the patterns in figure 3 correspond to water circulating clockwise, so that the wavefronts should be more closely spaced where the waves travel against the current than when they travel with the current, as expected on the basis of the Doppler effect.

5 Experiment†

A rectangular perspex tank was constructed, with dimensions $1.0\,\text{m} \times 0.6\,\text{m} \times 0.15\,\text{m}$. Surface waves were excited on water in the tank by vibrating a straight horizontal dipper $0.15\,\text{m}$ long connected to a variable-speed electric motor. The useful frequency range was from 7 to 70 Hz. For the lower frequencies the gravity-wave term gk in equation (48) dominated, whilst for the higher frequencies the surface-tension ripple term $\gamma k^3/\rho$ dominated; the water was sufficiently deep for the term $\tanh(kd)$ to be set equal to unity for all frequencies.

A vortex formed spontaneously on letting the water pour out through a hole 6 mm in diameter in the middle of the bottom of the tank. For the wave-making dipper to operate efficiently it was essential to maintain a constant water level; this

† The results reported in this section were obtained by two of us (MDL and JCW) in an undergraduate research project.

was achieved by pumping the outflowing water back into the tank, taking care to let the pumped water re-enter via submerged perforated tubes, so as to inhibit the development of bulk rotation.

In order to 'freeze' the motion of the wave crests, the water surface was illuminated from below by a stroboscope set to the wave frequency. The light refracted by the wave crests formed bright caustic lines on a screen just above the water. The patterns on the screen were photographed; figure 4 shows a series of such pictures for dislocation strengths ranging from 0 to 3.

According to §3 the dislocation structure of the waves depends on the flux parameter α. For water waves this quantity is given by equation (47), which shows that in order to predict the dislocation strength it is necessary to know not only the angular frequency ω of the waves but also the circulation $\oint U \cdot dr$ of the vortex. To measure the circulation, small paper discs were allowed to circulate on the water surface near the vortex, and photographed under continuous illumination with an exposure of 0.5 s. The circulating paper discs gave tracks whose lengths were used to estimate the water velocity and hence the circulation. With the dipper running, the streamlines were approximately circular to a distance of about 50 mm from the vortex core. Each photograph contained about ten useful tracks, and measurement gave the same circulation for these tracks to an accuracy of about 20%, which enabled α to be estimated with a standard error of less than 10%. On the other hand, inspection of the photographs of wave crests, and comparison with the results of §3, enabled α to be estimated with an absolute accuracy of about 0.25 (figure 4, cf figure 3). More than fifty comparisons of the values of α obtained by these two different methods were made, using frequencies between 8 Hz and 67 Hz to generate the waves, and a range of water outflow rates to generate the vortex. The values of α ranged from 0 to 2, and in every case the two methods gave complete agreement within the quoted accuracy, thus confirming the theory of §3 and §4.

6 Discussion

From a mathematical point of view, there can be no doubt that ψ_{AB}, given by equation (11), is the correct solution of the AB wave equation (4). Moreover ψ_{AB} correctly models observable features of waves in physical systems: the periodic shifting of interference fringes as the flux varies, confirmed by Chambers (1960) and Möllenstedt and Bayh (1962), and the changing of the wavefront topology as the circulation varies, which we have confirmed and studied (§5). But the physical interpretation of the mathematics and the experiments continues to cause controversy. The question is, are the predicted and observed fringe shifts really effects of potentials in the absence of electromagnetic fields

accessible to the particles, or is it possible to explain them entirely in terms of accessible fields?

Weisskopf (1961) considers the flux to be switched on from zero to its final value Φ. He shows that the evolution of the wavefunction to ψ_{AB} can be explained entirely in terms of the accessible electric field induced outside the cylinder, even if this field is made arbitrarily small by changing the flux very slowly. Casati and Guarneri (1979) consider the cylinder to be slightly penetrable, and modelled by a high potential V for $r < R$. They show that for any finite V, however large, the fringe shifts can be explained entirely in terms of the magnetic field within the cylinder, which is now accessible to the particles (their analysis is in terms of the hydrodynamical formulation of quantum mechanics, rather than the wave equation, but their conclusions do not depend on this). Roy (1980) considers the cylinder to be a solenoid of finite length L. He shows that no matter how large L is, a gauge can be found in which the vector potential can be expressed entirely in terms of the weak magnetic field leaking out of the ends of the cylinder into the space accessible to the particles.

In view of these arguments we must agree that if any electromagnetic fields, however small, are or have ever been accessible to the particles, then the state of the particles, in quantum as well as classical mechanics, can be described entirely in terms of these fields. It could be argued that since such fields always exist in practice, there is no AB effect, i.e. no observable consequence of inaccessible fields. But in view of the fact that the wave always tends to ψ_{AB}, no matter how the limiting process is performed, those who argue thus must also believe that infinitesimal causes (fields) can produce finite effects (fringe shifts). On the other hand, if we consider the vector potential (as embodied in the gauge-invariant and quantum-mechanically observable deviation of the flux parameter from an integer) as a primary causative agent, then there is no infinite discordance between the magnitudes of causes and effects in the limit of inaccessibility. We therefore consider the question of the reality of the AB effect to be metaphysical and devoid of observational implication.

Appendix
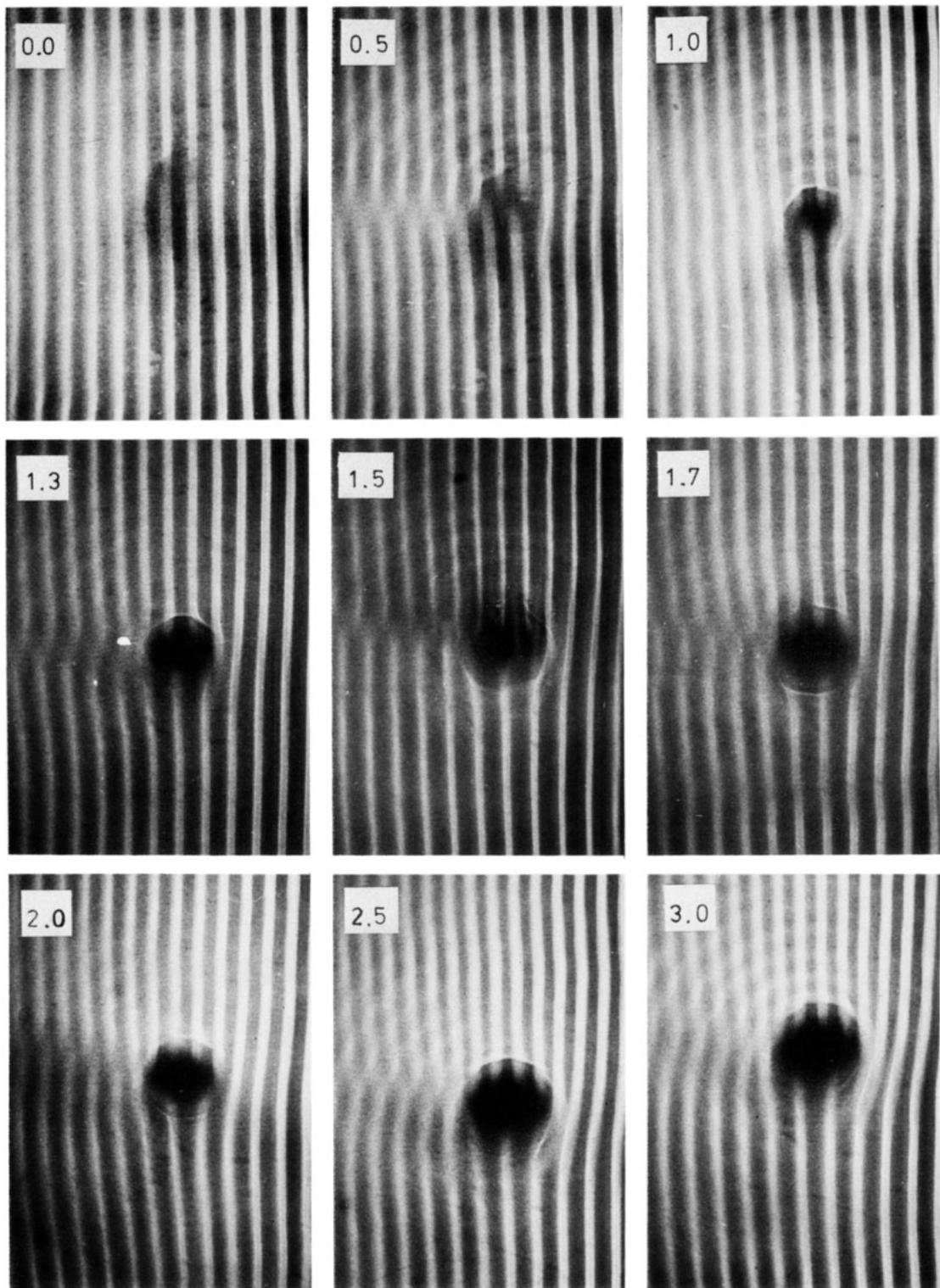The first step in studying the asymptotics of ψ_{AB} as defined by equation (11) is to replace the Bessel functions by the integral representation

$$J_\nu(z) = \frac{1}{2\pi} \int_C \exp\{i(\nu t - 2 \sin t)\} \, dt \qquad (A1)$$

(Gradshteyn and Ryzhik 1965), where C is the contour shown in figure A1(*a*). Then equation (11) becomes

Figure 4 Water wave crests passing an irrotational vortex, giving rise to different strengths of wavefront dislocation. Waves are incident from the right and the water is circulating clockwise. Estimated values of the flux parameter α are indicated.
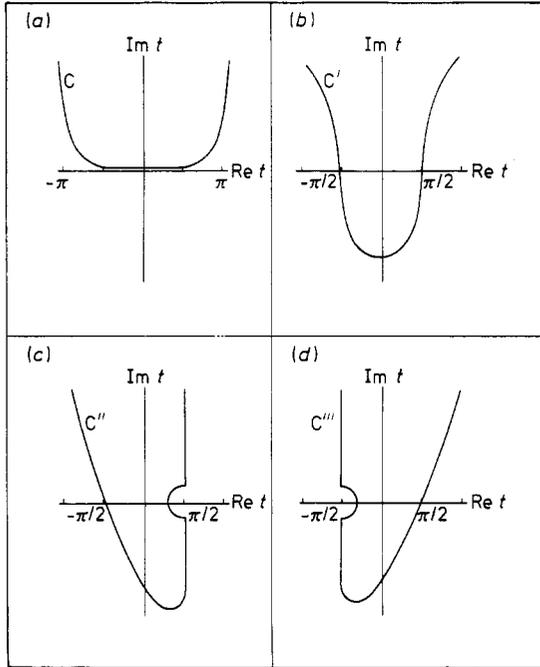
Figure A1 Integration contours in the *t* plane.

$$\psi_{AB} = \frac{1}{2\pi} \int_C dt \, \exp(-ikr \sin t)$$
$$\times \sum_{l=-\infty}^{\infty} \exp\left\{i\left(l\theta - |l-\alpha|\frac{\pi}{2} + |l-\alpha|t\right)\right\}. \quad (A2)$$

The sum converges because Im $t > 0$, and can be evaluated by splitting the range into $l \le [\alpha]$ and $l \ge [\alpha] + 1$, to give

$$\psi_{AB} = \frac{1}{2\pi} \int_C dt \, \exp(-ikr \sin t)$$
$$\times \left(\frac{\exp\{i([\alpha]\theta + (t - \pi/2)(\alpha - [\alpha]))\}}{1 - \exp(i(t - \pi/2 - \theta))}\right.$$
$$\left. + \frac{\exp\{i([\alpha] + 1)\theta + (t - \pi/2)([\alpha] - \alpha + 1))\}}{1 - \exp(i(t - \pi/2 + \theta))}\right). \quad (A3)$$

For large kr the integral is dominated by its poles and saddle points. The saddle points occur where $(d/dt)(kr \sin t) = 0$, i.e. at $t = \pm \pi/2$. The integration contour can be deformed into C' (figure A1(b)) so as to pass through these. During the deformation, the contour crosses a pole, whose contribution must be included. The first term in (A3) has a pole if $t - \pi/2 - \theta = 0$, and this contributes if $-\pi < \theta < 0$; the second term in (A3) has a pole if $t - \pi/2 + \theta = 0$, and this contributes if $0 < \theta < \pi$. Both contributions are of the same form, namely

$$\psi_{AB}(\text{pole}) = \exp\{i(-kr \cos \theta + \alpha\theta)\} \qquad (0 < |\theta| < \pi). \quad (A4)$$

This is precisely the incident wave in equation (14).

Next, the contributions to the integral over C' from the saddle points at $t = \pm \pi/2$ must be evaluated by the method of steepest descents (De Bruijn 1970). A straightforward calculation for $t = -\pi/2$ gives

ψ_{AB}(saddlepoint at $t = -\pi/2$)

$$= \frac{\exp(ikr) \sin(\pi\alpha)}{(2\pi i kr)^{1/2} \cos(\theta/2)} (-1)^{[\alpha]} \exp\{i([\alpha] + \tfrac{1}{2})\theta\}. \quad (A5)$$

This is precisely the outgoing scattered wave in equation (14). The saddle point at $t = +\pi/2$ would give an incoming 'scattered' wave, but substitution into equation (A3) shows that the coefficient of this wave is zero.

Our arguments leading to the asymptotic form (14) for ψ_{AB} fail when θ is close to 0 or π, because then the pole is close to one of the saddle points and the contributions of these two points cannot be separated.

In the backward direction $\theta = 0$, the saddle point at $\pi/2$, which gave no incoming wave, coincides with a pole in each term of (A3); the contour C'' on figure A1(c) gives a pole contribution identical with (A4) (with $\theta = 0$), as well as a principal-value integral which vanishes as $kr \to \infty$. Therefore equation (14) remains valid as $\theta \to 0$.

In the forward direction $\theta = \pi$, the saddle point at $t = -\pi/2$, which gave rise to the outgoing wave (A5), coincides with a pole in each term of (A3); the contour C''' on figure A1(d) gives a pole contribution

ψ_{AB}(saddle point and pole at $t = -\pi/2$)

$$= \exp\{i(kr + \pi[\alpha])\} \cos\{\pi(\alpha - [\alpha])\}, \qquad \theta = \pi, \quad (A6)$$

together with a principal-value integral which vanishes as $r \to \infty$. This justifies equation (18).

Next we must estimate the angular width of the 'diffraction shadow', near the forward direction, within which (A4) and (A5) are not valid. The 'domain' Δt of the stationary point at $t = -\pi/2$ is given by

$$\Delta(kr \sin t) \sim 1 \text{ radian}, \quad \text{i.e.} \quad \Delta t \sim (2/kr)^{1/2}. \quad (A7)$$

In order for a pole of (A3) to give a separate contribution, it must lie outside the domain Δt. This excludes the sector of width $2\Delta\theta$ defined by

$$\Delta\theta = \pi - |\theta| < O(kr)^{-1/2}, \quad (A8)$$

hence the restriction on $|\theta|$ in equation (14).

For the important transitional case of half-integer flux, equation (A3) can, after some calculation, be reduced to the simple form

$$\psi_{AB}(r, \theta) = \frac{2}{\pi^{1/2}} \exp\{i((N + \tfrac{1}{2})\theta - kr \cos \theta - \pi/4)\}$$
$$\times \int_0^{(2kr)^{1/2}\cos(\theta/2)} \exp(it^2) \, dt \qquad \text{when } \alpha = N + \tfrac{1}{2}, \quad (A9)$$

which is (apart from a slight correction) the same as equation (23) in the original paper by Aharonov and Bohm (1959). It is evident from (A9) that ψ_{AB} is single-valued.

References

Abramowitz M and Stegun I A (eds) 1964 *Handbook of Mathematical Functions* (New York: Dover)

Aharonov Y and Bohm D 1959 *Phys. Rev.* **115** 485–91

Casati G and Guarneri I 1979 *Phys. Rev. Lett.* **42** 1579–81

Chambers R G 1960 *Phys. Rev. Lett.* **5** 3–5

De Bruijn N G 1970 *Asymptotic Methods in Analysis* (Amsterdam: North-Holland)

Gradshteyn I S and Ryzhik I M 1965 *Tables of Integrals, Series, and Products* (New York: Academic)

Ingraham R L 1972 *Am. J. Phys.* **40** 1449–52
Keller J B 1958 *Proc. Symp. Appl. Math.* **8** 27–52
Kretschmar M 1965a *Z. Phys.* **185** 73–83
—— 1965b *Z. Phys.* **185** 84–96
Lamb H 1945 *Hydrodynamics* (Cambridge: CUP)
Merzbacher E 1962 *Am. J. Phys.* **30** 237–47
Möllenstedt G and Bayh W 1962 *Naturwiss.* **49** 81–2
Nye J F and Berry M V 1974 *Proc. R. Soc.* A **336** 165–90
Roy S M 1980 *Phys. Rev. Lett.* **44** 111–4
Tassie L J and Peshkin M 1961 *Ann. Phys., NY* **16** 177–84
Weisskopf V F 1961 *Lectures in Theoretical Physics* vol III, ed W E Brittin (New York: Interscience) p67
Wu T T and Yang C N 1975 *Phys. Rev.* D **12** 3845–57

Non-linear partial differential equations in physics:
I. Earlier progress

G A Toombs

Department of Physics, University of Nottingham, University Park, Nottingham NG7 2RD, England

Abstract Non-linear parital differential equations in physics are discussed using water waves in a channel as an illustration. Weakly non-linear waves, wave breaking, shocks, and solitary waves are considered in this paper. Although these topics have been understood for many years, they are unfamiliar to the majority of physicists. They are also the background to the recent advances in non-linear physics which will be discussed in a companion paper.

Résumé L'utilisation en physique des équations aux dérivées partielles non-linéaires est discutée sur l'exemple illustratif des ondes à la surface de l'eau dans un canal. Sont en particulier examinées les questions suivantes: ondes faiblement non-linéaires, déferlement, chocs, ondes solitaires. Ces questions, bien que comprises depuis longtemps, sont restées peu connues de la majorité des physiciens; elles présentent de plus l'intérêt d'être à la base des progrès récents en physique des phénomènes non linéaires, qui sont examinés dans un article du même journal.

1 Introduction

Non-linear physics has largely been ignored in the education of physicists. The reasons for this stem from the attitude of physicists to their subject. Physicists are very cautious people: they always solve the simplest problems first. They only move on to more complex problems when they have a thorough understanding of the easiest ones. Therefore, the problems studied during the early development of classical physics were linear, either by nature or design. In a linear problem, the behaviour of the unknown variables of interest is governed by partial differential equations (PDEs) which are particularly simple: the variables occur in the PDE only to first order. One of the most important properties of a linear problem is that solutions may be superposed to obtain new solutions. Superposition makes possible a general technique for