PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Semi-Classical Mechanics in Phase Space: A Study of Wigner's Function

M. V. Berry

Phil. Trans. R. Soc. Lond. A 1977 **287**, 237-271
doi: 10.1098/rsta.1977.0145

SEMI-CLASSICAL MECHANICS IN PHASE SPACE: A STUDY OF WIGNER'S FUNCTION

By M. V. BERRY†
Indian Institute of Science, Bangalore 560012, *India*

(*Communicated by J. M. Ziman, F.R.S. – Received* 4 *November* 1976)

CONTENTS

We explore the semi-classical structure of the Wigner functions $\Psi(\boldsymbol{q}, \boldsymbol{p})$ representing bound energy eigenstates $|\psi\rangle$ for systems with f degrees of freedom. If the classical motion is *integrable*, the classical limit of Ψ is a delta function on the f-dimensional torus to which classical trajectories corresponding to $|\psi\rangle$ are confined in the $2f$-dimensional phase space. In the semi-classical limit of Ψ ($\hbar$ small but not zero) the delta function softens to a peak of order $\hbar^{-\frac{2}{3}f}$ and the torus develops fringes of a characteristic 'Airy' form. Away from the torus, Ψ can have semi-classical singularities that are not delta functions; these are discussed (in full detail when $f = 1$) using Thom's theory of catastrophes. Brief consideration is given to problems raised when Ψ is calculated in a representation based on operators derived from angle coordinates and their conjugate momenta.

† Permanent address: H. H. Wills Physical Laboratory, Tyndall Avenue, Bristol BS8 1TL.

Vol. 287. A 1343. 30 [Published 4 October 1977

When the classical motion is *non-integrable*, the phase space is not filled with tori and existing semi-classical methods fail. We conjecture that

(*a*) For a given value of non-integrability parameter ϵ, the system passes through three semi-classical régimes as $\hbar$ diminishes.

(*b*) For states $|\psi\rangle$ associated with regions in phase space filled with irregular trajectories, Ψ will be a random function confined near that region of the 'energy shell' explored by these trajectories (this region has more than f dimensions).

(*c*) For $\epsilon \neq 0$, $\hbar$ blurs the infinitely fine classical path structure, in contrast to the integrable case $\epsilon = 0$, where $\hbar$ imposes oscillatory quantum detail on a smooth classical path structure.

1. Introduction

The celebrated function $\Psi(\boldsymbol{q}, \boldsymbol{p})$, introduced by Wigner (1932) to represent the quantum state $|\psi\rangle$, has stimulated a number of attempts (Groenewold 1946; Moyal 1949; Takabayasi 1954; Baker 1958) to formulate the mechanics of quantum systems in terms of functions of variables $\boldsymbol{q}, \boldsymbol{p}$ resembling the conjugate coordinates and momenta of the corresponding classical systems. In spite of this, studies of the semi-classical behaviour of $\Psi(\boldsymbol{q}, \boldsymbol{p})$ itself have been either inconclusive (Schipper 1969) or limited in scope (Balazs & Zipfel 1973; Siegel 1976; Voros 1976).

It is the purpose of the present work to investigate in some detail both the classical limit and the semi-classical limiting form (the 'semi-classical limit') of Wigner's function. (These two limits are different because $\Psi(\boldsymbol{q}, \boldsymbol{p})$, like the wavefunction $\langle \boldsymbol{q}|\psi\rangle$, is a highly non-analytic function of Planck's constant $\hbar$ at $\hbar = 0$.) What emerges from the analysis is a rich asymptotic structure to $\Psi(\boldsymbol{q}, \boldsymbol{p})$, in which strong but strangely distorted echoes of the more familiar semi-classical mechanics of wavefunctions can be discerned.

In order that the paper be not too long the following three restrictions are made: First, only those Wigner functions Ψ representing *pure states* $|\Psi\rangle$ are considered, because much of the interesting non-analyticity in $\hbar$ suffers 'thermal quenching' (Heller 1976) in the general case where Ψ represents a density matrix corresponding to a mixture of quantum states. Secondly, only those Ψ representing *energy eigenstates* are considered; extension to more general states $|\Psi\rangle$ should not present fundamental difficulties (a discussion of the important problems posed by the time evolution of Ψ is given by Heller 1976). And thirdly, only *bound systems* are considered; in fact the results apply almost unchanged to scattering systems.

The central results concern the form of $\Psi(\boldsymbol{q}, \boldsymbol{p})$ for bound quantum systems whose classical motion is *integrable*. If the system has f degrees of freedom, this means that f classical constants of the motion exist, so that the system inhabits f-dimensional regions of the $2f$-dimensional phase space, which can be shown (appendix 26 of Arnol'd & Avez 1968) to have the topology of tori. (In one dimension the single constant of motion is the energy, and the 'tori' are just the familiar closed curves in the phase plane q, p.) In the classical limit $\Psi(\boldsymbol{q}, \boldsymbol{p})$ collapses onto a delta function (§ 3) confined to the torus defined by the state $|\psi\rangle$ under consideration. In the semi-classical limit (§ 4) the delta function softens into a Wigner function large on the torus, oscillating on its 'concave' side and decaying on its 'convex' side. (In one-dimensional cases fringes were obtained by Balazs & Zipfel 1973 – but see § 4.) Quantization, that is the selection of particular tori to represent quantum states, arises (see § 6) from the requirement that $\Psi(\boldsymbol{q}, \boldsymbol{p})$ be single-valued. Classical integrability does not imply quantum separability (§ 8).

A peculiar feature of the Wigner function is the existence of 'catastrophes' far from the classical tori, on which $\Psi(\boldsymbol{q}, \boldsymbol{p})$ can become infinite as $\hbar \to 0$ (§ 5); however, these divergences do not have the character of delta functions.

The representation of states $|\psi\rangle$ given by the Wigner function depends crucially on the choice of operators $\hat{\boldsymbol{q}}$, $\hat{\boldsymbol{p}}$, to which the variables $\boldsymbol{q}$ and $\boldsymbol{p}$ correspond. The simplest situation, assumed throughout most of this paper, occurs when the eigenvalues of $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{p}}$ form continuous unbounded f-dimensional Euclidean spaces. When $\hat{\boldsymbol{q}}$ corresponds to an angle coordinate, however, notorious difficulties arise (Carruthers & Nieto 1968; Leaf 1968, 1969), whose semi-classical implications are discussed briefly in § 7.

When the classical motion is *non-integrable* (a situation that can only arise if $f \geqslant 2$), the nature of the stationary states and the distribution of quantized energy levels are unknown (Percival 1973). This is the most profound problem facing semi-classical mechanics at present. Here, in § 9, the case is made for the potential usefulness of the Wigner function in studying these classically non-integrable systems; this possibility was suggested by Dr P. Lloyd (private communication), and also Nordholm & Rice (1974).

The important physical conclusion emerges that Planck's constant plays a fundamentally different rôle in integrable and non-integrable systems. For integrable systems $\hbar$ imposes on oscillatory quantum fine structure onto a smooth classical background. For non-integrable systems $\hbar$ imposes a quantum smoothing onto a classical background with fine structure on arbitrarily small scales.

Superficially the 'phase space' methods employed in this paper resemble those of Maslov (1972) (see also Kravtsov 1968; Duistermaat 1974; Voros 1976). For three reasons this similarity is apparent rather than real: (i) Maslov approximates states $|\psi\rangle$, while we approximate Wigner functions which depend quadratically on $|\psi\rangle$ and generate probability densities, e.g. $|\langle \boldsymbol{q}|\,\psi\rangle|^2$, (ii) Maslov gives integral representations for $\langle \boldsymbol{q}|\psi\rangle$ while we give explicit functional forms for $\Psi(\boldsymbol{q}, \boldsymbol{p})$; (iii) Maslov's mathematics is rigorous, while ours is heuristic.

2. Weyl–Wigner formalism

Let $\hat{a}$ be an operator expressible in terms of the fundamental coordinate operators $\hat{\boldsymbol{q}} \equiv (\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_f)$ and momentum operators $\hat{\boldsymbol{p}} \equiv (\hat{p}_1, \hat{p}_2 \ldots \hat{p}_f)$ whose eigenvalues are continuous and unbounded and which obey the usual commutation rules. The Weyl correspondence associates with $\hat{a}$ a function $A(\boldsymbol{q}, \boldsymbol{p})$ as follows:

$$A(\boldsymbol{q}, \boldsymbol{p}) \equiv \frac{1}{\hbar^f} \mathrm{Tr}\left[\hat{a} \int \mathrm{d}\boldsymbol{Q} \int \mathrm{d}\boldsymbol{\Pi} \exp \frac{\mathrm{i}}{\hbar}\{(\hat{\boldsymbol{p}} - \boldsymbol{p}) \cdot \boldsymbol{Q} + (\hat{\boldsymbol{q}} - \boldsymbol{q}) \cdot \boldsymbol{\Pi}\}\right]. \tag{2.1}$$

The inverse association is

$$\hat{a} = \frac{1}{\hbar^{2f}} \int \mathrm{d}\boldsymbol{q} \int \mathrm{d}\boldsymbol{p} A(\boldsymbol{q}, \boldsymbol{p}) \int \mathrm{d}\boldsymbol{Q} \int \mathrm{d}\boldsymbol{\Pi} \exp \frac{\mathrm{i}}{\hbar}\{(\boldsymbol{p} - \hat{\boldsymbol{p}}) \cdot \boldsymbol{Q} + (\boldsymbol{q} - \hat{\boldsymbol{q}}) \cdot \boldsymbol{\Pi}\}. \tag{2.2}$$

(All integrations will run from $-\infty$ to $+\infty$ over all f components of each variable.)

Three things are important about the Weyl correspondence: first, $A(\boldsymbol{q}, \boldsymbol{p})$ is a *quantum function*; it depends on $\hbar$ and in general differs from the classical function corresponding to $\hat{a}$ except when $\hbar = 0$. Secondly, contrary to folklore, a unitary transformation to new fundamental operators $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{p}}$ is in general not equivalent to a canonical transformation in the 'phase space' $\boldsymbol{q}, \boldsymbol{p}$ (Leaf 1968) but leads to a completely new representation of $\hat{a}$. Thirdly, Weyl's rule (2.1) is not the only way to define a function of $\boldsymbol{q}$ and $\boldsymbol{p}$ that represents an operator $\hat{a}$, since several other rules can be devised (Mehta 1964), corresponding to different ordering conventions for the operators $\hat{\boldsymbol{q}}$, $\hat{\boldsymbol{p}}$; however, it is shown in appendix A that these alternative possibilities are all unsatisfactory for exploring the classical limit.

Wigner's function – the central object of study here – is the function $\Psi(\boldsymbol{q},\boldsymbol{p})$ corresponding to the operator

$$\hat{a} = \frac{|\psi\rangle\langle\psi|}{h^f}, \tag{2.3}$$

where $|\psi\rangle$ is the quantum state that Ψ represents. Thus

$$\Psi(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{h^{2f}}\int \mathrm{d}\boldsymbol{Q}\int \mathrm{d}\boldsymbol{\Pi}\exp\left\{-\frac{\mathrm{i}}{\hbar}(\boldsymbol{p}\cdot\boldsymbol{Q}+\boldsymbol{q}\cdot\boldsymbol{\Pi})\right\}\langle\psi|\exp\frac{\mathrm{i}}{\hbar}(\hat{\boldsymbol{q}}\cdot\boldsymbol{\Pi}+\hat{\boldsymbol{p}}\cdot\boldsymbol{Q})\,|\psi\rangle. \tag{2.4}$$

One of the integrations can be eliminated by using the fact that $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{p}}$ each commute with their commutator; this gives

$$\Psi(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{(\pi\hbar)^f}\int \mathrm{d}\boldsymbol{X}\exp\left(-\frac{2\mathrm{i}}{\hbar}\boldsymbol{p}.\boldsymbol{X}\right)\langle\boldsymbol{q}+\boldsymbol{X}|\psi\rangle\langle\psi|\boldsymbol{q}-\boldsymbol{X}\rangle, \tag{2.5}$$

from which it follows that Wigner's function is real.

It is obvious from (2.4) that in Wigner's function a complete formal symmetry exists between $\boldsymbol{q}$ and $\boldsymbol{p}$. This formal symmetry is not manifest in (2.5). However, the unsymmetrical form, involving the wavefunctions in the position representation, is more convenient for the purpose of calculating Ψ. The fact that all approximations thus obtained will exhibit the symmetry between $\boldsymbol{q}$ and $\boldsymbol{p}$ is a useful check on their correctness.

From (2.4) or (2.5) there follow

$$\left.\begin{aligned}\int \mathrm{d}\boldsymbol{q}\,\Psi(\boldsymbol{q},\boldsymbol{p}) &= |\langle\boldsymbol{p}|\psi\rangle|^2\\ \int \mathrm{d}\boldsymbol{p}\,\Psi(\boldsymbol{q},\boldsymbol{p}) &= |\langle\boldsymbol{q}|\psi\rangle|^2\end{aligned}\right\}. \tag{2.6}$$

This shows that the physically important 'projections' of Ψ are always positive, even though Ψ itself may be negative for some values of $\boldsymbol{q}$ and $\boldsymbol{p}$. If $|\psi\rangle$ is normalized to unity, it also follows that

$$\int \mathrm{d}\boldsymbol{q}\int \mathrm{d}\boldsymbol{p}\,\Psi(\boldsymbol{q},\boldsymbol{p}) = \langle\psi|\psi\rangle = 1. \tag{2.7}$$

The expectation value in the state $|\psi\rangle$ of an observable whose operator is $\hat{a}$ is, from (2.2) and (2.4),

$$\langle\psi|\hat{a}|\psi\rangle = \int \mathrm{d}\boldsymbol{q}\int \mathrm{d}\boldsymbol{p}\,A(\boldsymbol{q},\boldsymbol{p})\,\Psi(\boldsymbol{q},\boldsymbol{p}). \tag{2.8}$$

Some preliminary insight into the semi-classical behaviour of Ψ can be obtained from the operator identity

$$(|\psi\rangle\langle\psi|)^2 = |\psi\rangle\langle\psi| \tag{2.9}$$

whose Weyl correspondence is

$$\begin{aligned}\Psi(\boldsymbol{q},\boldsymbol{p}) = \frac{2^{2f}}{h^f}\int \mathrm{d}\boldsymbol{q}_1\int \mathrm{d}\boldsymbol{q}_2\int \mathrm{d}\boldsymbol{p}_1\int \mathrm{d}\boldsymbol{p}_2\,\Psi(\boldsymbol{q}_1,\boldsymbol{p}_1)\,\Psi(\boldsymbol{q}_2,\boldsymbol{p}_2)\\ \times\cos\left[\frac{2}{\hbar}\{\boldsymbol{p}_1\cdot(\boldsymbol{q}-\boldsymbol{q}_2)+\boldsymbol{p}_2\cdot(\boldsymbol{q}_1-\boldsymbol{q})+\boldsymbol{p}\cdot(\boldsymbol{q}_2-\boldsymbol{q}_1)\}\right].\end{aligned} \tag{2.10}$$

Takabayasi (1954) gives a detailed discussion of this relation, which is the condition Ψ must satisfy in order to represent a quantum pure state rather than a mixed state. Baker (1958) shows that it implies the inequality

$$|\Psi(\boldsymbol{q},\boldsymbol{p})| \leqslant (2/h)^f; \tag{2.11}$$

therefore Ψ must always be bounded, except possibly in the classical limit itself. Integration of (2.10) over $\boldsymbol{q}$ and $\boldsymbol{p}$ gives

$$\int \mathrm{d}\boldsymbol{q} \int \mathrm{d}\boldsymbol{p}\, \Psi^2(\boldsymbol{q}, \boldsymbol{p}) = 1/h^f. \tag{2.12}$$

This diverges in the classical limit, in contrast to the integral of Ψ itself (2.7) which remains bounded. The divergence could come either from a long-range 'tail' of Ψ^2 for large $|\boldsymbol{q}|$ and $|\boldsymbol{p}|$ or from singularities of Ψ^2. In fact the latter possibility is the one that actually occurs; this follows from the classical limit of (2.10), also derived by Baker (1958):

$$\Psi(\boldsymbol{q}, \boldsymbol{p}) \to h^f \Psi^2(\boldsymbol{q}, \boldsymbol{p}), \quad \text{as} \quad h \to 0. \tag{2.13}$$

Therefore when $h = 0$, Ψ is either zero or positive infinite. The divergence must be weak enough for (2.7) to hold and strong enough for (2.12) to hold.

These results whilst giving a useful indication of the semi-classical nature of Ψ are inadequate as a description of it. Even the dimensionality of the singular distribution onto which Ψ condenses is not specified (obviously it is less than $2f$); it could take any value from zero (isolated point in phase space, corresponding to a single classical orbit at one instant) to $2f-1$ (the whole 'energy shell' in phase space). Moreover, no indication is given of the manner in which Ψ attains the singular distribution. These are the subjects with which this paper deals.

3. Integrable systems: the classical limit

An integrable system has f independent constants of motion; these are functions of the classical phase space variables $\boldsymbol{q}$ and $\boldsymbol{p}$ and all their mutual Poisson brackets must vanish. If the system is bound, integrability implies that trajectories are forever confined to f-dimensional tori in phase space (appendix 26 of Arnol'd & Avez 1968). For a given torus the momenta $\boldsymbol{p}$ are multivalued functions of the coordinates $\boldsymbol{q}$. In the simplest one-dimensional case where a particle of mass m moves non-relativistically in a potential $V(q)$, the constant of motion is the energy E and the two-valued function $p(q, E)$ is

$$p(q, E) = \pm\sqrt{\{2m(E - V(q))\}}. \tag{3.1}$$

It is convenient to label tori by their *f action variables* $\boldsymbol{I}(\boldsymbol{q}, \boldsymbol{p})$ defined by

$$\boldsymbol{I}_i(\boldsymbol{q}, \boldsymbol{p}) \equiv \frac{1}{2\pi} \oint_{\gamma_i} \boldsymbol{p}(\boldsymbol{q}) \cdot \mathrm{d}\boldsymbol{q}, \tag{3.2}$$

where γ_i is the ith irreducible circuit of the forus. The I's are combinations of the original constants of motion and are obviously constants themselves. Points *on* the torus $\boldsymbol{I}$ are labelled by f *angle variables* $\boldsymbol{\theta}(\boldsymbol{q}, \boldsymbol{p})$, defined by

$$\boldsymbol{\theta} = \nabla_{\boldsymbol{I}} S(\boldsymbol{q}, \boldsymbol{I}), \tag{3.3}$$

where S is the position dependent action, namely

$$S(\boldsymbol{q}, \boldsymbol{I}) = \int_{\boldsymbol{q}_0}^{\boldsymbol{q}} \boldsymbol{p}(\boldsymbol{q}', \boldsymbol{I}) \cdot \mathrm{d}\boldsymbol{q}', \tag{3.4}$$

$\boldsymbol{q}_0$ being an arbitrary origin in the classically accessible region. $\boldsymbol{\theta}$ and $\boldsymbol{I}$ are conjugate coordinates and momenta in phase space, related to $\boldsymbol{q}$ and $\boldsymbol{p}$ by a canonical transformation with generating function $S(\boldsymbol{q}, \boldsymbol{I})$.

Energy eigenstates $|\psi_m\rangle$ correspond to particular tori $\boldsymbol{I}_m$ selected by a quantum condition and labelled by f quantum numbers $\boldsymbol{m}$. Quantization will be discussed in detail in §§ 6 and 7. Here it will be assumed that a particular $\boldsymbol{I}_m$ has been selected, and the classical limit of the corresponding Wigner function $\Psi_m(\boldsymbol{q}, \boldsymbol{p})$ will be calculated.

The starting point is equation (2.5). The eigenfunctions $\langle \boldsymbol{q}|\psi_m\rangle$ are linear combinations of primitive W.K.B. functions

$$\langle \boldsymbol{q}|\psi_m\rangle_i = C\left|\det \frac{\partial^2 S_i(\boldsymbol{q}, \boldsymbol{I}_m)}{\partial q_j \partial I_k}\right|^{\frac{1}{2}} \exp \frac{\mathrm{i}}{\hbar} S_i(\boldsymbol{q}, \boldsymbol{I}_m), \tag{3.5}$$

which can be shown (Van Vleck 1928; Dirac 1947) to be the semi-classical solutions of the time-independent Schrodinger equation. C is a normalization constant and the subscript i labels the values of S corresponding to the different choices of the multivalued function $\boldsymbol{p}(\boldsymbol{q}, \boldsymbol{I})$ in (3.4). If $\Psi_m(\boldsymbol{q}, \boldsymbol{p})$ is being evaluated at a point $\boldsymbol{q}, \boldsymbol{p}$ near the torus $\boldsymbol{I}_m$, $\boldsymbol{p}$ is close to a branch of $\boldsymbol{p}(\boldsymbol{q})$ which depends on where $\boldsymbol{q}, \boldsymbol{p}$ is. In these circumstances only the function $\langle \boldsymbol{q}|\psi_m\rangle_i$ corresponding to this branch need be inserted into (2.5), since the other functions that go to make up $\langle \boldsymbol{q}|\psi_m\rangle$ give a purely oscillatory contribution to the integrand in (2.5) which is classically negligible (this will become clear in the next section). Therefore (2.5) can be written as

$$\begin{aligned}\Psi_m(\boldsymbol{q}, \boldsymbol{p}) = \frac{C^2}{(\pi\hbar)^f}\int \mathrm{d}\boldsymbol{X} \exp\left\{\frac{\mathrm{i}}{\hbar}[S(\boldsymbol{q}+\boldsymbol{X}, \boldsymbol{I}_m) - S(\boldsymbol{q}-\boldsymbol{X}, \boldsymbol{I}_m) - 2\boldsymbol{p}\cdot\boldsymbol{X}]\right\} \\ \times\left|\det \frac{\partial^2 S(\boldsymbol{q}+\boldsymbol{X}, \boldsymbol{I}_m)}{\partial q_j \partial I_k} \det \frac{\partial^2 S(\boldsymbol{q}-\boldsymbol{X}, \boldsymbol{I}_m)}{\partial q_i \partial I_k}\right|^{\frac{1}{2}}.\end{aligned} \tag{3.6}$$

From this equation will be derived all subsequent semi-classical formulae for $\Psi(\boldsymbol{q}, \boldsymbol{p})$. These will be obtained in a form displaying complete formal symmetry in $\boldsymbol{q}$ and $\boldsymbol{p}$ and so will remain valid even near caustics in $\boldsymbol{q}$ space, where (3.5) no longer holds.

The classical limit itself is obtained by setting $\boldsymbol{X} = 0$ in the determinants in (3.6) and expanding the exponent to first order in $\boldsymbol{X}$, by using the formula, which follows from (3.4)

$$\nabla_{\boldsymbol{q}} S(\boldsymbol{q}, \boldsymbol{I}) = \boldsymbol{p}(\boldsymbol{q}, \boldsymbol{I}). \tag{3.7}$$

Then (3.6) becomes

$$\begin{aligned}\Psi_m(\boldsymbol{q}, \boldsymbol{p}) &= \frac{C^2}{(\pi\hbar)^f}\left|\det \frac{\partial p_j}{\partial I_k}\right| \int \mathrm{d}\boldsymbol{X} \exp\left\{\frac{2\mathrm{i}}{\hbar}[\boldsymbol{p}(\boldsymbol{q}, \boldsymbol{I}_m) - \boldsymbol{p}]\cdot\boldsymbol{X}\right\} \\ &= C^2\left|\det \frac{\partial p_j}{\partial I_k}\right| \delta(\boldsymbol{p} - \boldsymbol{p}(\boldsymbol{q}, \boldsymbol{I}_m)).\end{aligned} \tag{3.8}$$

This is an unsymmetrical expression, reflecting the manner of its derivation. However, the determinantal factor suggests changing variables in the delta function from $\boldsymbol{p}$ to $\boldsymbol{I}(\boldsymbol{q}, \boldsymbol{p})$. After setting C^2 equal to $(2\pi)^{-f}$, this gives the formally symmetric result

$$\Psi_m(\boldsymbol{q}, \boldsymbol{p}) = \frac{\delta(\boldsymbol{I}(\boldsymbol{q}, \boldsymbol{p}) - \boldsymbol{I}_m)}{(2\pi)^f}. \tag{3.9}$$

First observe that this expression is correctly normalized; this follows on integrating over $\boldsymbol{q}$ and $\boldsymbol{p}$ by changing variables to $\boldsymbol{\theta}$ and $\boldsymbol{I}$, by using the fact that the Jacobian of a canonical transformation is unity. Next, notice that this limiting form of Wigner's function is indeed a delta function on the torus $\boldsymbol{I}_m$ corresponding to the quantum state $|\psi_m\rangle$; moreover, Ψ_m is *uniform* on the torus, since the angle variables $\boldsymbol{\theta}$, which distinguish points on the torus, do not appear in (3.9). Next, recall

that no justification was given for the expansion in X that led to this result; the reason is that the correct semi-classical evaluation of (3.6), to be carried out in §§ 4 and 8, which will include (3.9) as its limit when $\hbar = 0$, involves more careful treatment of the rapid oscillations in the integrand. Finally, appreciate that (3.9) and the more accurate expressions to be derived in §§ 4 and 8 are *in no sense* Wigner functions in an 'action' representation, but rather Wigner functions in the Cartesian $\boldsymbol{q}, \boldsymbol{p}$ representation, expressed in action variables for clarity of presentation; the question of an 'action' representation of Ψ will be discussed in §7.

4. One-dimensional systems: the semi-classical limiting form

In order to see as clearly as possible how (3.6) can be evaluated to give the correct semi-classical limiting form of Ψ, attention will first be concentrated on the one-dimensional case. Let the state $|\psi_m\rangle$ have energy E and let the corresponding 'torus' in the phase space q, p, i.e. the energy shell $H(q, p) = E$, be a smooth convex curve denoted by $\mathscr{E}$ (figure 1) (what happens when $\mathscr{E}$ has inflexions will be discussed in §§ 5 and 7). For this quantum state (3.6) is

$$\Psi(q,p) = \frac{1}{\pi\hbar}\int_{-\infty}^{\infty} \frac{\mathrm{d}X \exp\left\{\frac{\mathrm{i}}{\hbar}\left[\int_{q-X}^{q+X}\mathrm{d}\xi\, p(\xi) - 2pX\right]\right\}}{\left|\frac{\partial I}{\partial p}(q+X, p(q+X))\frac{\partial I}{\partial p}(q-X, p(q-X))\right|^{\frac{1}{2}}}, \tag{4.1}$$

where the dependence of the momentum $p(q)$ on E (or, what amounts to the same thing, on the action I) is not written explicitly.

In the semi-classical limit the integrand oscillates rapidly, and the dominant contributions to Ψ come from those values of X for which the exponent is stationary, i.e. where

$$\tfrac{1}{2}[p(q+X) + p(q-X)] = p. \tag{4.2}$$

Let a positive solution of this equation be $X_0(q,p)$; then there will also be a negative solution $-X_0(q,p)$. The pair of points $q \pm X_0, p(q \pm X_0)$ lies on $\mathscr{E}$; let the points be labelled 1, 2 in the order they would be encountered by the classical system moving around E (figure 1). Then a symmetric way of expressing (4.2) is as follows: the points 1 and 2 that contribute to $\Psi(q, p)$ are the ends of that chord of E whose midpoint is q, p (figure 1). For points within and not too far from $\mathscr{E}$ there will be just one such chord; more complicated possibilities will be discussed in the next two sections.

The phase of the integrand in (4.1) at the stationary points 1 and 2 is simply

$$\left[\int_{q\mp X_0}^{q\pm X_0}\mathrm{d}\xi\, p(\xi) \mp 2pX_0\right] = \left.\begin{array}{l} -A(q,p) \text{ at point 1} \\ +A(q,p) \text{ at point 2}\end{array}\right\}, \tag{4.3}$$

where $A(q, p)$ is the (positive) area (shaded in figure 1) between $\mathscr{E}$ and the chord 12. This result holds whether Ψ is being evaluated in the 'upper' part of $\mathscr{E}$ (q, p in figure 1) or the 'lower' part of $\mathscr{E}$ ($q'p'$ in figure 1), because of the way the points 1 and 2 have been defined.

The second derivatives of the phase at the stationary points are

$$\frac{\partial p}{\partial q}(q \pm X_0) - \frac{\partial p}{\partial q}(q \mp X_0) = \left.\begin{array}{l}\text{positive at point 1} \\ \text{negative at point 2}\end{array}\right\}. \tag{4.4}$$

Again this result holds whether q, p is in the 'upper' or 'lower' parts of $\mathscr{E}$.

These are the ingredients necessary for the evaluation of (4.1) by the method of stationary phase, which will be valid when points 1 and 2 are not too close, i.e. when qp is not too near $\mathscr{E}$. The denominator can be put into a symmetric form by making use of the fact that on $\mathscr{E}$

$$\frac{\mathrm{d}I}{\mathrm{d}q}(q,p(q)) = \frac{\partial I}{\partial q}+\frac{\partial I}{\partial p}\frac{\partial p}{\partial q} = 0. \tag{4.5}$$

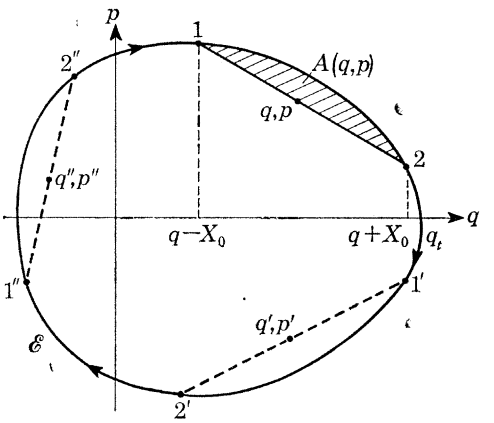

FIGURE 1. Points 1 and 2, on the 'torus' $\mathscr{E}$, contributing to the Wigner function Ψ at q, p.

Then, taking careful account of phases, and denoting derivatives of $I(q, p)$ by subscripts, the following expression is obtained for Wigner's function:

$$\Psi(q,p) = \frac{2\cos[A(q,p)/\hbar-\frac{1}{4}\pi]}{\pi\sqrt{h}[I_q(2)\,I_p(1)-I_p(2)\,I_q(1)]^{\frac{1}{2}}}, \tag{4.6}$$

where the arguments 1 and 2 indicate the points on $\mathscr{E}$ where the derivatives are to be evaluated. Being formally symmetric in q and p this result will also hold for points such as $q''p''$ on figure 1 where the derivation based on (3.5) would fail owing to the presence of a turning point of q between the chord ends 1 and 2.

What (4.6) shows is the presence of *fringes* (oscillations of Ψ) within the torus $\mathscr{E}$. Moreover, the expression is square integrable and satisfies the pure state condition (2.12); this is shown in appendix B. Nevertheless, it is unsatisfactory for the following reasons: first, it diverges on $\mathscr{E}$ itself, thus violating the inequality (2.11); this happens because points 1 and 2 coalesce as q, p moves onto $\mathscr{E}$, so that the simple method of stationary phase becomes inapplicable. Secondly, it does not satisfy the normalization condition (2.7); this is shown in appendix C. And thirdly, it predicts that Ψ is zero for points q, p outside $\mathscr{E}$ where (4.1) has no stationary points.

All these deficiencies of (4.6) can be remedied quite easily by employing the method of *uniform approximation* (Chester, Friedman & Ursell 1957) to evaluate (4.1) in a manner that remains valid as q, p moves onto $\mathscr{E}$. This gives the following expression, which is the central result of this section:

$$\Psi(q,p) = \frac{\sqrt{2}\,[\frac{3}{2}A(q,p)]^{\frac{1}{6}}\,\mathrm{Ai}\left(-\left[\frac{3A(q,p)}{2\hbar}\right]^{\frac{2}{3}}\right)}{\pi\hbar^{\frac{2}{3}}[I_q(2)\,I_p(1)-I_p(2)\,I_q(1)]^{\frac{1}{2}}}. \tag{4.7}$$

Ai denotes the Airy function (Abramowitz & Stegun 1964). This semi-classical limiting form of Wigner's function is rich with interesting properties, which will now be enumerated.

(i) There is manifest formal symmetry between q and p. This property is not possessed by the result obtained by Balazs & Zipfel (1973). Their formula (7) (which is actually for the integral of Ψ over energy rather than Ψ itself) was derived from the uniform approximation for $\langle q|\psi\rangle$ (cf. (viii) below) rather than the primitive W.K.B. approximation (3.5) used here, but this apparently greater accuracy is thrown away by illegitimately expanding the expression in the integrand of (2.5) to only first order in X rather than employing the correct stationary phase method to evaluate the integral.

(ii) When q, p is inside $\mathscr{E}$ and not close to $\mathscr{E}$, $A(q,p)$ is large in comparison with $\hbar$. Therefore the Airy function can be replaced by its asymptotic form for large negative argument, and (4.6) is recovered.

(iii) When q, p is outside $\mathscr{E}$ and not close to $\mathscr{E}$ there are no real stationary phase points, but (4.2) has complex solutions, so that q, p is the midpoint of a 'complex chord' joining two complex points 1 and 2 on the analytic continuation of $\mathscr{E}$. Then $A(q,p)$ is imaginary with phase $\frac{3}{2}\pi$ and the argument of the Airy function is large and positive. The appropriate asymptotic form shows that $\Psi(q,p)$ decays exponentially on the convex side of $\mathscr{E}$, while remaining real.

(iv) When q, p is very close to $\mathscr{E}$ the functions in (4.7) can be expanded to lowest order in the action difference $I(q,p)-I(\mathscr{E})$. This is tedious but straightforward, and leads to the following 'transitional approximation' for Ψ (which can also be derived by expanding the phase in (4.1) to third order in X – cf. appendix H):

$$\Psi(q,p) = \frac{1}{\pi}\left(\frac{1}{\hbar^2 B(q,p)}\right)^{\frac{1}{3}} \mathrm{Ai}\left[2\{I(q,p)-I(\mathscr{E})\}\left(\frac{1}{\hbar^2 B(q,p)}\right)^{\frac{1}{3}}\right], \tag{4.8}$$

where

$$B(q,p) \equiv I_q^2 I_{pp} + I_p^2 I_{qq} - 2 I_{pq} I_p I_q, \tag{4.9}$$

all derivatives being evaluated at q, p. B remains finite as q, p moves onto $\mathscr{E}$.

(v) On $\mathscr{E}$, Ψ as given by (4.7) rises to a value of order $\hbar^{-\frac{2}{3}}$ and so does not violate (2.11). In fact (4.7) shows precisely how the classical limit of a delta function on the torus $\mathscr{E}$ (equation 3.9) gets softened in the semi-classical limit when $\hbar$ is small but non-zero. In strict mathematical terms $\mathscr{E}$ is a *fold catastrophe* of the Wigner function (Thom 1975; Duistermaat 1974; Berry 1976), because it is on $\mathscr{E}$ that the mapping (4.2) induced by the gradient of the phase in (4.1) is singular. These catastrophes resemble the caustics (Berry 1976) of families of trajectories in coordinate space or momentum space; however, that term would be very confusing if used in the present context, because owing to Liouville's theorem there are no caustics in phase space. It is possible for Ψ to have higher catastrophes, and catastrophes not on $\mathscr{E}$; these will be examined in the next two sections.

(vi) The semi-classical limiting form (4.7) shows that Ψ at any point q, p depends only on the properties of the torus $\mathscr{E}$, and not on the properties of the classical trajectory that passes through q, p. This implies that to the present order of approximation (which is extremely accurate – cf. (viii) below) *arbitrary alterations* may be made to the Hamiltonian at points not on $\mathscr{E}$ without in any way affecting the Wigner function representing the eigenstate whose energy is E!.

(vii) The uniform approximation shares the property of the cruder expression (4.6) of being square-integrable to h^{-1} (appendix B). In addition, however, (4.7) is correctly normalized and satisfies the condition (2.7); this is shown in appendix C.

(viii) In view of the manner in which (4.7) was derived, it is a remarkable fact that when integrated over p it gives the correct semi-classical position probability density $|\langle q|\psi\rangle|^2$ (equation

(2.6)) *even near turning points*. This will now be shown. Let the Hamiltonian be even in p and let the turning point lie at $q = q_t$ (figure 1). Then

$$|\langle q|\psi\rangle|^2 = \frac{2\sqrt{2}}{\pi\hbar^{\frac{2}{3}}}\int_0^\infty \mathrm{d}p\, \frac{[\frac{3}{2}A(q,p)]^{\frac{1}{6}}\,\mathrm{Ai}\left(-\left[\frac{3A(q,p)}{2\hbar}\right]^{\frac{2}{3}}\right)}{[I_q(2)\,I_p(1) - I_p(2)\,I_q(1)]^{\frac{1}{2}}}. \tag{4.10}$$

To evaluate the integral the variable is changed from p to V, where

$$V(p) \equiv -[\tfrac{3}{2}A(q,p)]^{\frac{2}{3}}. \tag{4.11}$$

As p varies from 0 to ∞, V varies from

$$V(0) = -[\tfrac{3}{2}A(q,0)]^{\frac{2}{3}} = -\left[3\int_q^{q_t} p(q')\,\mathrm{d}q'\right]^{\frac{2}{3}} \tag{4.12}$$

to ∞. The most important region of the integrand is the neighbourhood of the lower limit $V(0)$. This follows from two facts: the first (appendix D) is that near $p = 0$

$$\mathrm{d}p = \tfrac{1}{2}[\tfrac{3}{2}A(q,0)]^{\frac{1}{6}}\left|\frac{\partial p}{\partial q}\right|^{\frac{1}{2}}\frac{\mathrm{d}V}{\sqrt{[V + (\frac{3}{2}A(q,0))^{\frac{2}{3}}]}}, \tag{4.13}$$

which diverges at $V(0)$. The second fact is that the denominator in (4.10) remains bounded as $V \to V(0)$ – in fact it takes a simple form, obtained as follows: first note that

$$I_q(2)\,I_p(1) - I_p(2)\,I_q(1) \to 2I_q(1)\,I_p(1), \quad \text{as} \quad p \to 0. \tag{4.14}$$

Next note that because, the Hamiltonian can be written either in the original variables q, p or in terms of the action I, the action derivatives can be expressed in the form

$$\left.\begin{aligned} \omega I_p(1) &= \frac{\partial H}{\partial p} = \dot{q}(1) \\ \omega I_q(1) &= -\frac{\partial H}{\partial q} = -\dot{p}(1) = +\dot{q}(1)\left|\frac{\partial p}{\partial q}\right| \end{aligned}\right\}, \tag{4.15}$$

where ω is the frequency of classical motion round $\mathscr{E}$ and dots denote differentiation with respect to time. Therefore

$$I_q(2)\,I_p(1) - I_p(2)\,I_q(1) \to 2\left|\frac{\partial p}{\partial q}\right|\frac{\dot{q}^2}{\omega^2}, \quad \text{as} \quad p \to 0. \tag{4.16}$$

The probability density (4.10) now becomes, in an approximation accurate as $\hbar \to 0$

$$|\langle q|\psi\rangle|^2 = \frac{\omega(\frac{3}{2}A(q,0))^{\frac{1}{3}}}{\pi\hbar^{\frac{2}{3}}|\dot{q}(q)|}\int_{-[\frac{3}{2}A(q,0)]^{\frac{2}{3}}}^\infty \mathrm{d}V\, \frac{\mathrm{Ai}\,(V/\hbar^{\frac{2}{3}}),}{\sqrt{\{V + [\frac{3}{2}A(q,0)]^{\frac{2}{3}}\}}}. \tag{4.17}$$

The integral can be evaluated exactly because of the following surprising 'projection identity' (appendix E):

$$\int_{-y}^\infty \mathrm{d}x\, \frac{\mathrm{Ai}\,(x)}{\sqrt{(x+y)}} = 2^{\frac{2}{3}}\pi\,\mathrm{Ai}^2\left(\frac{y}{2^{\frac{2}{3}}}\right). \tag{4.18}$$

This gives, finally,

$$|\langle q|\psi\rangle|^2 = \frac{2\omega}{|\dot{q}(q)|}\left[\frac{3}{2\hbar}\int_q^{q_t}\mathrm{d}q'p(q')\right]^{\frac{1}{3}}\mathrm{Ai}^2\left(-\left[\frac{3}{2\hbar}\int_q^{q_t}\mathrm{d}q'p(q')\right]^{\frac{2}{3}}\right), \tag{4.19}$$

which is not only positive definite but indeed precisely the uniform semi-classical approximation for $|\langle q|\psi\rangle|^2$ (Langer 1937; see also Berry & Mount 1972), and moreover in a form that is correctly normalized.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

5. Catastrophes of Wigner functions in non-classical regions

Catastrophes of Ψ occur at points q, p where the simple stationary phase approximation (4.6) diverges. It has already been explained how this happens on $\mathscr{E}$, and how (4.7) provides a uniform approximation in this case. To examine more general possibilities it is convenient to transform the quantity whose square root appears in (4.6); denoting this by $D(q, p)$ and using the relations (4.15) we get

$$D(q,p) \equiv I_p(1)\, I_q(2) - I_q(1)\, I_p(2) = [\dot{p}(1)\, \dot{q}(2) - \dot{q}(1)\, \dot{p}(2)]/\omega^2. \tag{5.1}$$

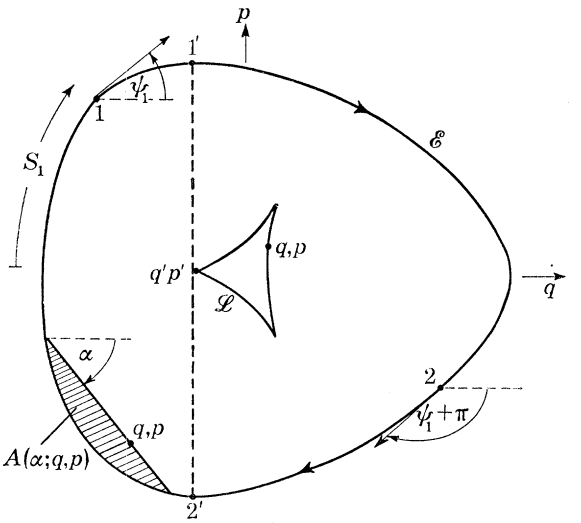

Figure 2. Catastrophe line $\mathscr{L}$ where Ψ is semi-classically large. Most of $\mathscr{L}$ is smooth (e.g. q, p), but there are at least three cusps (e.g. q', p') at the midpoints of diameters at whose ends the curvatures of $\mathscr{E}$ are equal. Also illustrated is the definition of $A(\alpha; q, p)$.

Now let s be arc length on $\mathscr{E}$ (measured from some arbitrary origin), and let ψ be the angle made by the 'forward' tangent at $\mathscr{E}$ with the q axis (figure 2). Then (5.1) becomes

$$D(q,p) = \frac{\dot{s}_1\, \dot{s}_2}{\omega^2} \sin(\psi_1 - \psi_2), \tag{5.2}$$

where $\dot{s}_1$ and $\dot{s}_2$ are the 'speeds' with which the classical system is describing $\mathscr{E}$ at 1 and 2. From this it is obvious that catastrophes occur at points q, p in phase space for which ψ_1 and ψ_2 differ by an integer multiple of π. Alternatively stated, catastrophes occur at the *midpoints of chords joining parallel parts of* $\mathscr{E}$. In order to understand Ψ it is necessary to understand these catastrophes in some detail.

In the special case already discussed (q, p on $\mathscr{E}$), the chord is of zero length, so that $\mathscr{E}$ is obviously parallel at its ends. In general (5.2) shows that catastrophes occur on lines $\mathscr{L}$ in the q, p plane defined as the *loci of midpoints of 'diameters' of* $\mathscr{E}$, where the term 'diameter' will be understood to mean a chord joining parallel parts of $\mathscr{E}$. The lines $\mathscr{L}$ will be smooth except at points q, p at the ends of whose diameters the *curvatures* of $\mathscr{E}$ are equal; at such a point $\mathscr{L}$ has a *cusp* (figure 2), in the generic case (this follows from the theory of Thom (1975)). It is shown in appendix F that for any closed convex curve $\mathscr{E}$ there must be at least three such cusps and in general an odd number; figure 2 shows an example of this.

At points of $\mathscr{L}$ that do not lie on $\mathscr{E}$ *two* coalescences of pairs of stationary points of the integral (4.1) occur, at $X = \pm X_1$, say. This contrasts with what happens on $\mathscr{E}$ where a single coalescence occurs at $X = 0$. To examine the global aspects of these 'coalescences off $\mathscr{E}$' for the case where $\mathscr{E}$ is a closed convex curve (general anharmonic oscillator) it is convenient to introduce an interesting geometrical representation that does not have the disadvantage of (4.1) of failing whenever a turning point of q lies between points 1 and 2 on $\mathscr{E}$. This is as follows: Let $A(\alpha; q, p)$ be the area cut off (figure 2) by the chord of $\mathscr{E}$ through q, p that makes an angle α with the q axis. Then elementary geometry shows that the condition

$$\partial A(\alpha; q, p)/\partial\alpha = 0, \tag{5.3}$$

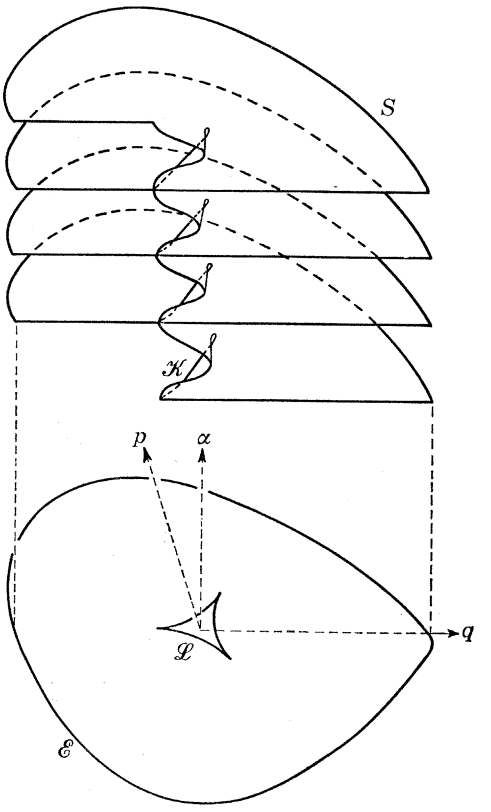

FIGURE 3. The 'rear' half ($p > 0$) of the stationary point surface S (equation 5.3) in the q, p, α space. The fold line $\mathscr{K}$ projects onto the catastrophe line $\mathscr{L}$.

which defines a surface S in the space q, p, α is precisely equivalent to the stationary point condition (4.2) in that it selects that chord of which q, p is the midpoint. In the language of catastrophe theory (Thom 1975) (5.3) is a 'gradient map' with 'state variable' α and 'control variables' q, p, and S is the 'critical manifold'. The line $\mathscr{L}$ is formed by the singularities of the projection of S onto the control plane q, p.

S has a fascinating structure. It has period π in the variable α (because $A(\alpha; q, p)$ is unchanged (figure 2) when α changes by π). S touches the 'cylinder' whose projection along α is $\mathscr{E}$ in the two right-handed helices (figure 3)

$$\alpha = \psi + 2n\pi \quad \text{and} \quad \alpha = \psi + (2n+1)\,\pi. \tag{5.4}$$

Away from this cylinder S consists of two helicoids each of pitch 2π that join smoothly on a 'fold' curve $\mathscr{H}$ which has the form of a left-handed helix whose projection is $\mathscr{L}$.

Now we turn to the case where $\mathscr{E}$ has *points of inflexion*. Then the diameters can intersect $\mathscr{E}$ and their midpoints, which form the catastrophe line $\mathscr{L}$, can lie close to $\mathscr{E}$. It is convenient to analyse the specific example where $\mathscr{E}$ (figure 4) has two points of inflexion and is given by the equation

$$p(q) = p_0 - \alpha q^2 + \beta q^4 \quad (\alpha > 0, \beta > 0). \tag{5.5}$$

The inflexions are at

$$\left.\begin{aligned} q &= \pm\sqrt{(\alpha/6\beta)} \equiv \pm q_i \\ p &= p_0 - 5\alpha^2/36\beta \equiv p_i. \end{aligned}\right\} \tag{5.6}$$

The stationary phase condition (4.2) can be solved exactly in this case, and shows that there may be up to four stationary points X of (4.1), given by

$$X^2 = \frac{(\alpha - 6\beta q^2) \pm \sqrt{[(\alpha - 6\beta q^2)^2 - 4\beta(p(q) - p)]}}{2\beta}. \tag{5.7}$$

As q, p moves onto $\mathscr{E}$, two of these X-values coalesce at $X = 0$. As q, p moves onto the line $\mathscr{L}$ (figure 4) whose equation is

$$p = p(q) - (\alpha - 6\beta q^2)^2/4\beta \tag{5.8}$$

the X-values coalesce in pairs at $X = \pm X_1$, where

$$X_1 = \sqrt{[(\alpha - 6\beta q^2)/2\beta]}. \tag{5.9}$$

At the inflexions themselves all four roots coalesce at $X = 0$.

This understanding of the catastrophe structure of the q, p plane enables the semi-classical behaviour of Ψ to be understood in some detail. On *smooth parts* of $\mathscr{L}$, where just two stationary points coalesce, it is obvious that (4.6) must be replaced by a uniform approximation involving Airy functions and resembling (4.7). But there is a crucial difference: the two areas, A_a and A_b, say, corresponding to the two stationary points X coalescing at X_1 on $\mathscr{L}$, are not equal and opposite as is the case (equation (4.3)) with coalescences on $\mathscr{E}$. This has the effect of introducing a factor

$$F(q, p) = \cos(A_a + A_b/2\hbar) \tag{5.10}$$

into Ψ, modulating the slow variations of the Airy function (whose argument involves $A_a - A_b$ – cf. Berry 1966) with rapid oscillations as the point q, p moves parallel or perpendicular to $\mathscr{L}$. In these rapid oscillations Ψ takes negative as well as positive values, and this prevents the occurrence of a delta function on $\mathscr{L}$ in the classical limit – it is only on $\mathscr{E}$, where the factor F is unity, that the Airy functions (4.7) or (4.8) grow to arbitrarily large values, always positive, as $\hbar \to 0$.

On *cusps* of $\mathscr{L}$ (figure 2) three stationary points coalesce, and the Airy approximation itself diverges. The appropriate uniform approximation in this case will involve the function of Pearcey (1946) (for an experimental illustration of this type of diffraction, see Berry 1975). At a cusp, Ψ rises to a value of order $\hbar^{-\frac{3}{4}}$ and so does not violate (2.11).

At *inflexions of* $\mathscr{E}$, where $\mathscr{L}$ meets $\mathscr{E}$ (figure 4) four stationary points coalesce and the Pearcey approximation diverges. The appropriate uniform approximation in this case will involve a particular section through the diffraction function corresponding to the 'swallow tail' catastrophe (the section $\Omega_2 = 0$ of Berry 1976), and Ψ would show an intricate fringe structure. At an inflexion of $\mathscr{E}$, Ψ rises to a value of order $\hbar^{-\frac{1}{5}}$ and still does not violate (2.11).

This exhausts the description of the semi-classical behaviour of Ψ in generic one-dimensional systems. However, there are two important non-generic (i.e., 'infinitely special') cases where Ψ has catastrophes of infinite order at isolated points q, p. The first case is where $\mathscr{E}$ is a closed convex curve with a *centre of symmetry*. Then the catastrophe $\mathscr{L}$ (figure 2) collapses to a point, and the

helix $\mathscr{K}$ (figure 3) degenerates to a line parallel to the α axis. This occurs whenever a particle moves in a potential $V(q)$ which is an even function of q. A special case is the *harmonic oscillator* (for which $\mathscr{E}$ is an ellipse). It was shown by Groenewold (1946) that if the oscillator has unit frequency and unit mass the Wigner function of the nth eigenstate ($E_n = (n+\frac{1}{2})\,\hbar$) is

$$\Psi_n(q,p) = \frac{(-1)^n}{\pi\hbar}\exp[-(p^2+q^2)/\hbar]\,L_n\left(\frac{2(q^2+p^2)}{\hbar}\right), \tag{5.11}$$

where L_n is the nth Laguerre polynomial normalized to unity at $x = 0$. As well as the 'Airy' peak on the torus $\mathscr{E}$ where the argument and order of the polynomial are equal, $|\Psi|$ also has a maximum at the degenerate 'catastrophe' where $q = p = 0$. This maximum has the value

$$\Psi_n(0,0) = 2(-1)^n/h, \tag{5.12}$$

which exceeds the values found for the generic fold, cusp and swallowtail catastrophes, but still (just!) satisfies (2.11).

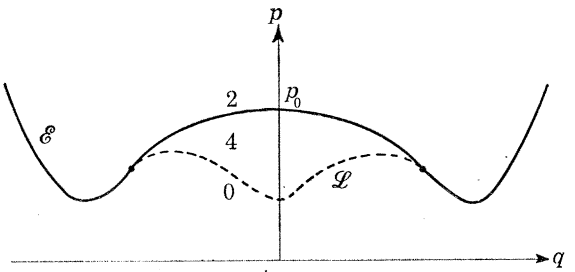

FIGURE 4. Catastrophe line $\mathscr{L}$ for a case (equation 5.5) where $\mathscr{E}$ has two inflexions. The number of real stationary points of (4.1) in each region is as indicated.

The second non-generic catastrophe, whose relevance will appear in § 7 when angle coordinates are studied, occurs when $\mathscr{E}$ is *antisymmetric about its inflexions*. Such a case is

$$p(q) = p_0 + A\cos q, \tag{5.13}$$

which could describe, for example, the 'hindered rotator' with Hamiltonian

$$H(q,p) = \omega(|p| - A\cos q) \equiv \omega p_0 \tag{5.14}$$

There are inflexions where $q = (n+\frac{1}{2})\,\pi$. In this case the midpoints of all diameters whose length $2X$ is less than 2π lie on these inflexions, so that the catastrophe lines $\mathscr{L}$ (figure 4) condense onto isolated 'catastrophe points of infinite order'.

6. Quantization

It is obvious (e.g. from the explicit expression 2.4) that $\Psi(\boldsymbol{q},\boldsymbol{p})$ must be a single-valued function of position $\boldsymbol{q}$, $\boldsymbol{p}$ in phase space. Now we show how this requirement, together with the simple stationary phase expression (4.6), leads to the familiar semi-classical quantization rule for one-dimensional anharmonic oscillators. The derivation is based on the fact, which we have purposely not mentioned before, that each chord of $\mathscr{E}$ defines *two* areas. One is the area $A(q,p)$ defined on figure 1. The other is the 'complementary' area $A'(q,p)$, given by

$$A' = \oint p\,\mathrm{d}q - A, \tag{6.1}$$

where $\oint p\,\mathrm{d}q$ is the total area within $\mathscr{E}$. The area A' is defined in exactly the same way as A, but *with the points* 1 *and* 2 *interchanged* (since 1 always precedes 2 by definition, A' the area on figure 1 between the chord and that part of $\mathscr{E}$ transversed while going 'the long way round' between points 1 and 2).

Now, any point q, p within $\mathscr{E}$ can be taken round a closed path in two fundamentally different ways. In the first, the path avoids the catastrophe curve (figure 5*a*) and points 1 and 2 end up in their original positions. Referring to figure 3, this corresponds to staying on a single helicoid so that α changes by 2π. In the second path – the one of principal interest – the path *touches* $\mathscr{L}$ (figure 5*b*) and points 1 and 2 end up reversed, with $A'(q, p)$ replacing $A(q, p)$. On figure 3 this corresponds to changing from one helicoid to the other by smoothly crossing the fold $\mathscr{K}$, so that α changes by π.

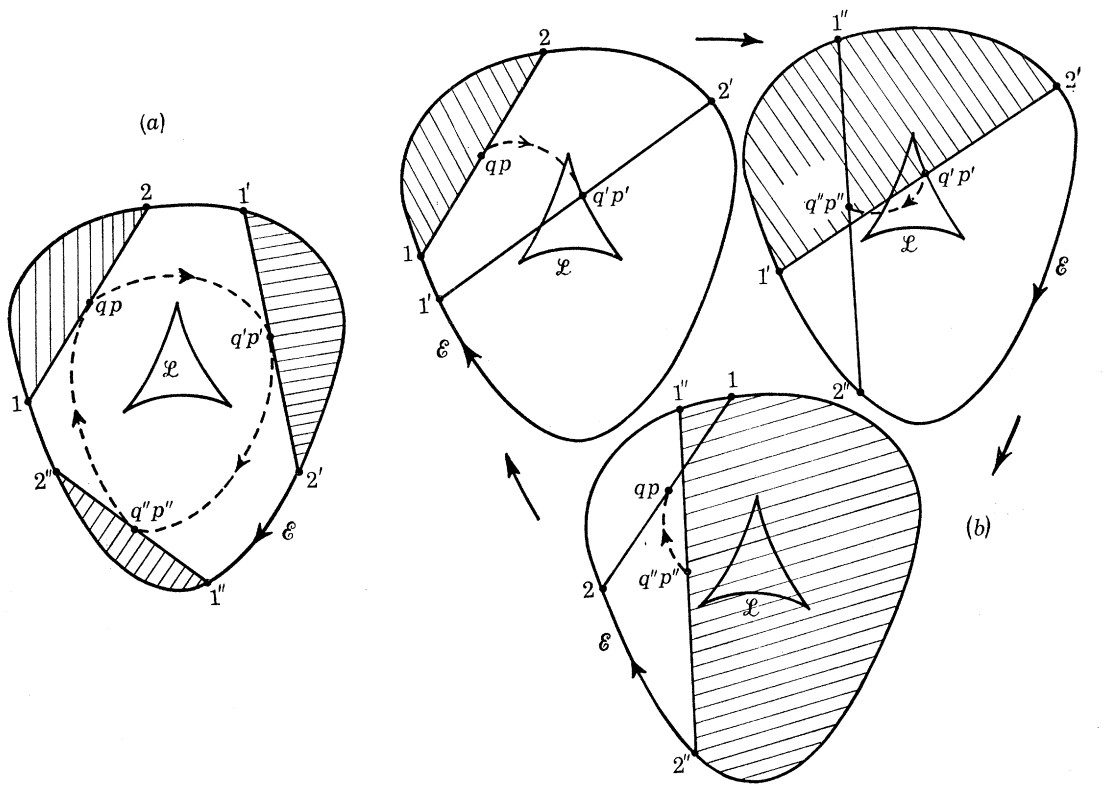

FIGURE 5. Closed paths ---▶--- of q, p. The areas A are shaded. (*a*) path avoids catastrophe line $\mathscr{L}$, (*b*) path touches $\mathscr{L}$, so that the area A changes to its complement A' and points 1 and 2 are interchanged.

Wigner's function (4.6) must remain unchanged after any such circuit of q, p. To understand the implications of this, (4.6) must be written in a slightly different form, employing (5.1) and (5.2) and making explicit the fundamental fact that Ψ is real:

$$\Psi(q,p) = \frac{2\omega}{\pi\sqrt{(h\dot{s}_1\dot{s}_2)}}\,\mathrm{Re}\,\frac{\exp \mathrm{i}[A(q,p)/\hbar - \frac{1}{4}\pi]}{[\sin(\psi_1-\psi_2)]^{\frac{1}{2}}}. \tag{6.2}$$

During a circuit in which 1 and 2 are interchanged and A is replaced by A', it is obvious that $\sin(\psi_1-\psi_2)$ must change sign. This sign change occurs when the point q, p touches $\mathscr{L}$, and is associated with a phase change of $+\pi$. The fact that this is $+\pi$ rather than $-\pi$ follows from the requirement that Ψ behaves as an Airy function near a smooth part of $\mathscr{L}$ and the consequent

restrictions on the phases of the contributions to Ψ on the two helicoids near $\mathscr{K}$ on figure 3. This can also be deduced from the fact that the phase jump being considered here is mathematically identical to the well-known phase advance of $\frac{1}{2}\pi$ as a ray touches a caustic (see Berry & Mount 1972 and references therein).

After the circuit, then, $\Psi(q, p)$ becomes $\Psi'(q, p)$ where

$$\Psi'(q,p) = \frac{2\omega}{\pi\sqrt{(h\dot{s}_1\dot{s}_2)}}\,\mathrm{Re}\,\frac{\exp \mathrm{i}[A'(q,p)/\hbar - \frac{1}{4}\pi - \frac{1}{2}\pi]}{[\sin(\psi_1-\psi_2)]^{\frac{1}{2}}}$$
$$= \frac{2\omega}{\pi\sqrt{(h\dot{s}_1\dot{s}_2)}}\,\frac{\cos[A(q,p)/\hbar - \frac{1}{4}\pi - \oint p\,\mathrm{d}q/\hbar + \pi]}{[\sin(\psi_1-\psi_2)]^{\frac{1}{2}}}, \tag{6.3}$$

where (6.1) has been used. But this must, equal $\Psi(q, p)$, the value of Wigner's function before the circuit, and comparison with (4.6) shows that this is possible only for tori $\mathscr{E}$ satisfying

$$\oint p\,\mathrm{d}q = (n+\tfrac{1}{2})\,h, \tag{6.4}$$

where n is an integer. Of course this is the correct semi-classical quantization rule that we set out to obtain.

7. Systems with angle coordinates

If one or more components $\hat{q}$ of the fundamental coordinate operators $\hat{\boldsymbol{q}}$ has *bounded* eigenvalues q, then the eigenvalues p of the conjugate momentum operator $\hat{p}$ are quantized. Henceforth, and without loss of generality, the domain of $\hat{q}$ will be taken to be

$$-\pi < q \leqslant +\pi \tag{7.1}$$

and q will be referred to as an 'angle' coordinate. q may be a rotation angle in real space, or one of the 'angle' variables $\boldsymbol{\theta}$ of classical mechanics (equation 3.3). When (7.1) holds, the discrete eigenvalues p_n of $\hat{p}$ will have spacing $\hbar$.

Suppose that in these circumstances it is desired to represent operators and quantum states in the phase space $\boldsymbol{q}$, $\boldsymbol{p}$, where the fundamental operators $\hat{\boldsymbol{q}}$ correspond to angles. Then the Weyl correspondence and the Wigner function as defined in § 2 must be modified in the following simple way: all functions in phase space must be evaluated only in the strip (7.1) for $\boldsymbol{q}$ and at the discrete eigenvalues $\boldsymbol{p}_n$. Moreover, all integrations over $\boldsymbol{q}$ must be confined to the range (7.1) and integrations over all quantized components of $\boldsymbol{p}$ replaced by summations as follows

$$\int \mathrm{d}p \to \hbar \sum_{p_n}. \tag{7.2}$$

Finally, the integration in (2.5) is over the range $-\frac{1}{2}\pi < X_i < \frac{1}{2}\pi$ for each component X_i of $\boldsymbol{X}$. These modifications preserve the consistency of all equations in § 2.

This procedure presupposes the existence of angle operators $\hat{\boldsymbol{q}}$. If $\hat{\boldsymbol{q}}$ corresponds to a *rotation angle in real space*, so that the eigenvalues of $\hat{\boldsymbol{p}}$ are

$$p_n = n\hbar \quad (-\infty < n < +\infty), \tag{7.3}$$

then no inconsistency will result from the use of the operator $\hat{\boldsymbol{q}}$ provided its spectrum is restricted by (7.1) (see Carruthers & Nieto 1968). Wigner's function for one degree of freedom now becomes

$$\Psi(q,p_n) = \frac{1}{\pi\hbar}\int_{-\frac{1}{2}\pi}^{+\frac{1}{2}\pi} \mathrm{d}X\,\mathrm{e}^{-2\mathrm{i}nX}\langle q+X|\psi\rangle\langle\psi|q-X\rangle, \tag{7.4}$$

where the arguments $q \pm X$ lie in the range (7.1) modulo 2π. Equation (7.4) has been independently derived on a rigorous group-theoretical basis by Dr N. Mukunda (private communication).

However, if q is a classical angle variable (3.3) corresponding to a *libration in real space* (oscillator system), then the conjugate momenta, which are the actions I, are also quantized, but the quantization restricts I to *positive values only*, namely

$$I_n = (n+\tfrac{1}{2})\,\hbar \quad (0 \leqslant n < \infty). \tag{7.5}$$

It has been shown by Leaf (1969) that this 'onesidedness' of I means that no angle variable $\hat{q}$ exists for such systems (a pleasant physical argument to this effect is given by Fanelli & Struzynski 1969). Nevertheless, as a *semi-classical approximation*, 'action-angle operators' can be very useful, as is shown by the interesting application and references in Marcus (1971).

The approximate Wigner function in action-angle phase space will be discussed later. Now, however, some of the interesting properties of the precisely defined 'rotational' Wigner function (7.4) will be considered. Let Ψ_m represent the mth energy eigenstate $|\psi_m\rangle$ when the Hamiltonian is smoothly periodic in q (hindered rotator). The W.K.B. eigenfunction is simply

$$\langle q|\psi_m\rangle = \sqrt{\left[\frac{\omega(E_m)}{2\pi\dot{q}(q,E_m)}\right]}\exp\left\{\frac{\mathrm{i}}{\hbar}\int_0^q p(q',E_m)\,\mathrm{d}q'\right\}, \tag{7.6}$$

where ω and $\dot{q}$ are the classical frequency and angular velocity of the rotation with energy E_m. The angular momentum p can differ only by a constant from a periodic function of q, so that the 'torus' $\mathscr{E}$ takes the form

$$p(q,E) = p_0(E) + p_1(q,E), \tag{7.7}$$

where $p_1(q)$ is periodic with period 2π. Single-valuedness of (7.6) then gives the quantization condition

$$p_0(E_m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} p(q',E_m)\,\mathrm{d}q' = m\hbar. \tag{7.8}$$

Wigner's function (7.4) now becomes

$$\Psi_m(q,p_n) = \frac{2\omega}{(2\pi)^2\hbar}\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi}\mathrm{d}X\,\frac{\exp\left[\mathrm{i}(m-n)\,X\right]\exp\left\{\frac{\mathrm{i}}{\hbar}\int_{q-X}^{q+X} p_1(q')\,\mathrm{d}q'\right\}}{\sqrt{[\dot{q}(q+X)\,\dot{q}(q-X)]}}, \tag{7.9}$$

where the functional dependences on E_m are omitted.

This is the analogue of the expression (4.1) for oscillators. The method of stationary phase can be applied as in § 4 and the same results obtained with the following modifications: (i) the p axis is quantized according to (7.3), so that the fringes of Ψ in concavities near the torus $\mathscr{E}$ are sampled on lines whose spacing is $\hbar$. This sampling is semi-classically dense, because the fringes themselves have spacing $\hbar^{\frac{2}{3}}$ near $\mathscr{E}$, as the transitional approximation (4.8) shows. (ii) $\mathscr{E}$ must have inflexions (figure 4) since p_1 (equation 7.7) is periodic. Therefore the 'swallowtail' catastrophe behaviour of Ψ discussed in § 5 is generic for these rotational systems. (iii) The limited range of integration in (7.9) means that midpoints q, p_n of diameters of $\mathscr{E}$ longer than $X = \pi$ will not give stationary point contributions to Ψ. Therefore the fringes near $\mathscr{E}$ will fade out near the locus of midpoints of diameters of $\mathscr{E}$, with length π, the fading being given analytically in terms of the Fresnel integrals that describe the transition from light to shadow behind an illuminated edge (the mathematical origin of this similarity is the presence in both problems of a stationary point that can coalesce with an endpoint of the range of integration). The manner in which these 'shadow boundaries' of Ψ

interact with the catastrophes along the line $\mathscr{L}$ (figure 4) has not been considered here; it would make an interesting study.

A most instructive Wigner function is that representing the eigenstates of the non-generic rotator whose Hamiltonian is (5.14), for which the torus has the form (5.13). For this special case (7.9) gives

$$\Psi_m(q,p_n) = \frac{\mathrm{Re}}{\pi^2\hbar}\int_0^{\frac{1}{2}\pi} \mathrm{d}X \exp\left\{2\mathrm{i}(m-n)\,X + \frac{2\mathrm{i}\,A}{\hbar}\cos q \sin X\right\}. \tag{7.10}$$

Despite appearances this is not a standard Bessel function but belongs to the wider class of 'incomplete cylindrical functions' (Agrest & Maksimov 1971). Elementary manipulations reduce Ψ_m to the following alternative forms:

$$\Psi_m(q,p_n) = \frac{1}{\hbar}J_{2(m-n)}\left(\frac{2A\cos q}{\hbar}\right) - \frac{1}{\pi^2\hbar}\int_0^{\frac{1}{2}\pi} \mathrm{d}X \sin\left[2(m-n)\,X\right]\sin\left(\frac{2A}{\hbar}\cos q \sin X\right) \tag{7.11}$$

or

$$\Psi_m(q,p_n) = \frac{1}{\hbar}J_{2(m-n)}\left(\frac{2A\cos q}{\hbar}\right) - \frac{2(m-n)}{\pi^2\hbar}\frac{\partial}{\partial B}\prod_{k=1}^{|m-n|-1}\left(1+\frac{\partial/\partial B}{\sin^2\left(\dfrac{k\pi}{2|m-n|}\right)}\right)\frac{\sin B}{B}\left(B \equiv \frac{2A}{\hbar}\cos q\right). \tag{7.12}$$

When 'projected' along the momentum axis, (7.10) gives the correct coordinate probability density, namely

$$\begin{aligned}|\langle q|\psi_m\rangle|^2 &= \hbar\sum_n \Psi_m(q,p_n)\\ &= \frac{1}{2\pi}\end{aligned} \tag{7.13}$$

while 'projection' along the coordinate axis gives the correct momentum probability density, namely

$$\begin{aligned}|\langle p_n|\psi_m\rangle|^2 &= \int_{-\pi}^{\pi}\mathrm{d}q\,\Psi_m(q,p_n)\\ &= \frac{1}{\hbar}J_{m-n}^2\left(\frac{A}{\hbar}\right).\end{aligned} \tag{7.14}$$

Now, the surprising fact is that both of these results can be obtained from the first term alone of (7.11) or (7.12). It seems that the second terms are 'ghosts' with no observable consequences. Nevertheless they do play the important rôle of ensuring that Ψ_m has the correct classical limit of a delta function on the torus $\mathscr{E}$. To see this, first realize that for given q the torus lies at the momentum p_n given by

$$p_n = p(q, E_m), \tag{7.15}$$

which from (7.7), (7.8), (7.3) and (5.13) corresponds to

$$n = m + \frac{A\cos q}{\hbar}. \tag{7.16}$$

Now, in the semi-classical limit when n and m are both large the Bessel function in (7.11) reaches its largest values near places where its argument equals $\pm$ its order; these two peaks take the form of Airy functions (Abramowitz & Stegun 1964), situated at

$$n = m \pm \frac{A\cos q}{\hbar}. \tag{7.17}$$

The negative sign indicates that the Bessel function terms in (7.11) and (7.12) peak not only on $\mathscr{E}$ (equation (7.16)) but also on the 'anti-torus' $m - A\cos q$ (figure 6). Moreover, the two peaks have the same strength, equal to 'half a delta function', because the ghost terms do not contribute to the normalization of Ψ_m.

However, the ghost terms (which are odd functions of $m-n$), just cancel the delta function on the 'anti-torus' and double the strength of the torus on $\mathscr{E}$ (figure 6). This can be shown by calculating the 'mass' of the ghost terms in that half of momentum space for which $n > m$ using the form (7.11) as follows:

$$\begin{aligned}
&\hbar \sum_{n=m}^{\infty} \frac{1}{\pi^2\hbar}\int_0^{\frac{1}{2}\pi} \mathrm{d}X \sin\left[2(n-m)\,X\right] \sin\left[\frac{2A}{\hbar}\cos q \sin X\right] \\
&\quad = \frac{1}{\pi^2}\int_0^{\frac{1}{2}\pi} \mathrm{d}X \sin\left(\frac{2A}{\hbar}\cos q \sin X\right) \sum_{\nu=0}^{\infty} \sin 2\nu X \\
&\quad = \frac{1}{2\pi^2}\int_0^{\frac{1}{2}\pi} \mathrm{d}X \sin\left(\frac{2A}{\hbar}\cos q \sin X\right) \cot X \\
&\quad = \frac{1}{2\pi^2}\int_0^{1} \mathrm{d}y \frac{\sin\left(\frac{2A}{\hbar}\cos qy\right)}{y} \\
&\quad \xrightarrow{\hbar\to 0} \pm\frac{1}{4\pi} \quad \text{if} \quad \cos q \gtrless 0.
\end{aligned} \tag{7.18}$$

The delta function would correspond to $\frac{1}{2}\pi$ (to allow for subsequent integration over q), so that this is precisely the stated result, showing that the ghost terms are essential if Wigner's function is to have the correct classical limit. Deep in the shadow Ψ possesses weak fine fringes (figure 6), as shown in appendix G.

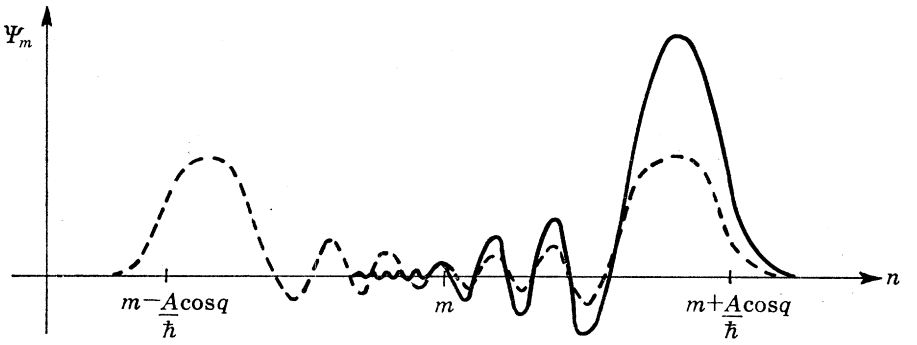

FIGURE 6. The full curve is a sketch of Wigner's function Ψ_m for the mth state of a hindered rotator (equation (7.10)), as a function of the momentum quantum number n. The dashed curve is Ψ with the 'ghost' terms in (7.11) or (7.12) left out.

Now we leave this special case, and return to the general f-dimensional case where as a semi-classical approximation the fundamental operators employed to evaluate Ψ are chosen to correspond to the classical action-angle variables $\boldsymbol{I}$ and $\boldsymbol{\theta}$, and will be denoted by $\hat{\boldsymbol{I}}$ and $\hat{\boldsymbol{\theta}}$. Now the Hamiltonian $H(\hat{\boldsymbol{I}})$ is diagonal in $\hat{\boldsymbol{I}}$, since the classical actions are constants of the motion. To find the eigenvalues $\boldsymbol{I}_n$ and, eigenfunctions $|\psi_n\rangle$, write the operator $\hat{\boldsymbol{I}}$ in the $\hat{\boldsymbol{\theta}}$ representation, as follows:

$$\hat{\boldsymbol{I}} = -\mathrm{i}\hbar\nabla_{\boldsymbol{\theta}} + \boldsymbol{\beta}\hbar, \tag{7.19}$$

where the components of $\boldsymbol{\beta}$ are f constants related to the number of caustics of the system's orbit in the original $\boldsymbol{q}$-space encountered during the irreducible circuits of the torus $\boldsymbol{I}$ (cf. § 3) corresponding to the angles $\boldsymbol{\theta}$ (see Maslov 1972; Percival 1977). The form (7.19) satisfies the basic commutation law

$$[\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{I}}] = \mathrm{i}\hbar. \tag{7.20}$$

Then the eigenvalues $\boldsymbol{I}_n$ are easily found to be

$$\boldsymbol{I}_n = (\boldsymbol{n} + \boldsymbol{\beta})\,\hbar \tag{7.21}$$

and the eigenfunctions of both $\hat{\boldsymbol{I}}$ and $\hat{\boldsymbol{H}}$ are

$$\langle \boldsymbol{\theta} | \psi_n \rangle = \frac{\mathrm{e}^{\mathrm{i}\boldsymbol{n}\cdot\boldsymbol{\theta}}}{\sqrt{(2\pi)}}. \tag{7.22}$$

The energy eigenvalues are $$E_n = H(\boldsymbol{I} = (\boldsymbol{n} + \boldsymbol{\beta})\,\hbar). \tag{7.23}$$

In this representation Wigner's function Ψ only has meaning for the quantized values $\boldsymbol{I}_m$ of $\boldsymbol{I}$, and for the state $|\psi_m\rangle$ takes a particularly simple form if the origin in $\boldsymbol{I}$ space is shifted to $\boldsymbol{I}_0 = \boldsymbol{\beta}$, namely

$$\begin{aligned}\Psi_m(\boldsymbol{\theta}, \boldsymbol{I}_n) &= \frac{1}{(\pi\hbar)^f}\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \mathrm{d}\Theta_1 \ldots \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \mathrm{d}\Theta_f \exp\left[-\frac{2\mathrm{i}}{\hbar}(\boldsymbol{I}_n - \boldsymbol{\beta})\cdot\boldsymbol{\Theta}\right] \langle \boldsymbol{\theta} + \boldsymbol{\Theta} | \psi_m \rangle \langle \psi_m | \boldsymbol{\theta} - \boldsymbol{\Theta} \rangle \\ &= \frac{1}{(2\pi)^f (\pi\hbar)^f}\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \mathrm{d}\Theta_1 \ldots \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \mathrm{d}\Theta_f \exp\left[2\mathrm{i}(\boldsymbol{m} - \boldsymbol{n})\cdot\boldsymbol{\theta}\right] \\ &= \frac{\delta_{m,n}}{(2\pi\hbar)^f}.\end{aligned} \tag{7.24}$$

This shows that when actions and angles are employed as fundamental operators the phase space consists only of the toroidal 'shells' $\boldsymbol{I} = \boldsymbol{I}_n$, and Wigner's function Ψ_m for the state $|\psi_m\rangle$ is entirely confined to the $\boldsymbol{m}$th torus. This behaviour contrasts strongly with that found in § 4 for Cartesian variables: there the phase space $\boldsymbol{q}, \boldsymbol{p}$ was continuous, and Ψ_m for the same quantum state took non-zero values over the whole phase space – even the 'first bright fringe' (equation (4.8)) has an 'action width' of order $\hbar^{\frac{2}{3}}$ and so includes many of the aforementioned toroidal shells (whose separation is $\hbar$). We are merely illustrating here the well-known fact that Wigner's function is not invariant under canonical transformations of the phase space $\boldsymbol{q}, \boldsymbol{p}$. (It should be pointed out that *any* representation of Ψ can be employed to find the expectation value of *any* operator $\hat{a}$ in the state $|\psi\rangle$ using (2.8), provided the Weyl function $A(\boldsymbol{q}, \boldsymbol{p})$ (equation 2.1) is evaluated in the same representation as Ψ.)

8. Integrable systems: the semi-classical limiting form

Now we return to the case where the fundamental operators $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{p}}$ do not correspond to angles and angular momenta but have continuous unbounded spectra. Our intention is to derive from the basic semi-classical integral representation (3.6) multi-dimensional analogues for general integrable systems of the simple semi-classical approximation (4.6), the transitional approximation (4.8) and the uniform approximation (4.7), but we have only partially succeeded in carrying out this programme.

Classical motion corresponding to the energy eigenstate $|\psi_m\rangle$ is confined to the torus $\boldsymbol{I} = \boldsymbol{I}_m$, which will be called $\mathscr{T}$: henceforth the dependence on $\boldsymbol{I}_m$ in (3.6) will not be written explicitly.

As in § 4, the first step in finding the semi-classical limiting form of $\Psi_m(q, p)$ is the determination of the stationary points of the phase in (3.6). These are given by

$$\tfrac{1}{2}[\boldsymbol{p}(\boldsymbol{q}+\boldsymbol{X})+\boldsymbol{p}(\boldsymbol{q}-\boldsymbol{X})] = \boldsymbol{p}, \tag{8.1}$$

which is an obvious generalization of (4.2) and has an analogous 'symmetric' interpretation, namely that the points on $\mathscr{T}$ contributing to Ψ are the ends of those chords of $\mathscr{T}$ whose midpoints are $\boldsymbol{q}, \boldsymbol{p}$.

As before, each chord defines two solutions of (8.1), $\boldsymbol{X} = +\boldsymbol{X}_0(\boldsymbol{q}, \boldsymbol{p})$ and $\boldsymbol{X} = -\boldsymbol{X}_0(\boldsymbol{q}, \boldsymbol{p})$. Now, however, there will in general be *many* chords of $\mathscr{T}$ whose midpoints are $\boldsymbol{q}, \boldsymbol{p}$. This can be seen by considering the *separable* case, where each component $p_i(\boldsymbol{q})$ depends only on the corresponding q_i. Then equation (8.1) decouples into f separate equations, and *each component* of $\boldsymbol{X}$ can be positive or negative, giving 2^f distinct solutions altogether, or 2^{f-1} vectors $\boldsymbol{X}_0$ together with their negatives $-\boldsymbol{X}_0$. By continuity, all these solutions will persist when the system is perturbed to make it *non-separable* (but still integrable).

Let the 2^f solutions be denoted by

$$\boldsymbol{X} = \pm \boldsymbol{X}_\alpha(\boldsymbol{q}, \boldsymbol{p}) \; (1 \leqslant \alpha \leqslant 2^{f-1}), \tag{8.2}$$

and define

$$\mathscr{A}_\alpha(\boldsymbol{q}, \boldsymbol{p}) \equiv \int_{\boldsymbol{q}-\boldsymbol{X}_\alpha}^{\boldsymbol{q}+\boldsymbol{X}_\alpha} \boldsymbol{p}\cdot \mathrm{d}\boldsymbol{q} - 2\boldsymbol{p}\cdot\boldsymbol{X}_\alpha \tag{8.3}$$

(cf. 4.3). Also denote by 1 and 2 the points on $\mathscr{T}$ corresponding to $\boldsymbol{X} = -\boldsymbol{X}_\alpha$ and $\boldsymbol{X} = +\boldsymbol{X}_\alpha$. Now the method of stationary phase can be applied to (3.6), the denominator being simplified with the aid of the multi-dimensional analogue of (4.5). This gives, for the case where $\boldsymbol{X}_\alpha$ is small (i.e. $\boldsymbol{q}, \boldsymbol{p}$ not near $\mathscr{T}$)

$$\Psi_m(\boldsymbol{q}, \boldsymbol{p}) = \frac{2}{\pi^f \hbar^{f/2}} \sum_{\alpha=1}^{2^{f-1}} \frac{\cos\left[(\mathscr{A}_\alpha(\boldsymbol{q}, \boldsymbol{p})/\hbar) + n_\alpha \pi/4\right]}{\left|\det\left[\left(\frac{\partial I_i}{\partial q_l}\right)_2 \left(\frac{\partial I_l}{\partial p_j}\right)_1 - \left(\frac{\partial I_i}{\partial q_l}\right)_1 \left(\frac{\partial I_l}{\partial p_j}\right)_2\right]\right|^{\frac{1}{2}}}, \tag{8.4}$$

which is the generalization of (4.6). n_α is the excess of positive over negative eigenvalues of the matrix whose determinant appears in the denominator. The phase $\mathscr{A}_\alpha$ has the following interpretation: it is the action around the closed curve C_α consisting of the segment $1 \to 2$ on $\mathscr{T}$ and the straight segment $2 \to 1$ through $\boldsymbol{q}, \boldsymbol{p}$. Alternatively, $\mathscr{A}_\alpha$ is the sum of the areas of the projections of C_α on the f separate $q_i p_i$ planes.

In the special case where the system is *separable*, the determinant factorizes, because each component I_i depends only on the corresponding q_i and p_i. Moreover, the A_α and n_α can be written as follows:

$$\left.\begin{aligned} \mathscr{A}_\alpha &= \sum_{i=1}^{f} A(q_i, p_i)\,(-1)^{u_{i\alpha}} \\ n_\alpha &= -\sum_{i=1}^{f} (-1)^{u_{i\alpha}} \end{aligned}\right\}, \tag{8.5}$$

where $A(q_i, p_i)$ is given by (4.3) with q_i, p_i replacing q, p, $u_{i\alpha}$ for $2 \leqslant i \leqslant f$ is one of the 2^{f-1} permutations of $f-1$ zeroes and ones and $u_{1\alpha} = 0$. Then with the aid of the identity

$$2\sum_{\alpha=1}^{2^{f-1}} \cos\left[\sum_{i=1}^{f} a_i(-1)^{u_{i\alpha}}\right] = 2^f \prod_{i=1}^{f} \cos a_i \tag{8.6}$$

(which can be proved by writing the cosines in terms of exponentials), (8.4) becomes

$$\Psi_m(\boldsymbol{q},\boldsymbol{p}) = \frac{2^f}{\pi^f h^{f/2}} \prod_{i=1}^{f} \frac{\cos\left[A(q_i,p_i)/\hbar - \tfrac{1}{4}\pi\right]}{\left[\left(\frac{\partial I_i}{\partial q_l}\right)_1 \left(\frac{\partial I_l}{\partial p_j}\right)_2 - \left(\frac{\partial I_i}{\partial q_l}\right)_2 \left(\frac{\partial I_l}{\partial p_j}\right)_1\right]^{\frac{1}{2}}}; \tag{8.7}$$

this can also be derived directly from (3.6)

The following question naturally arises: it is possible to factorize (8.4) into a form like (8.7) in the general non-separable case? If the answer were affirmative, then the extension of validity of (8.4) to the points $\boldsymbol{q}, \boldsymbol{p}$ near and on $\mathscr{T}$ could be accomplished simply by substituting the uniform 'Airy' expression (4.7) for each factor in (8.7). Unfortunately we are not able to settle this question, but we can make two remarks about it.

First, it is always possible, by making a simple canonical transformation in $\boldsymbol{q}, \boldsymbol{p}$ corresponding to a rotation of coordinates $\boldsymbol{q}$, to diagonalize the quadratic form that arises when the phase in (3.6) is expanded as a function of $\boldsymbol{X}$ about $\boldsymbol{X}_\alpha$. This will cause the determinant in the term α in (8.4) to factorize. However, it is not obvious, and indeed seems unlikely, that the same coordinate rotation will cause the determinants in the other terms α to factorize.

Secondly, it is nevertheless likely at least for $f = 2$ that a 'local separability' holds *close to the torus* $\mathscr{T}$, i.e. to lowest order in the 'action distance' $\boldsymbol{I}(\boldsymbol{q},\boldsymbol{p}) - \boldsymbol{I}_m$. To see this it is first necessary to seek a 'transitional approximation' analogous to (4.8) by expanding the phase in (3.6) to third order in $\boldsymbol{X}$ and setting $\boldsymbol{X} = 0$ in the determinants. Then a natural change of variables is made, from $\boldsymbol{X}$ to the 'local' angle coordinate $\boldsymbol{\Theta}$, given by (cf. 3.3)

$$\boldsymbol{\Theta} \equiv \boldsymbol{X} \cdot \nabla_{\boldsymbol{I}} \boldsymbol{p}(\boldsymbol{q}, \boldsymbol{I}_m). \tag{8.8}$$

A tedious transformation (appendix H) now gives

$$\Psi_m(\boldsymbol{q},\boldsymbol{p}) = \left(\frac{1}{\pi\hbar}\right)^f \int \mathrm{d}\boldsymbol{\Theta} \exp\left\{\frac{\mathrm{i}}{\hbar}\left[2(\boldsymbol{I}_m - \boldsymbol{I}(\boldsymbol{q},\boldsymbol{p}))\cdot\boldsymbol{\Theta} - \tfrac{1}{3}T_{abc}\,\Theta_a\,\Theta_b\,\Theta_c\right]\right\}, \tag{8.9}$$

where the matrix T_{abc}, defined by

$$T_{abc} \equiv \frac{\partial^2 I_a}{\partial p_i \partial p_j}\frac{\partial I_b}{\partial q_i}\frac{\partial I_c}{\partial q_j} + \frac{\partial^2 I_a}{\partial q_i \partial q_j}\frac{\partial I_b}{\partial p_i}\frac{\partial I_c}{\partial p_j} - 2\frac{\partial^2 I_a}{\partial p_i \partial q_j}\frac{\partial I_b}{\partial q_i}\frac{\partial I_c}{\partial p_j}, \tag{8.10}$$

is the generalization of the quantity B in equation (4.9).

Now the question of factorizability hinges on whether the symmetric cubic form in (8.9) can be expressed as a sum of cubes of the separate components of $\boldsymbol{\Theta}$ by rotating axes in $\boldsymbol{\Theta}$ space. In general such 'triagonalization' is not possible. In the case $f = 2$, for example, any cubic form can be transformed by real linear transformations into one of the following two inequivalent forms:

$$\Theta_1^3 + \Theta_2^3 \quad \text{or} \quad \Theta_1^3 - 3\Theta_1\Theta_2^2. \tag{8.11}$$

These are, respectively, the 'germs' of the hyperbolic and elliptic umbilic catastrophes (Thom 1975). Only the hyperbolic umbilic has the required triagonal form. (It is possible to transform the second form in (8.11) into the first with a *complex* transformation, but this is of no use here because the resulting triagonalization would involve an impermissable contour transformation in the integral (8.9)). For $f > 2$ the situation is worse because there are *infinitely many* inequivalent cubic forms (see for example the unimodal 'parabolic' germ P_8 in Arnol'd 1974, 1975).

However, if the non-separable system is a perturbed separable system (for which T_{abc} is triagonal) then if $f = 2$ it is obvious that the cubic form in (8.9) will remain equivalent to the hyperbolic rather than the elliptic form in (8.11), provided the perturbation is not so large

that a 'parabolic umbilic' catastrophe (Thom 1975) occurs with a sudden switch to the elliptic form. For such 'perturbed' systems where triagonalization is possible we can write

$$T_{abc} \equiv T_a(\boldsymbol{q}, \boldsymbol{p})\, \delta_{a,b}\, \delta_{b,c} \tag{8.12}$$

in the appropriately rotated coordinates $\boldsymbol{\Theta}$, and Wigner's function near the torus becomes

$$\Psi_m(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{\pi^f \hbar^{\frac{2}{3}f}} \prod_{a=1}^{f} \frac{\mathrm{Ai}\,[2\{I_a(\boldsymbol{q}, \boldsymbol{p}) - I_{ma}\}\{\hbar^2 T_a(\boldsymbol{q}, \boldsymbol{p})\}^{-\frac{2}{3}}]}{\{T_a(\boldsymbol{q}, \boldsymbol{p})\}^{\frac{1}{3}}}; \tag{8.13}$$

this generalizes (4.8), at least for 'perturbed' systems with two degrees of freedom and possibly for a wider class of cases.

Now we consider briefly the *multi-dimensional catastrophes* of Ψ, i.e. the places where (8.4) diverges because of the vanishing of the determinants in the denominators. Such a catastrophe occurs on $\mathscr{T}$ itself, and we have just found that at least for 'perturbed' systems with $f = 2$ the catastrophe is given by the 'generalized hyperbolic umbilic' whose germ is a sum of cubes and whose 'diffraction function' (Berry 1976) is (8.13); this generalizes the 'fold' catastrophe function (4.8) for one-dimensional systems.

However, the catastrophes on $\mathscr{T}$ are *non-generic*, because the full 'unfolding' of the generalized hyperbolic umbilics would involve quadratic terms in Θ and these are absent from (8.9) because the phase in (3.6) is an odd function of $\boldsymbol{X}$ (a non-generic property). The actual catastrophes of Ψ are special sections of the generalized hyperbolic umbilics. They exist in the f-dimensional subspace of $\boldsymbol{q}$, $\boldsymbol{p}$ whose 'control' variables are the actions $\boldsymbol{I}$ – the angles $\boldsymbol{\theta}$ play no rôle because Ψ varies smoothly round each torus (we ignore the situation where $\mathscr{T}$ has 'multi-dimensional inflexions' corresponding to zeros of T_a (equation (8.12) and generating 'multi-dimensional swallowtails' (cf. § 5)). The result (8.13) shows that the Airy function (fold catastrophe) behaviour can appear off $\mathscr{T}$, when *any* of the components of $\boldsymbol{I}(\boldsymbol{p}, \boldsymbol{q})$ equals the corresponding component of $\boldsymbol{I}_m$. In action space, therefore, the torus $\mathscr{T}$ (i.e. $\boldsymbol{I} = \boldsymbol{I}_m$) is the corner of a 'rectangular' catastrophe surface; a two-dimensional example of the diffraction function (8.13) near such a corner can be seen on figure 9 of Berry (1976), if the axes of that figure are relabelled I_1 and I_2.

There will also be catastrophes far from $\mathscr{T}$, analogous to the line $\mathscr{L}$ on figure 3. These will inhabit the full $2f$-dimensional control space $\boldsymbol{q}$, $\boldsymbol{p}$, and will rapidly increase in complexity with f. Even when $f = 2$ all seven 'elementary catastrophes' (Thom 1975) can occur generically. A full description of any such situation would involve the multi-dimensional generalization of figure 3 and is beyond our present understanding.

9. Non-integrable systems

For the integrable systems so far considered the f-dimensional tori (§ 3) to which classical systems are confined have been central in understanding the semi-classical structure of Wigner's function. Now we turn our attention to non-integrable systems, where f constants of motion do not exist (this implies $f \geqslant 2$ because energy is always conserved for systems of interest here), so that tori do not fill the phase space and some orbits can explore regions whose dimensionality exceeds f. In recent decades major advances in classical mechanics (reviewed by Arnol'd & Avez 1968; Moser 1973 and Ford 1975) have led to some understanding of the main features of non-integrable motion. In particular, motion in regions where tori do not exist is 'stochastic' in nature; orbits are unstable and rapidly lose all 'memory' of their initial conditions; such regions of phase space are called 'irregular'.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Irregular regions pose profound difficulties for the semi-classical limit of quantum mechanics: if there are no tori there are no actions $\boldsymbol{I}$ either, and hence no quantum numbers $\boldsymbol{n}$ (equation (7.21)), and the semi-classical eigenvalue formula (7.23) become meaningless. This was first pointed out by Einstein (1917) in a paper whose importance was first appreciated by Percival (1973, 1976).

Nevertheless, quantum states (defined by Schrodinger's equation plus boundary conditions) must still exist in irregular regions. Moreover, their mean density in energy $\rho(E)$ is known to be

$$\rho(E) = \frac{1}{h^f}\int \mathrm{d}\boldsymbol{q}\int \mathrm{d}\boldsymbol{p}\,\delta(E-H(\boldsymbol{q},\boldsymbol{p})), \tag{9.1}$$

where H is the classical Hamiltonian. This is most easily obtained as the classical limit of the trace of (2.2), where $\hat{a}$ is the operator $\delta(E-\hat{H})$. Equation (9.1) simply formalizes the old rule that each quantum state 'occupies' a volume h^f in phase space. The rule holds for integrable and non-integrable systems alike. For integrable systems it follows from the quantization rule (7.21): each state has volume $\hbar^f$ in action space, but each torus (i.e. each value of $\boldsymbol{I}$) has an 'angle' volume $(2\pi)^f$, so that the total phase space volume of the state is h^f. (This result might seem to conflict with the semi-classical formula (8.13) for Wigner's function, according to which even the first bright fringe of Ψ occupies a much larger volume, of order $\hbar^{2f/3}$. However, this is an artefact of the $\boldsymbol{q}, \boldsymbol{p}$ representation, and Ψ for an energy eigenstate will attain its minimum spread in a representation in which $\hat{H}$ is diagonal, as equation (7.24) shows.)

Percival (1973) conjectures that quantum states associated with irregular motion will themselves be very irregularly distributed and sensitive to perturbation; computations of Pomphrey (1974) lend support to the latter part of this conjecture. Lloyd (private communication) suggests that since the distinction between integrable and non-integrable motion is manifest in phase space, the natural quantum object for studying the corresponding semi-classical mechanics is Wigner's function $\Psi(\boldsymbol{q}, \boldsymbol{p})$. Nordholm & Rice (1974) make the same suggestion. This has led us to some conjectures about the semi-classical limit of Ψ, which we now describe.

In the simplest non-trivial case, the system has two degrees of freedom, and a simple pictorial device is available for studying the phase space. This is *Poincaré's surface of section S*, defined as follows: each energy shell $\mathscr{E}$, given by

$$H(q_1\,q_2 p_1 p_2) = E, \tag{9.2}$$

is a three-dimensional manifold in the four-dimensional phase space. S is the two-dimensional section $q_2 = 0$ of $\mathscr{E}$, with coordinates q_1, p_1. Specifying a point on S completely specifies the system, because q_2 is defined as zero and p_2 is obtained from (9.2) as a function of E, q_1 and p_1 up to a sign which can be defined as positive. For example, if q_1 and q_2 are the radial and azimuthal polar coordinates of a particle moving non-relativistically in a potential $V(q_1, q_2)$, then

$$p_2(q_1, p_1; E) = +q_1\sqrt{[2m(E-V(q_1, 0)) - p_1^2]}. \tag{9.3}$$

A corresponding section can be taken through Wigner's function $\Psi_m(\boldsymbol{q}, \boldsymbol{p})$, giving a density $W_m(q_1, p_1)$ on the surface S whose energy is E_m i.e.

$$W_m(q_1, p_1) \equiv \Psi_m(q_1, 0, p_1, p_2(q_1, p_1; E_m)). \tag{9.4}$$

Consider first an integrable system, for example, a particle bound in a central potential $V(q_1)$ of Lennard–Jones type. Then phase space is filled with tori $\boldsymbol{I}$ which are two-dimensional submanifolds of the energy shells $\mathscr{E}$, cutting Poincaré's surface S in *closed curves*. If a particle starts out from

S on a point on such a curve, all the subsequent intersections of its orbit with S will lie on the same curve. Figure 7*a* shows the 'invariant curves' on S for the example just given; the energy of this section is chosen to equal the energy E_m of a quantum state, and the curve C shown dotted is the intersection of the torus $\boldsymbol{I}_m$ with S. The corresponding Wigner density $W_m(q_1, p_1)$ is concentrated in fringes near C, shown on figure 7*b*, whose form can be obtained from equation (8.13).

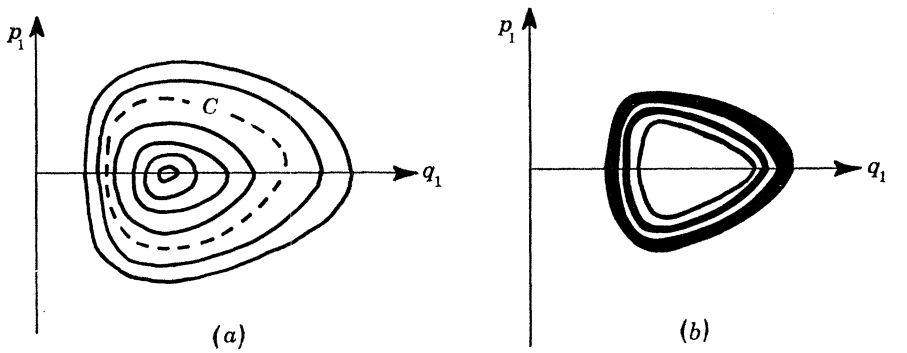

FIGURE 7. (*a*) Poincaré's surface of section S for the radial coordinate and momentum of a particle bounded by a central potential in two dimensions, corresponding to the energy E_m of the quantum state $|\psi_m\rangle$. The outermost curve has zero angular momentum, and the innermost curve has the maximum angular momentum for which orbits exist at E_m. C is the intersection of the torus $\boldsymbol{I}^m$ with S. (*b*) Intensity map of Wigner's density $W_m(q_1, p_1)$ on S.

Now make the system non-integrable by a perturbation whose strength is governed by a parameter ϵ. Then it was proved by Kolmogoroff (1954), Arnol'd (1963) and Moser (1962) (see also the references cited earlier) that for small but finite ϵ most tori continue to exist, albeit distorted. Therefore most of S is still covered with invariant curves. However, some tori are destroyed; these lie near 'resonant' tori $\boldsymbol{I}$ whose fundamental unperturbed frequencies

$$\boldsymbol{\omega} = \nabla_{\boldsymbol{I}} H(\boldsymbol{I}) \tag{9.5}$$

are *commensurable*, corresponding to closed orbits on $\mathscr{T}$. For two-dimensional systems with frequency ratio ω_1/ω_2 the resonances are given by rational fractions r/s (in their lowest terms) and the destroyed tori near r/s are those for which

$$|\omega_1/\omega_2 - r/s| < K(\epsilon)/s^{2.5}, \tag{9.6}$$

where $K(\epsilon)$ vanishes with ϵ. This can be depicted in unperturbed action space $\boldsymbol{I}$ as on figure 8: commensurable frequencies correspond to rational directions of $\boldsymbol{\omega}$, and from equation (9.5) this implies that resonant tori lie on lines in $\boldsymbol{I}$ which are the loci of points where the 'energy contours' $H = E$ have rational normals; the 'resonant zones' (9.6) are narrow sectors containing these lines. Of course there are infinitely many rational normals arbitrarily close to any point on an energy contour, so that the resonant zones are pathologically distributed; however, the higher-order zones are narrow (because of the factor $s^{-2.5}$ in 9.6) and the total measure of all the zones can be shown to be finite (in fact of order $K(\epsilon)$).

Phase space within each resonant zone has a very complicated structure. Part of the zone is filled with new tori centred on particular closed orbits of the original resonant tori. The rest of the zone is filled with irregular trajectories. Figure 9*a* gives some idea of the surface of section S. The whole structure repeats itself down to infinitely fine scales, because the system of new tori in the original resonant zones will contain its own resonant zones, filled with new tori and irregular

regions and so on ad infinitum. It should be emphasized that this marvellous richness of behaviour is *generic* for classical systems; the integrable case already discussed, and the 'ergodic' limit envisaged in statistical mechanics where all trajectories are irregular, are special limiting cases.

It is obvious that the semi-classical limit will be complicated for these perturbed integrable systems. For fixed ϵ, $\hbar$ must be made ever smaller and the changing behaviour of the eigenvalues and eigenfunctions investigated. There are *three régimes* distinguished by values of a parameter β which will now be defined. Surrounding any point $\boldsymbol{I}$ in unperturbed action space is the area $\hbar^2$ corresponding to a quantum state. This area will be crossed (figure 8) by infinitely many resonance zones, the widest of which has the frequency ratio r/s with smallest s (equation 9.6).

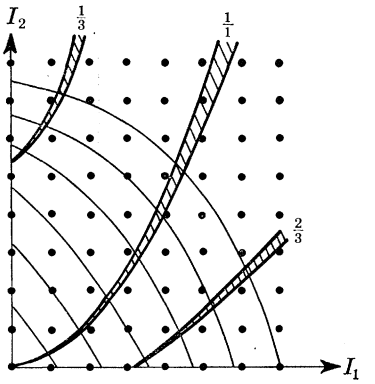

FIGURE 8. Low-order resonance zones (shaded), where normals to energy contours (full lines) have rational directions. Quantum states (lattice points) divide $\boldsymbol{I}$ into cells with volume $\hbar^2$.

Between two energy contours whose perpendicular separation $|\Delta \boldsymbol{I}|$ is $\hbar$ this widest resonance will occupy an action area A. Then β is defined as $A/\hbar^2$. When β is small it is a measure of the proportion of the quantum area $\hbar^2$ occupied by irregular trajectories. Large values of β indicate that the resonant zones near $\boldsymbol{I}$ contain many quantum states. Elementary geometry and use of (9.6) give

$$\beta \equiv \frac{A}{\hbar^2} = \frac{\hbar \times \text{width of widest resonance}}{\hbar^2}$$

$$\approx \frac{\hbar|\boldsymbol{I}| \times \text{angular width of widest resonance}}{\hbar^2} \approx \frac{K(\epsilon)|\boldsymbol{I}|}{\hbar s^{2.5}} \tag{9.7}$$

as a rough estimate for β. Therefore β is large in the semi-classical limit $\hbar \to 0$, and also for large perturbations ϵ, low-order resonances (small s) and high excited states (large $|\boldsymbol{I}|$).

In the *first semi-classical régime*, $\hbar$ is small enough for a semi-classical treatment of the unperturbed system to be valid, but ϵ is also small enough for β to be small for all $\boldsymbol{I}$ in the energy region of interest, even those crossed by the lowest-order resonance zones ($s = 1$). The irregular regions occupy only a small fraction of any quantum area $\hbar^2$ and so do not affect the form of the states $|\psi\rangle$ or Wigner functions Ψ – Planck's constant $\hbar$ blurs all the classical fine structure. In these circumstances the semi-classical formula (7.23), with approximately calculated actions $\boldsymbol{I}$, can be employed to locate the perturbed quantum levels E_m. We believe that recent highly successful computations based on (7.23) for non-integrable systems, by Chapman, Garrett & Miller (1976), Noid & Marcus (1975), Percival & Pomphrey (1976) and Handy, Colwell & Miller (1976), were all in this first semi-classical régime. (In these studies the unperturbed systems were harmonic oscillators, for which the energy contours (figure 8) are flat, and no low-order resonances existed.)

Making $\hbar$ smaller leads to the *second semi-classical régime*, in which β is of order unity for states whose actions $\boldsymbol{I}$ lie in the lowest resonance zones. Then there will be a few states whose quantum area is dominated by irregular trajectories. The energies of these states will be given approximately by (7.23), but their wave functions can no longer be approximated by the W.K.B. method, and their Wigner functions Ψ will no longer be concentrated on the fringed tori discussed earlier, because the tori in the relevant regions of phase space have been destroyed. Instead we conjecture that the Wigner density $W(q_1, p_1)$ on the surface of section *will spread over the region* (figure 9*a*) *occupied by the irregular trajectories, taking the form of a series of random maxima and minima* (figure 9*b*) resembling the diffraction structure of random wave fields (see e.g. Cochran 1973). (In the whole phase space such a Wigner function would still be concentrated near the energy shell $\mathscr{E}$, and so need not violate the basic inequality (2.11).) A similar conjecture was made, and to a certain extent supported by numerical evidence, by Nordholm & Rice (1974) for the non-generic case where the unperturbed system is a set of harmonic oscillators.

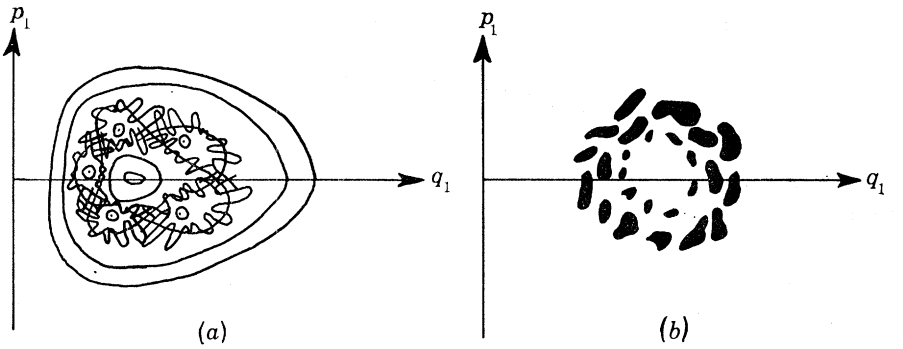

FIGURE 9. (*a*) Surface of section corresponding to figure 7*a* with a non-integrable perturbation, showing irregular trajectories and new tori in the destroyed zone around a low-order resonance (after Arnol'd & Avez 1968). (*b*) Conjectured intensity map of Wigner's density $W(q_1, p_1)$ for a state whose corresponding classical motion is predominantly irregular.

Further diminishing $\hbar$ leads to the *third semi-classical régime*; which is the semi-classical limit proper. Now β is large for $\boldsymbol{I}$ in all lower-order resonance zones. The corresponding irregular regions in phase space will be densely populated with quantum states, corresponding to the lattice spacing in figure 8 getting very small so that each zone is dense with lattice points. The group of states in each irregular region cannot now be individually labelled with quantum numbers (Percival 1973), although they may be said to share a 'vague quantum number' corresponding to the destroyed region $\boldsymbol{I}$ of action space. The semi-classical quantization rule (7.23) will not now be applicable in any sense. The Wigner function Ψ for any single state will spread over the whole of the irregular region of the energy shell, and the Wigner density $W(q_1, p_1)$ on S is conjectured to resemble figure 9*b* but with a much finer granularity in the random structure.

This does not exhaust the description of the generic structure of the third semi-classical régime, because there will be points $\boldsymbol{I}$ in high-order resonance zones where β is of order unity, and points $\boldsymbol{I}$ in still higher-order zones where β is small. Therefore along with the groups of 'irregular' states just described there will also be states of the type described for the first and second régime. What seems to be happening is that the smaller values of $\hbar$ expose more of the *infinite heterogeneity of the classical structure*, so that the quantum states are more varied in structure as well as more numerous.

When ϵ is zero this heterogeneity of structure is absent, because the system is integrable and there are no irregular regions; the Wigner density for any state resembles figure 7*b*. When ϵ is

large this heterogeneity is also absent, because the resonance zones have expanded and eaten away all the tori and all motions are irregular; the Wigner density for any state is then expected to resemble figure 9*b* with the disorder spread over the whole accessible area of S. The schematic figure 10 summarizes our picture of the generic semi-classical limit.

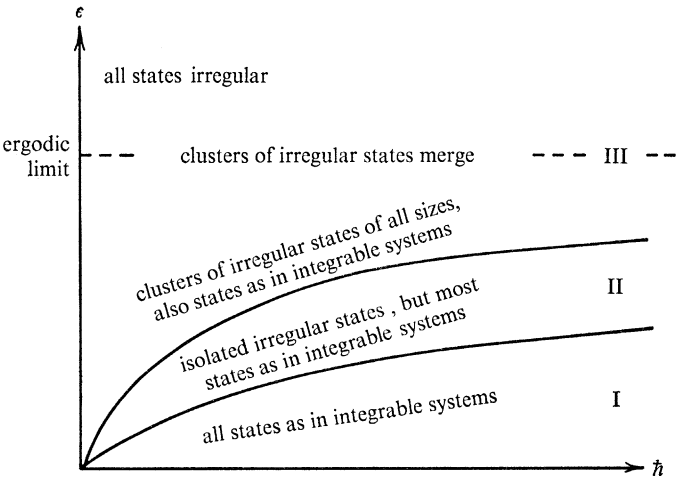

FIGURE 10. Schematic structure of semi-classical limit ($\hbar \to 0$) for systems with non-integrability parameter ϵ. The first, second and third semi-classical régimes are labelled I, II and III.

To test these conjectures, it would be necessary to compute exact solutions of Schrodinger's equation for states $|\psi\rangle$ in the second and third semi-classical régimes, and then calculate Wigner's density $W(q_1, p_1)$ for these states. From the fact that (except for the work of Nordholm & Rice (1974)) the most sophisticated computations to date (see the references cited) have been limited to the first semi-classical régime it is clear that the determination of these states $|\psi\rangle$ will not be easy. Once such $|\psi\rangle$ have been found, however, the determination of the Wigner density should not present any great difficulties. This is because computations on non-integrable systems invariably obtain $|\psi\rangle$ in terms of some simple basis states $|\phi_n\rangle$, i.e.

$$|\psi\rangle = \sum_n \langle\phi_n|\psi\rangle|\phi_n\rangle, \tag{9.8}$$

in terms of which Wigner's function $\Psi(\boldsymbol{q}, \boldsymbol{p})$ is

$$\Psi(\boldsymbol{q},\boldsymbol{p}) = \sum_{n_1}\sum_{n_2}\langle\phi_{n_1}|\psi\rangle\langle\psi|\phi_{n_2}\rangle\Phi_{n_1 n_2}(\boldsymbol{q},\boldsymbol{p}), \tag{9.9}$$

where $\Phi_{n_1 n_2}$ are the 'phase space eigenfunctions' of Moyal (1949), defined by (cf. 2.5)

$$\Phi_{n_1 n_2}(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{(\pi\hbar)^f}\int \mathrm{d}\boldsymbol{X}\exp\left(-\frac{2\mathrm{i}}{\hbar}\boldsymbol{p}\cdot\boldsymbol{X}\right)\langle\boldsymbol{q}+\boldsymbol{X}|\phi_{n_1}\rangle\langle\phi_{n_2}|\boldsymbol{q}-\boldsymbol{X}\rangle. \tag{9.10}$$

For simple basis sets $|\phi_n\rangle$, such as harmonic oscillators (Bartlett & Moyal 1949) and particles in boxes, the phase space eigenfunctions can be calculated analytically. Then the summation (9.9) is trivial given the expansion coefficients $\langle\phi_n|\psi\rangle$, and the Wigner density $W(q_1, p_1)$ is given by (9.4).

10. Conclusions

This paper in no way exhausts its subject, and in now summarizing our principal conclusions we shall emphasize the problems that remain.

For classically *integrable systems* we found that Wigner's function Ψ_m condenses in the classical limit onto the delta function (3.9) on the torus $\mathscr{T}$ (selected by the quantum condition 7.21) to which classical motion is confined. In the semi-classical limit (i.e. $\hbar$ small but not zero) this delta function softens into a set of fringes near $\mathscr{T}$, whose analytic form is given in one-dimensional systems by (4.8). These results were obtained, and their significance understood, with the aid of the geometrical 'diameter' construction of § 4 (see figure 1).

Thom's beautiful theory of catastrophes (Thom 1975) was invoked in §§ 5 and 6 to give a detailed description of the generic semi-classical singularities of Ψ and an understanding of those singularities that are non-generic (as with the harmonic oscillator in equation (5.11)).

For multi-dimensional integrable systems the discussion (§ 8) was incomplete owing to the need for an assumption of 'local separability' in order to derive the transitional approximation (8.13). Somebody should study this assumption, to see whether it holds always, or (as we suspect) only near $\mathscr{T}$ for $f \leqslant 2$, or not at all.

Fundamental operators corresponding to angle coordinates (§ 7) presented no problems if the angles represented spatial rotations, provided Ψ was evaluated only at quantized values of the conjugate angular momentum.

When extended to fundamental operators $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{I}}$ corresponding to the angle and action variables of classical mechanics (§ 3) the theory gave results (equations (7.19)–(7.24)) that seem to be consistent. However, since it is known (Leaf 1969) that these operators do not exist, the procedure adopted here should be subjected to a close scrutiny to determine the circumstances in which it is a valid approximation. (In particular the shift of origin to $\boldsymbol{I} = \boldsymbol{\beta}$ in the derivation of 7.24 should be examined.)

The deepest problems, however, concern classically *non-integrable systems*, studied in § 9. It seems that the semi-classical limit in such cases is very complicated, the main point (cf. figure 10) being that the limits $\epsilon \to 0$ (switching off non-integrability)and $\hbar \to 0$ (switching off quantum mechanics) cannot be interchanged. We conjectured that Wigner's function Ψ for a state in the irregular spectrum of Percival (1973) will have a random structure that could be computed on the Poincaré surface of section as the density $W(q_1, p_1)$ of equation (9.4). This should look like figure 9*b*, in contrast with the behaviour of a state that is not irregular, whose W function should look like figure 7*b*.

These conjectures can and should be tested by direct computation of quantum eigenstates in the three semi-classical régimes identified in § 9, along the lines of the pioneering work by Pomphrey (1974) and Nordholm & Rice (1974). However, the crying need is for an *analytical* theory of 'non-integrable semi-classical mechanics'. Since the classical trajectories associated with an irregular state $|\psi\rangle$ are not smoothly distributed in phase space on the scale of $\hbar$, diffraction will frustrate any attempt to use W.K.B. solutions to describe $|\psi\rangle$ and hence Ψ; another way of expressing this is to say that no global solution of the Hamilton–Jacobi equation exists (cf. Einstein 1917). Two alternative approaches suggest themselves. The first is based on the possibility that $\langle \boldsymbol{q}|\psi\rangle$ is some kind of random wave function, and would be an adaptation of techniques being developed (Tatarski 1961; Prokhorov *et al.* 1975; Elliott *et al.* 1974) to study waves in systems with random Hamiltonians (I am indebted to Dr G. V. Woolley for this suggestion). The

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

second approach is based on the strong similarity between the third semi-classical régime of § 9 (cf. III on figure 10), and the region near a critical point in statistical mechanics (i.e. the presence in both problems of structure on all scales), and would be an adaptation of the 'renormalization group' technique being developed by Wilson (1975) (see also Fisher 1974).

I thank Mr M. O. Hongler, Dr P. Lloyd, Dr N. Mukunda and Dr M. Tabor for helpful discussions, and Professor K. P. Sinha for the hospitality of the Indian Institute of Science where this work was carried out.

References

Abramowitz, M. & Stegun, I. A. 1964 *Handbook of mathematical functions*. Washington: US National Bureau of Standards.
Agrest, M. M. & Maksimov, M. S. 1971 *Theory of incomplete cylindrical functions and applications*. Berlin: Springer.
Arnol'd, V. I. 1963 *Usp. Mat. Nauk*. **18**, no. 5, 13–39; no. 6, 91–196. (English translations in *Russ. Math. Surv.* **18**, no. 5, 9–36; no. 6, 85–191.)
Arnol'd, V. I. 1974 *Usp. Mat. Nauk*, **29**, no. 2, 11–49 (English translation in *Russ. math. surv.* **29**, no. 2, 10–50).
Arnol'd, V. I. 1975 *Usp. Mat. Nauk*, **30**, no. 5, 3–65 (English translation in *Russ. math. surv.* **30**, no. 5, 1–75).
Arnol'd, V. I. & Avez 1968 *Ergodic problems of classical mechanics*. New York: W. A. Benjamin.
Baker, G. A. Jr 1958 *Phys. Rev.* **109**, 2198–2206.
Balazs, N. L. & Zipfel, G. G. Jr. 1973 *Annls. Phys.* **77**, 139–156.
Bartlett, M. S., & Moyal, J. E. 1949 *Proc. Camb. Phil. Soc. math. phys. Sci.* **45**, 545–553.
Berry, M. V. 1966 *Proc. phys. Soc.* **89**, 479–490.
Berry, M. V. 1975 *J. Phys.* A **8**, 566–584.
Berry, M. V. 1976 *Adv. Phys.* **25**, 1–26.
Berry, M. V. & Mount, K. E. 1972 *Rep. Prog. Phys.* **85**, 315–397.
Carruthers, P. & Nieto, M. M. 1968 *Rev. mod. Phys.* **40**, 411–440.
Chapman, S., Garrett, B. C. & Miller, W. H. 1976 *J. chem. Phys.* **64**, 502.
Chester, C., Friedman, B. & Ursell, F. 1957 *Proc. Camb. Phil. Soc. math. phys. Sci.* **53**, 599–611.
Cochran, W. 1973 *Phys. Rev.* B **8**, 623–629.
Dirac, P. A. M. 1947 *The principles of quantum mechanics*. Oxford: Clarendon Press.
Duistermaat, J. J. 1974 *Communs pure appl. Math.* **27**, 207–281.
Einstein, A 1917 *Verh. dt. Phys. Ges.* **19**, 82–92.
Elliott, R. J., Krumhansl, J. A. & Leath, P. L. 1974 *Rev. mod. Phys.* **46**, 465–543.
Fanelli, R. & Struzynski, R. E. 1969 *Am. J. Phys.* **37**, 928.
Fisher, M. E. 1974 *Rev. mod. Phys.* **46**, 597–616.
Ford, J. 1975 In *Fundamental problems in statistical mechanics*, (ed. Cohen), vol. III, pp. 215–255. North-Holland.
Groenewold, H. J. 1946 *Physica* **12**, 405–460.
Handy, N. C., Colwell, S. M. & Miller, W. H. 1976 *Faraday Discussion*, no. **62** (to be published).
Heller, E. J. 1976 *J. chem. Phys.* **65**, 1289–1298.
Kravtsov, Yu. A. 1968 *Sov. Phys. Acoust.* **14**, 1–17.
Kolmogoroff, A. N. 1954 *Dokl. Akad. Nauk.* **98**, 527–530.
Langer, R. E. 1937 *Phys. Rev.* **51**, 669–676.
Leaf, B. 1968 *J. Math. Phys.* **9**, 65–72, 769–781.
Leaf, B. 1969 *J. Math. Phys.* **10**, 1971–1979, 1980–1987.
Marcus, R. A. 1971 *J. chem. Phys.* **54**, 3965–3979.
Maslov, V. P. 1972 *Theorie des perturbations et methodes asymptotiques*. Paris: Dunod (Original Russian Publication 1965).
Mehta, C. L. 1964 *J. Math. Phys.* **5**, 677–686.
Moser, J. 1962 *Nachr. Akad. Wiss. Gottingen* **1**, 1–20.
Moser, J. 1973 *Stable and random motions in dynamical systems*. Princeton: University Press.
Moyal, J. E. 1949 *Proc. Camb. Phil Soc. math. phys. Sci.* **45**, 99–124.
Noid, D. W. & Marcus, R. A. 1975 *J. chem. Phys.* **62**, 2119–2124.
Nordholm, K. S. J. & Rice, S. A. 1974 *J. chem. Phys.* **61**, 203–223, 768–779.
Pearcey, T. 1946 *Phil. Mag.* **37**, 311–317.
Percival, I. C. 1973 *J. Phys.* B, **6** L 229–232.
Percival, I. C. 1977 *Adv. chem. Phys.* (to be published).
Percival, I. C. & Pomphrey, N. 1976 *Molec. Phys.* **31**, 917.
Pomphrey, N. 1974 *J. Phys.* B **7**, 1909–1915.
Prokhorov, A. M., Bunkin, F. V., Gochelashvili, K. S. & Shishov, V. I. 1975 *Sov. Phys. Usp.* **17**, 826–847.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Schipper, J. F. 1969 *Phys. Rev.* **184**, 1283–1304.
Siegel, W. 1976 *Acta Phys. Pol.* B **7**, 29–38.
Takabayasi, T. 1954 *Prog. Theor. Phys. Jap.* **11**, 341–373.
Tatarski, V. I. 1961 *Wave propagation in a turbulent medium*. New York: McGraw-Hill.
Thom, R. 1975 *Structural stability and morphogenesis*. Reading, Mass: Benjamin.
Van Vleck, J. H. 1928 *Proc. math. Acad. Sci. U.S.A.* **14**, 178–188.
Voros, A. 1976 *Annls. Inst. Poincaré* **XXIV**, 31–90.
Wigner, E. P. 1932 *Phys. Rev.* **40**, 749–759.
Wilson, K. G. 1975 *Adv. Math.* **16**, 170–186.

Appendix A

This is the discussion of alternatives to Weyl's rule (2.1), as classified by Mehta (1964).

(i) Standard ordering. The operator $\hat{a}$ is written as a sum of products in which all factors $\hat{\boldsymbol{q}}$ precede all factors $\hat{\boldsymbol{p}}$. Then $A(\boldsymbol{q},\boldsymbol{p})$ is obtained by substituting $\boldsymbol{q}$ for $\hat{\boldsymbol{q}}$ and $\boldsymbol{p}$ for $\hat{\boldsymbol{p}}$. The resulting Wigner function is

$$\Psi(\boldsymbol{q},\boldsymbol{p}) = \frac{\langle \boldsymbol{p}|\psi\rangle\langle\psi|\boldsymbol{q}\rangle}{h^{f/2}} e^{i\boldsymbol{p}\cdot\boldsymbol{q}/\hbar}, \tag{A 1}$$

and is unsatisfactory for the following reasons: first, it is not real. Secondly, and more important, its classical limit for an integrable system is large not on the classical torus in phase space but on the union of caustics in $\boldsymbol{q}$ and $\boldsymbol{p}$ (for a one-dimensional oscillator, this is the rectangle in the q, p plane defined by the two turning points in q and two turning points in p).

(ii) Rivier ordering. This is simply (i), symmetrized in q and p. Ψ is simply the real part of (A 1) and is unsatisfactory for the second reason given under (i).

(iii) Normal ordering. The operator $\hat{a}$ is written as a sum of products of operators $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\alpha}}^*$ involving the product β of the mass and frequency of a reference oscillator where,

$$\left.\begin{aligned} \hat{\boldsymbol{\alpha}} &\equiv (\hat{\boldsymbol{q}} + i\beta\hat{\boldsymbol{p}})/(2\beta\hbar)^{f/2} \\ \hat{\boldsymbol{\alpha}}^* &\equiv (\hat{\boldsymbol{q}} - i\beta\hat{\boldsymbol{p}})/(2\beta\hbar)^{f/2} \end{aligned}\right\}, \tag{A 2}$$

with $\hat{\boldsymbol{\alpha}}^*$ always preceding $\hat{\boldsymbol{\alpha}}$. The resulting Wigner function is

$$\Psi(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{(\pi\hbar)^f} \exp\left\{-\frac{\hbar}{4}\left(\beta\nabla_p^2 + \frac{1}{\beta}\nabla_q^2\right)\right\} \int d\boldsymbol{X} \exp\left(-\frac{2i}{\hbar}\boldsymbol{p}\cdot\boldsymbol{X}\right) \langle \boldsymbol{q}+\boldsymbol{X}|\psi\rangle\langle\psi|\boldsymbol{q}-\boldsymbol{X}\rangle. \tag{A 3}$$

This is always real, but unsatisfactory because its projections do not satisfy (2.6) (and indeed need not even be positive), and moreover the value of β is arbitrary.

Appendix B

This is the proof that (4.6) and (4.7) are semi-classically square-integrable to h^{-1} as equation (2.12) requires. Since (4.6) and (4.7) coincide except near $\mathscr{E}$ and the integral of Ψ^2 will not be dominated by the region near $\mathscr{E}$ it suffices to give the proof for (4.6), confining the region of integration over q, p to the interior of $\mathscr{E}$. It is convenient to replace the variables q and p by the tangent directions ψ_1 and ψ_2 (figure 2) of the chord ends 1 and 2. Elementary geometry gives the transformation as

$$\begin{aligned} \int_{q,\,p \text{ within } \mathscr{E}} dq\, dp &= \frac{1}{4}\int_0^{2\pi} d\psi_2 \int_{\psi_2}^{2\pi} d\psi_1 \frac{\partial S}{\partial \psi}(\psi_1)\frac{\partial S}{\partial \psi}(\psi_2)\sin(\psi_1-\psi_2) \\ &= \frac{1}{8}\int_0^{2\pi} d\psi_2 \int_0^{2\pi} d\psi_1 \frac{\partial S}{\partial \psi}(\psi_1)\frac{\partial S}{\partial \psi}(\psi_2)\,|\sin(\psi_2-\psi_1)|. \end{aligned} \tag{B 1}$$

For the integration over Ψ the more convenient from (6.2) is employed rather than (4.6), and the factor $\cos^2$ is replaced by its mean value of $\frac{1}{2}$. The result is

$$\int \mathrm{d}q \int \mathrm{d}p\, \Psi^2(q,p) = \frac{\omega^2}{4\pi^2 h}\int_0^{2\pi} \mathrm{d}\psi_1 \int_0^{2\pi} \mathrm{d}\psi_2 \frac{\partial S}{\partial \psi}(\psi_1)\frac{\partial S}{\partial \psi}(\psi_2)\Big/ \dot{S}_1 \dot{S}_2. \tag{B 2}$$

The integrals separate, and each equals

$$\int_0^{2\pi} \mathrm{d}\psi \frac{\partial S/\partial \psi}{\dot{S}} = \int \mathrm{d}t = \text{period of rotation} = \frac{2\pi}{\omega}, \tag{B 3}$$

so that

$$\int \mathrm{d}q \int \mathrm{d}p\, \Psi^2(q,p) = 1/h, \tag{B 4}$$

Q.E.D.

Appendix C

This is the proof that (4.6) is not correctly normalized to unity whereas (4.7) is. The unsquated Wigner function Ψ oscillates positively and negatively within $\mathscr{E}$, so that semi-classically only the region of phase space near $\mathscr{E}$ will contribute significantly to the normalization integral.

Consider first (4.6), expressed in the form (6.2), and change the variables from q, p to ψ_1, ψ_2 using the first equality (B 1). The area $A(q, p)$ can be approximated by the sagittal area of a quadratic curve when q, p is near $\mathscr{E}$, and becomes

$$A(q,p) \approx \left[\frac{\partial S(\psi_2)}{\partial \psi}\right]^2 \frac{(\psi_1-\psi_2)^3}{12}. \tag{C 1}$$

The integral over (6.2) now becomes

$$\int \mathrm{d}q \int \mathrm{d}p\, \Psi(q,p) = \frac{\omega}{2\pi\sqrt{h}} \mathrm{Re} \int_0^{2\pi} \mathrm{d}\psi_2 \int_{\psi_2}^{2\pi} \mathrm{d}\psi_1 \frac{[\sin(\psi_1-\psi_2)]^{\frac{1}{2}}}{\sqrt{(\dot{S}_1\dot{S}_2)}} \frac{\partial S}{\partial \psi}(\psi_1)\frac{\partial S}{\partial \psi}(\psi_2)$$
$$\times \exp \mathrm{i}\left[\frac{(\psi_1-\psi_2)^3}{12\hbar}\left[\frac{\partial S}{\partial \psi}(\psi_2)\right]^2 - \frac{\pi}{4}\right]. \tag{C 2}$$

Only the region $\psi_1 \sim \psi_2$ contributes significantly (this corresponds to q, p near $\mathscr{E}$), so that, setting $\psi_1 - \psi_2 \equiv u$,

$$\int \mathrm{d}q \int \mathrm{d}p\, \Psi(q,p) = \frac{\omega}{2\pi\sqrt{h}} \mathrm{Re} \int_0^{2\pi} \mathrm{d}\psi_2 \frac{\left[\frac{\partial S}{\partial \psi}(\psi_2)\right]^2}{\dot{S}_2} \int_0^{\infty} \mathrm{d}u\, u^{\frac{1}{2}} \exp \mathrm{i}\left\{\frac{u^3}{12\hbar}\left[\frac{\partial S}{\partial \psi}(\psi_2)\right]^2 - \frac{\pi}{4}\right\}. \tag{C 3}$$

The u integral is elementary, and the ψ_2 integral can then be performed using (B. 3) to give the final result

$$\int \mathrm{d}q \int \mathrm{d}p\, \Psi(q,p) = \sqrt{\tfrac{2}{3}}, \tag{C 4}$$

which differs from unity, Q.E.D.

The normalization of (4.7) is much simpler: since the integral of Ψ is dominated by the region near $\mathscr{E}$, the transitional form (4.8) can be employed. We use the fact that

$$\int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{Ai}(x) \equiv \int_{-\infty}^{\infty} \mathrm{d}x \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}t \exp \mathrm{i}\left(\frac{t^3}{3} + xt\right) = 1. \tag{C 5}$$

Then (4.8) gives

$$\int \mathrm{d}q \int \mathrm{d}p\, \Psi(q,p) = \int_0^{2\pi} \mathrm{d}\Theta \int_0^{\infty} \mathrm{d}I\, \Psi(q,p)$$
$$= \int_{-2I(\mathscr{E})/(\hbar^2 B)^{\frac{1}{3}}}^{\infty} \mathrm{d}x\, \mathrm{Ai}(x) \to 1 \quad \text{when} \quad \hbar \to 0, \tag{C 6}$$

Q.E.D.

Appendix D

This is the proof of (4.13) given (4.11). It involves expanding $A(q, p)$ about $q = 0$. For the relevant points q, p a turning point of q lies on $\mathscr{E}$ between the chord ends 1 and 2 (as with the point q'', p'' on figure 2). Therefore (4.3) is not a convenient expression for $A(q, p)$ and we use instead

$$A(q,p) = \int_{p-P_0}^{p+P_0} \mathrm{d}p' q(p') - 2qP_0, \tag{D 1}$$

where $P_0(q, p)$ is defined by (cf. 4.2)

$$\tfrac{1}{2}[q(p+P_0) + q(p-P_0)] = q. \tag{D 2}$$

Thus $\mathscr{E}$ is now defined not by $p(q)$ but by the inverse function $q(p)$. Then

$$\begin{aligned}\frac{\partial A}{\partial p} &= -2q\frac{\partial P_0}{\partial p} + \left(1+\frac{\partial P_0}{\partial p}\right) q(p+P_0) - \left(1-\frac{\partial P_0}{\partial p}\right) q(p-P_0) \\ &= q(p+P_0) - q(p-P_0) \xrightarrow{p=0} 0,\end{aligned} \tag{D 3}$$

since $$q(p+P_0) = q(p-P_0) = q \quad \text{when} \quad p = 0.$$

The second derivative is

$$\frac{\partial^2 A}{\partial p^2} = \left(1+\frac{\partial P_0}{\partial p}\right)\frac{\partial q}{\partial p}(p+P_0) - \left(1-\frac{\partial P_0}{\partial p}\right)\frac{\partial q}{\partial p}(p-P_0). \tag{D 4}$$

This can be simplified by using the derivative of (D 2), namely

$$\begin{aligned}0 &= \left(1+\frac{\partial P_0}{\partial p}\right)\frac{\partial q}{\partial p}(p+P_0) + \left(1-\frac{\partial P_0}{\partial p}\right)\frac{\partial q}{\partial p}(p-P_0) \\ &\xrightarrow{p\to 0} 2\frac{\partial P_0}{\partial p}\frac{\partial q}{\partial p}(p = p(q)),\end{aligned} \tag{D 5}$$

since $\dfrac{\partial q}{\partial p}(p+P_0) = -\dfrac{\partial q}{\partial p}(p-P_0)$ when $p = 0$. Therefore $\dfrac{\partial P_0}{\partial p} = 0$ when $p = 0$ and

$$\frac{\partial^2 A}{\partial p^2} \xrightarrow{p\to 0} 2\frac{\partial q}{\partial p}(p = p(q)) = 2\Big/\frac{\partial p}{\partial q}(q). \tag{D 6}$$

For small p, then, $$A(q,p) = A(q,0) + p^2\Big/\frac{\partial p}{\partial q}(q) + \ldots. \tag{D 7}$$

With the use of (4.11), $V(p)$ can also be expanded to second order in p, whereupon (4.13) follows trivially, Q.E.D.

Appendix E

This is the proof of the 'projection integral' (4.18). By an elementary transformation, we obtain

$$\int_{-y}^{\infty} \mathrm{d}x \frac{\mathrm{Ai}(x)}{\sqrt{(x+y)}} = \int_{-\infty}^{\infty} \mathrm{d}u\, \mathrm{Ai}\,(u^2-y) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}u\int_{-\infty}^{\infty}\mathrm{d}v \exp\left(\mathrm{i}\left[\frac{v^3}{3} + (u^2-y)\,v\right]\right). \tag{E 1}$$

Now make the following linear change of variables:

$$\begin{aligned}u &\equiv (Y-X)/2^{\frac{2}{3}} \\ v &\equiv (Y+X)/2^{\frac{2}{3}}.\end{aligned} \tag{E 2}$$

This gives
$$\int_{-y}^{\infty} \mathrm{d}x \frac{\mathrm{Ai}(x)}{\sqrt{(x+y)}} = \frac{1}{2\pi 2^{\frac{1}{3}}} \int_{-\infty}^{\infty} \mathrm{d}X \int_{-\infty}^{\infty} \mathrm{d}Y \exp\left(\mathrm{i}\left(\frac{X^3}{3}+\frac{Y^3}{3}-\frac{y}{2^{\frac{2}{3}}}(X+Y)\right)\right)$$
$$= 2^{\frac{2}{3}}\pi \,\mathrm{Ai}^2(y/2^{\frac{2}{3}}), \tag{E 3}$$

Q.E.D. (A similar integral was evaluated by Balazs & Zipfel (1973).)

Appendix F

This is the proof that the locus of midpoints of diameters of any closed convex curve $\mathscr{C}$ must have at least three cusps and in general an odd number. Let $\mathscr{C}$ be given as the function $s(\psi)$ (figure 2). Cusps occur when the curvatures at the ends ψ and $\psi+\pi$ of a diameter are equal, i.e. when

$$ds(\psi_1)/d\psi = \mathrm{d}s(\psi_1+\pi)/\mathrm{d}\psi. \tag{F 1}$$

Now $\mathrm{d}s/\mathrm{d}\psi$ is a periodic function of ψ, restricted by the fact that $\mathscr{C}$ is a closed curve, so that

$$\left.\begin{aligned} \int \mathrm{d}x = \int \mathrm{d}s \cos\psi = \int_0^{2\pi} \mathrm{d}\psi \cos\psi \frac{\mathrm{d}s}{\mathrm{d}\psi} = 0 \\ \int \mathrm{d}y = \int_0^{2\pi} \mathrm{d}\psi \sin\psi \frac{\mathrm{d}s}{\mathrm{d}\psi} = 0 \end{aligned}\right\}. \tag{F 2}$$

This implies that in the Fourier expansion of $\mathrm{d}s/\mathrm{d}\psi$, namely

$$\frac{\mathrm{d}s}{\mathrm{d}\psi} = \sum_{n=0}^{\infty} (A_n \cos n\psi + B_n \sin n\psi), \tag{F 3}$$

the coefficients A_1 and B_1 are zero.

(F 1) can now be written

$$0 = \frac{\mathrm{d}s}{\mathrm{d}\psi}(\psi) - \frac{\mathrm{d}s}{\mathrm{d}\psi}(\psi+\pi) = 2\sum_{n=1}^{\infty} (A_{2n+1}\cos(2n+1)\psi + B_{2n+1}\sin(2n+1)\psi)$$
$$\equiv G(\psi). \tag{F 4}$$

The number of cusps is therefore the number of zeros of $G(\psi)$ in the range 0 to π. Since

$$G(0) = -G(\pi)$$

the number of cusps must be odd. Moreover, the number of cusps must be at least three, since the Fourier series (F 4) begins with terms in $\cos 3\psi$ and $\sin 3\psi$ (no function $G(\psi)$ with a single zero on 0 to π can have vanishing values of both A_1 and B_1) Q.E.D.

Appendix G

This is the discussion of the 'angle' Wigner function (7.10) deep in the shadow region. For definiteness consider the case $A\cos q > 1$; then the deep shadow region is (cf. figure 6)

$$\nu \equiv n-m \ll 0. \tag{G 1}$$

The integral in (7.10) can be transformed for negative ν by studying the following contour C in the complex plane $Z \equiv X+\mathrm{i}Y$; C runs from $\mathrm{i}\infty$ to 0, from 0 to $\frac{1}{2}\pi$ and from $\frac{1}{2}\pi$ to $\frac{1}{2}\pi+\mathrm{i}\infty$. The analogous integral to (7.10) taken over C is zero. On taking the real part, the contribution from $[\mathrm{i}\infty, 0]$ vanishes, and this gives

$$\Psi_m(q,p_n) = \frac{(-1)^\nu}{\pi^2\hbar} \int_0^{\infty} \mathrm{d}Y \,\mathrm{e}^{2\nu Y} \sin\left(\frac{2A}{\hbar}\cos q \cosh Y\right). \tag{G 2}$$

For large negative ν only the region $Y \approx 0$ contributes significantly and integration by parts yields the dominant term as

$$\Psi_m(q, p_n) = \frac{(-1)^{m-n}}{2\pi^2\hbar(m-n)} \sin\left(\frac{2A\cos q}{\hbar}\right) \quad (m \gg n). \tag{G 3}$$

Since this changes sign between successive quantized values p_n these 'fringes' of Ψ are as fine as is possible in a system where q is an angle.

It might seem as if (G 3) predicts that Ψ is of order $\hbar^{-1}$. However, $m-n$ must be large in order for this formula to be valid. When $m = n$ (7.10) can be evaluated exactly (cf. 7.11 and 7.12) and is

$$\Psi_m(q, p_n) = \frac{1}{\hbar} J_0\left(\frac{2A}{\hbar}\cos q\right). \tag{G 4}$$

This is semi-classically of order $h^{-\frac{1}{2}}$ and hence small in comparison with the order $h^{-\frac{2}{3}}$ (cf. §4) attained at the 'Airy' peak on $\mathscr{E}$, except at the infinitely catastrophic inflexions where $\cos q = 0$, discussed at the end of §5.

Appendix H

This is the proof of (8.9) and (8.10). The phase ϕ in (3.6), when expanded to third order in $\boldsymbol{X}$, is

$$\phi = 2(\boldsymbol{p}(\boldsymbol{q}) - \boldsymbol{p})\cdot\boldsymbol{X} + \frac{1}{3}\frac{\partial^2 p_k}{\partial q_i \partial q_j} X_k X_i X_j, \tag{H 1}$$

where (3.7) has been used. $\boldsymbol{X}$ can be expressed in terms of the variable $\boldsymbol{\Theta}$ defined in (8.8) by considering not $\boldsymbol{p}(\boldsymbol{I})$ but the inverse function $\boldsymbol{I}(\boldsymbol{p})$ ($\boldsymbol{q}$ is fixed here). This gives

$$\boldsymbol{X} = \boldsymbol{\Theta}\cdot\nabla_p \boldsymbol{I}. \tag{H 2}$$

The first term in (H 1) simplifies immediately, giving

$$\phi = 2(\boldsymbol{I}_m - \boldsymbol{I}(\boldsymbol{q}, \boldsymbol{p}))\cdot\boldsymbol{\Theta} - \tfrac{1}{3} T_{abc}\,\Theta_a\,\Theta_b\,\Theta_c, \tag{H 3}$$

where

$$T_{abc} = -\frac{\partial^2 p_k}{\partial q_i \partial q_j}\frac{\partial I_a}{\partial p_k}\frac{\partial I_b}{\partial q_i}\frac{\partial I_c}{\partial q_j}. \tag{H 4}$$

This can be put into a form symmetric in $\boldsymbol{q}$ and $\boldsymbol{p}$ by using the functional relation, valid on the torus $\boldsymbol{I}_m$,

$$\boldsymbol{I}(\boldsymbol{q}, \boldsymbol{p}(\boldsymbol{q})) = \boldsymbol{I}_m. \tag{H 5}$$

Successive differentiations with respect to $\boldsymbol{q}$ give

$$\frac{\partial I_q}{\partial q_i}\frac{\partial p_i}{\partial q_j} + \frac{\partial I_a}{\partial q_j} = 0, \tag{H 6}$$

and

$$\frac{\partial^2 I_a}{\partial p_i \partial p_k}\frac{\partial p_k}{\partial q_l}\frac{\partial p_i}{\partial q_j} + \frac{\partial^2 I_a}{\partial p_i \partial q_l}\frac{\partial p_i}{\partial q_j} + \frac{\partial^2 I_a}{\partial q_j \partial q_l} + \frac{\partial^2 I_a}{\partial q_j \partial p_k}\frac{\partial p_k}{\partial q_l} = -\frac{\partial I_a}{\partial p_i}\frac{\partial^2 p_i}{\partial q_j \partial q_l}. \tag{H 7}$$

Then (H 4) gives

$$T_{abc} = \frac{\partial I_b}{\partial p_j}\frac{\partial I_c}{\partial p_l}\left(\frac{\partial^2 I_a}{\partial p_i \partial p_k}\frac{\partial p_k}{\partial q_l}\frac{\partial p_i}{\partial q_j} + \frac{\partial^2 I_a}{\partial p_i \partial q_l}\frac{\partial p_i}{\partial q_j} + \frac{\partial^2 I_a}{\partial q_j \partial q_l} + \frac{\partial^2 I_a}{\partial q_j \partial p_k}\frac{\partial p_k}{\partial q_l}\right). \tag{H 8}$$

Now

$$\partial p_i/\partial q_j = \partial p_j/\partial q_i \tag{H 9}$$

(because of 3.7) and this interchange of indices within the bracket in (H 8) enables (H 6) to be used. Then (8.10) follows immediately, Q.E.D.