

# The inverse spectral problem

Steve Zelditch

## CONTENTS

1. Introduction	401
2. Expansions at $t = 0$	408
3. Dynamics and dynamical inverse problems	410
4. Wave invariants	419
5. Formulae for wave invariants	427
6. Calculation of wave invariants I: Birkhoff normal forms	432
7. Calculation of wave invariants II: Balian-Bloch approach	445
8. Surfaces and domains with integrable dynamics	452
References	460
9. Appendix: Quantum monodromy revisited (by Johannes Sjöstrand and Maciej Zworski)	464
References	466

## 1. Introduction

The inverse spectral problem on a Riemannian manifold  $(M, g)$ , possibly with boundary, is to determine as much as possible of the geometry of  $(M, g)$  from the spectrum of its Laplacian  $\Delta_g$  (with some given boundary conditions). The special inverse problem of Kac is to determine a Euclidean domain  $\Omega \subset \mathbb{R}^n$  up to isometry from the spectrum  $Spec_B(\Omega)$  of its Laplacian  $\Delta_B$  with Dirichlet, Neumann or more general boundary conditions  $B$ . The physical motivation is to identify physical objects from the light or sound they emit, which may be all that is observable of remote objects such as stars or atoms.

The inverse spectral problem is just one among many kinds of inverse problems whose goal is to determine a metric, domain or scatterer from physically relevant invariants. A comparison with other inverse problems shows just how small a set of invariants the spectrum is. For instance, the boundary inverse problem asks to determine the metric  $g$  on a fixed bounded domain  $\Omega \subset M$  of a Riemannian manifold  $(M, g)$  from the spectrum of the Dirichlet Laplacian on  $L^2(\Omega)$ , and from

---

Research partially supported by NSF grant #DMS-0071358.

the Cauchy data  $\frac{\partial\phi_\Delta}{\partial\nu}|_{\partial\Omega}$  of its eigenfunctions, or equivalently from its Dirichlet-to-Neumann operator [BK, KKL, LU]. The inverse scattering problem seeks to determine an obstacle from its scattering amplitude [Ma, Ma2], or from its scattering length spectrum [St, St2], a set of lengths parametrized by  $S^{n-1} \times S^{n-1}$ . By comparison, the inverse spectral problem has to make do with just a discrete ‘unformatted’ set of eigenvalues (or resonance poles in the open case).

It is probably a consequence of this relative poverty of invariants that most of the results in inverse spectral theory over the last two decades (especially since the appearance of Sunada’s article [Su]) are ‘negative results’, showing that one cannot determine metrics or domains by their spectra. The collection of non-isometric isospectral pairs of Riemannian manifolds would require a lengthy survey of its own and for that we refer to the lectures of C. Gordon (cf. [Gor, Gor2]). By comparison, the number of ‘positive results’ showing that one can indeed recover a domain or metric is rather small.

This survey is devoted to the positive results. The emphasis is on relations between the spectrum of the Laplacian and the dynamics of the geodesic flow  $G^t : S^*M \rightarrow S^*M$ . Most of the new material concerns wave trace invariants and their applications to solving concrete inverse spectral problems. The wave group is the quantization of the geodesic flow, and so wave trace methods often ‘reduce’ inverse spectral problems to inverse dynamical problems. Because of their relevance, we have attempted to describe inverse dynamical problems and results.

**1.1. Some basic inverse spectral problems.** Let us introduce some basic terminology. The spectrum of a compact Riemannian manifold defines a map

$$Spec : \mathcal{M} \rightarrow \mathbb{R}_+^{\mathbb{N}}, \quad (g, B) \rightarrow Spec(\Delta_{g,B}) = \{\lambda_0 < \lambda_1^2 \leq \lambda_2^2 \leq \dots\}$$

from some class of metrics  $\mathcal{M}$  on a manifold  $M$  to the spectrum of its Laplacian,

$$\begin{cases} \Delta\phi_j = \lambda_j^2\phi_j, & \langle\phi_i, \phi_j\rangle = \delta_{ij} \\ B\phi_j = 0 & \text{on } \partial M, \end{cases}$$

with boundary conditions  $B : C^\infty(M) \rightarrow C^\infty(\partial M)$  if  $\partial M \neq \emptyset$ . Here,  $\Delta$  denotes the *positive* Laplacian

$$\Delta = -\frac{1}{\sqrt{g}} \sum_{i,j=1}^n \frac{\partial}{\partial x_i} g^{ij} \sqrt{g} \frac{\partial}{\partial x_j}$$

of a Riemannian manifold  $(M, g)$ , where  $g_{ij} = g(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j})$ ,  $[g^{ij}]$  is the inverse matrix to  $[g_{ij}]$  and  $g = \det[g_{ij}]$ . We will only consider Dirichlet  $Bu = u|_{\partial M}$  and Neumann  $Bu = \partial_\nu u|_{\partial M}$ . Eigenvalues are repeated according to their multiplicities.

Two metrics or domains are called isospectral if they have the same spectrum. The main problem in inverse spectral theory is to describe the possible spectra  $\Lambda \subset \mathbb{R}^{\mathbb{N}}$  of Laplacians and, for each possible spectrum, to describe the metrics or domains in the spectral class

$$(1.1) \quad Spec^{-1}(\Lambda).$$

Somewhat simpler is to describe the possible smooth curves in the isospectral class, since it a priori eliminates irregular subsets. An isospectral deformation of a Riemannian manifold (possibly with boundary) is one-parameter family of metrics satisfying  $Spec(M, g_t) = Spec(M, g_0)$  for each  $t$ . Similarly, an isospectral deformation of a domain with a fixed background metric  $g_0$  and boundary conditions

$B$  is a family  $\Omega_t$  with  $Spec_B(\Omega_t) = Spec_B(\Omega)$ . One could also pose the inverse spectral problems for boundary conditions (while holding the other data fixed) as in [GM2, PT].

The inverse spectral and isospectral deformation problems are difficult because the map  $Spec$  is highly nonlinear. The linearization of the problem is to find infinitesimal isospectral deformations, i.e. deformations for which the eigenvalue variations vanish to first order. By first order perturbation theory, the variations of the eigenvalues under a variation of the metric are given by

$$(1.2) \quad \dot{\lambda}_j = \frac{d}{dt}\lambda_j(t)|_{t=0} = \langle \dot{\Delta}\phi_j, \phi_j \rangle,$$

where  $\dot{\Delta}$  is the variation of the Laplacian and where  $\phi_j = \phi_j(0)$  is an orthonormal basis of eigenfunctions which varies smoothly in  $t$  (such a basis exists by the Kato-Rellich theory). As will be recalled below, lengths  $L_\gamma$  of closed geodesics  $\gamma$  are spectral invariants (at least, when there is at most one closed geodesic of each length), so infinitesimal iso-length-spectral deformations are those where

$$(1.3) \quad \dot{L}_\gamma = 0, \quad \forall \gamma.$$

The deformation of the metric is a symmetric 2-tensor  $\dot{g}$ , so the linearized problem is to determine the space of  $\dot{g} \in S^2T^*M$  (modulo tensors arising from diffeomorphisms  $\phi_t^*(g)$ ) for which

$$(1.4) \quad \int_\gamma \dot{g}ds = 0 \quad \forall \gamma \iff \langle \dot{\Delta}\phi_j, \phi_j \rangle = 0, \quad (\forall j).$$

The operator  $\dot{\Delta} = Op(\dot{g})$  is the differential operator with symbol  $\dot{g}$ . The linearized problem is still very difficult because it requires a study of the asymptotic behavior of the expressions (1.4) as the lengths or eigenvalues tend to infinity. This is tantamount to the study of the equidistribution theory of closed geodesics and eigenfunctions.

The basic distinctions in inverse spectral theory are the following. We say that

- a metric or domain is *spectrally determined* (within  $\mathcal{M}$ ) if it is the unique element of  $\mathcal{M}$  with its spectrum;
- it is *locally spectrally determined* if there exists a neighborhood of the metric or domain in  $\mathcal{M}$  on which it is spectrally determined;
- a metric or domain is *spectrally rigid* in  $\mathcal{M}$  if it does not admit an isospectral deformation within the class;
- the inverse spectral problem is solvable in  $\mathcal{M}$  if  $Spec|_{\mathcal{M}}$  is 1 – 1, i.e. if any other metric or domain in  $\mathcal{M}$  with the same spectrum is isometric to it. If not, one has found a counterexample.

There are analogous problems for Laplacians on non-compact Riemannian manifolds, which often have continuous spectra as well as discretely occurring eigenvalues. In place of eigenvalues, one considers the *resonances*  $Res(\Delta)$  of  $\Delta$ , i.e. the poles of the analytic continuation of its resolvent

$$R(z) = (\Delta + z^2)^{-1}.$$

Depending on whether the dimension is odd or even,  $Res(\Delta)$  is a discrete subset of  $\mathbb{C}$  or of the logarithmic plane. The inverse spectral problems above have natural analogues for resonance poles. We refer to Zworski’s expository articles [Zw2, Zw3] for background.

To illustrate the current state of knowledge, we note that even the simplest special metrics are not known to be spectrally determined at the present time (at least to the author's knowledge). It is not known:

- if the standard metric  $g_0$  on  $S^n$  is determined by its spectrum (in dimensions  $\geq 7$ ), i.e. if  $(M, g)$  (or even  $(S^n, g)$ ) is isospectral to  $(S^n, g_0)$  then it is isometric to it. This has been proved in dimensions  $\leq 6$  [T].
- if ellipses in the plane are determined by their Dirichlet spectra, or even if they are spectrally rigid, i.e. if there exist isospectral deformations of ellipses (with Dirichlet boundary conditions).
- if hyperbolic manifolds are determined by their spectra in dimensions  $\geq 3$ . I.e. if  $(M_0, g_0)$  is hyperbolic and  $(M, g)$  is isospectral to it, then is  $(M, g)$  hyperbolic? This is of course true in dimension 2. Is  $(M, g)$  isometric to  $(M_0, g_0)$ ? In dimension 2, this is known to be false for some hyperbolic surfaces.
- if flat metrics are determined by their spectra in the sense that if  $(M, g_0)$  is flat and  $(M, g)$  is isospectral to it, then  $(M, g)$  is flat (it is known that this is true in dimensions  $\leq 6$  or in all dimensions if additionally  $g$  is assumed to lie in a sufficiently small neighborhood of  $g_0$  [Ku3]); it is also classical that there are non-isometric flat tori with the same spectra.

These special cases are tests of the strength of the known methods. Another test is given by the two-dimensional inverse spectral problem. One-dimensional problems are comparatively well understood because the eigenvalue problems are ordinary differential equations and the underlying dynamics consists of just one orbit (an interval)! Two-dimensional problems are already rich in spectral and dynamical complexities, as illustrated by the classical dynamics of twist maps or geodesic flows on Riemannian surfaces. In general, the inverse spectral problem grows rapidly in difficulty with the dimension, and is already quite open for analytic surfaces and domains in two dimensions. This motivates our concentration on two-dimensional problems for much of the survey. The following simple-sounding problems are still apparently beyond the reach of known methods:

- Are convex analytic domains determined by their spectra among other such domains? Are they spectrally rigid?
- Are convex analytic surfaces of revolution determined by their spectra among all metrics on  $S^2$ ? Are they spectrally rigid?

These problems are in some ways analogous to each other in that the unknown is a function of one variable (the boundary or the profile curve), and that is limit of what wave trace invariants at one orbit can hope to recover. Surfaces of revolution are simpler than plane domains since the geodesic flow is integrable, while billiards on plane domains could have any dynamical type. On the other hand, in the domain problem the class  $\mathcal{M}$  only consists of convex plane domains, whereas in the second problem we allow any other metric. If we similarly restricted the class of surfaces of the first problem entirely to analytic convex surfaces of revolution, then the answer is known to be 'yes' [Z2].

The inverse problem for surfaces of revolution might sound reasonably simple since the geodesic flow is completely integrable, and the feeling arises that one should be able to detect this property from the spectrum. This is one of many problems which relate spectral invariants to dynamics of the geodesic flow.

The relations between Laplace spectrum and dynamics have been at the center of at least the positive results in inverse spectral theory in the last thirty years, by comparison with the emphasis on heat invariants in the earlier period. Before going into the technical relations between spectral (in particular, wave) invariants and dynamics, it might be helpful to give some heuristic principles which suggest the relevance of dynamical inverse spectral problems to Laplace inverse spectral problems. The first is the relation of classical to quantum mechanics. Two Laplacians are isospectral if

$$(1.5) \quad \Delta_{g_1} = U \Delta_{g_2} U^*,$$

where  $U : L^2(M_1, g_1) \rightarrow L^2(M_2, g_2)$  is a unitary operator. In the Dirac dictionary of analogies, the classical analogue of this similarity is the symplectic conjugacy

$$|\xi|_{g_1} = \chi^* |\xi|_{g_2} \iff G_{g_1}^t = \chi \circ G_{g_2}^t \circ \chi^{-1}$$

of the corresponding geodesic flows (cf. (3.8)). Here,  $\chi : T^*M_1 \setminus 0 \rightarrow T^*M_2 \setminus 0$  is a homogeneous symplectic diffeomorphism.) This analogy should not be taken too literally, but it is useful in suggesting conjectures. In modern language, the analogue would hold if  $U$  were a unitary Fourier integral operator quantizing  $\chi$ . It would clearly be difficult to prove, even in special cases, that isospectral Laplacians are conjugate by unitary Fourier integral operators, though it was observed independently by Uribe and the author (see [Z6]) and by P. Bérard ([Be, Be2]) that the Sunada counterexamples [Su] have this property (the resulting Fourier integral operators were termed ‘transplantations’ by Bérard). It was also observed in [Z6] that such Fourier integral intertwining operators need not be quantizations of symplectic diffeomorphisms, but could be (and indeed they are, in the Sunada examples) quantizations of multi-valued symplectic correspondences.

But the analogy is suggestive and is fruitful on a local (or more accurately, formal local) level. One of the main results in inverse spectral theory in recent years is the theorem due to V. Guillemin [G, G3] (see also [Z3, Z4]) that isospectrality (with a simple length spectrum assumption) implies the *formal local* symplectic equivalence of the geodesic flows around corresponding pairs of closed geodesics, i.e. it implies the equality of their *Birkhoff normal forms*. This implies local symplectic equivalence around hyperbolic orbits, although not around elliptic orbits (see Problem 3.4). It is perhaps the closest that the above heuristic principle comes to being valid in a general setting.

A further heuristic principal is that much of the dynamics of a flow is encoded in the structure of the flow near closed orbits. To the extent that this is true, local equivalence would be a powerful fact and one should be able to obtain strong information about the metric  $g$  from studying the wave trace expansion around closed geodesics. This raises the possibility that the zeta functions of the flows might determine their dynamical type (see Problem 3.5).

- Can one determine the dynamical type of the geodesic flow from the spectrum of  $\Delta$ , i.e. whether the geodesic flow is integrable, ergodic or of some other type? The Ruelle zeta function of the geodesic flow is generically a Laplace spectral invariant. Can one determine dynamical type from its analytic properties.
- In the above problems, we almost always assume in addition that the length spectrum is multiplicity-free: i.e. that the set of closed geodesics of a fixed length, or more generally the fixed point sets of the geodesic flow, should

contain at most two components interchanged by the time reversal involution  $(x, \xi) \rightarrow (x, -\xi)$ . Are any of the known counterexamples, i.e. non-isometric isospectral pairs, multiplicity free?

The motivation for the second problem is that wave trace methods cannot get off the ground, and in particular the dynamical zeta function need not be a spectral invariant, unless the length spectrum is simple. The issue is that there could exist complicated cancellations among invariants of closed geodesics of the same length.

**1.2. Strategies for solving the inverse spectral problem.** As mentioned above, we are concerned here primarily with the positive results, ones which prove that certain geometric data is determined by the spectra. Roughly speaking, the strategy for obtaining positive results has long consisted of the the following steps:

- (A) Define a lot of spectral invariants;
- (B) Calculate them in terms of geometric or dynamical invariants;
- (C) Try to determine the metric or domain from the invariants.

A crucial limitation arises in step (B), which accounts for the relative paucity of positive results compared to negative results. It is easy to define a complete set of spectral invariants, namely the ‘special values’ of any one of

$$(1.6) \quad \left\{ \begin{array}{l} \text{The heat trace,} \quad Z(t) = \text{Tr} e^{-t\Delta} = \sum_{j=0}^{\infty} e^{-\lambda_j^2 t} \quad (t > 0), \\ \text{The zeta function} \quad \zeta(s) = \text{Tr} \Delta^{-s} = \sum_{j=0}^{\infty} \lambda_j^{-2s} \quad (\Re s > n) \\ \text{The wave trace} \quad S(t) = \text{Tr} e^{it\sqrt{\Delta}} = \sum_{j=0}^{\infty} e^{i\lambda_j t}, \text{ or } S_{ev}(t) = \text{Tr} \cos t\sqrt{\Delta}. \end{array} \right.$$

Of course,  $\zeta(s)$  must be meromorphically continued to  $\mathbb{C}$  and  $S(t)$  is a distribution rather than a function. But the point we are making is that special values are rarely computable in terms of the geometry and are therefore of limited use for positive results on the inverse spectral problem. By comparison, special values can be used to prove negative results by showing that the traces of any of the above operators are the same for two non-isometric  $(M, g)$ . A key step in obtaining positive results is to find computable invariants, and to give efficient algorithms for computing them in terms of the simplest possible geometric invariants. It should be mentioned that there exists spectral invariants such as  $\lambda_1$  and  $\log \det \Delta = -\zeta'(0)$  which are useful although not computable in the above sense. To maintain our theme of wave invariants and dynamics, we will not discuss such invariants.

Oversimplifying a bit, the computable invariants arise from the singularity (or related) asymptotic expansions of the traces defined above or, in another language, from non-commutative residues of functions of the Laplacian (we refer to [G, G2, Z9] for discussion of non-commutative residues). In fact, all of the computable invariants known to the author are wave invariants, i.e. arising from the singularities of the distribution trace or residues of the wave operator  $U(t) = e^{it\sqrt{\Delta}}$  at times  $t$  in the length spectrum of  $(M, g)$  (including  $t = 0$ ).

And it should not be forgotten that the goal of inverse spectral theory is step (C). There now exist a number of rather abstract results showing that various dynamical or quantum mechanical invariants (e.g. Birkhoff normal forms) are spectral invariants [G, GM2, ISZ, Z1, Z3, Z4]. But this only trades one inverse problem

for another, and there is relatively little work on the subsequent inverse problem of determining the domain or metric from these invariants. For instance, it is not hard to see that the classical Birkhoff normal form of the Poincaré map of a bouncing ball orbit *does not* determine all of the Taylor coefficients of the boundary at the endpoints of the orbit, even if the domain has one symmetry. Often step (C) is the deepest, requiring a separate study of inverse dynamical problems.

**1.3. Contents of the survey.** The focus of this survey is on the use of wave trace formula to derive information about the metric and geodesic flow around closed geodesics and to determine metrics or domains from the information. We survey in some detail the relation between wave invariants and Birkhoff normal form invariants on general Riemannian manifolds, initiated by V. Guillemin, and developed by the author and by Iantchenko-Sjöstrand-Zworski [**G**, **G3**, **ISZ**, **SjZ**, **Z3**, **Z4**]. In an appendix to this article J. Sjöstrand and M. Zworski describe in more detail how their general results on quantum monodromy apply to the Laplacian.

But as mentioned above, to succeed with step (C) we need to be able to determine a metric or domain from such invariants. In the end, the crucial problem is to compute wave trace invariants in the simplest and most efficient way, to analyze them in detail, and to reconstruct the domain or metric. This is most feasible in dimension two, so we review in some detail the articles which have succeeded in determining special families of domains or metrics from wave invariants, to wit, bounded plane domains [**CdV**, **P**, **Z1**, **Z2**, **Z5**, **Z7**, **Z10**, **ISZ**, **GM**, **MM**, **S1**] and surfaces of revolution [**Z2**]. One of our aims is to describe a new method for calculating wave invariants from [**Z5**] which so far has achieved better results than the Birkhoff normal form approach.

In addition, we provide a fair amount of background that hopefully puts the special problems in context. There already exist a number of surveys on the inverse spectral problem (e.g. [**Ber**, **Be3**, **C3**, **Me**, **Gor**, **Gor2**]) including our own expository articles [**Z9**, **Z10**], and we have tried to avoid duplication of material which already appears elsewhere. However, to make the survey more self-contained we quote from a number of prior surveys, including our own. We also follow the lecture notes of Melrose [**Me**] in our discussion of the Lifshits (Penrose mushroom) example of two domains with the same wave invariants, and also the (much better known) examples of domains with the same heat invariants.

We also omit a number of topics as being too far from our focus on wave invariants and dynamics. As mentioned above, we do not discuss counterexamples and negative results (cf. [**Gor**, **Gor2**]). We also omit discussion of compactness results of isospectral sets, of which there are many since the (unpublished) work of Melrose and the work of Osgood-Phillips-Sarnak [**OPS**] on isospectral sets of plane domains (for the resonance analogue, see [**HZel2**]). To avoid dissipation of energy, we do not discuss the inverse resonance problem in detail, but only mention some recent results closely related to the inverse spectral problems covered in this survey. We refer to [**Z10**] for further discussion of inverse resonance problems for exterior domains and to [**BJP**, **BP**] for geometric scattering settings.

The author would like to thank V. Baladi, G. Besson, Y. Colin de Verdière, G. Courtois, R. Kuwabara, R. de La Llave, G. Lebeau, M. Rouleux, K. F. Siburg and M. Zworski for informative remarks on the contents of this survey. Of course, errors and omissions are the author's responsibility.

## 2. Expansions at $t = 0$

The most classical spectral invariants are the heat invariants, namely the coefficients of the expansion at  $t = 0$  of the trace of the heat kernel. They are closely related to the coefficients of the trace of the wave group at  $t = 0$ , although it should be noted that the powers  $t^{-\frac{n}{2}+m}$  of the heat kernel expansion at  $t = 0$  are not singular if  $m$  is even and if  $m \geq \frac{n}{2}$  (where  $n = \dim M$ ). Hence, the heat kernel expansion contains more information than the singularity expansion of the wave trace at  $t = 0$ .

**2.1. Boundaryless case.** The earliest work in inverse spectral was based on calculations of heat invariants in terms of curvature invariants, and the recovery of special metrics or domains from these curvature invariants [Ber, Pa, T, T2, Ku2, Ku3].

We recall that the heat trace expansion in dimension  $n$  on a boundaryless manifold has the asymptotic expansion,

$$(2.1) \quad \text{Tr}e^{t\Delta_g} \sim t^{-n/2} \sum_{j=0}^{\infty} a_j t^j.$$

The coefficients  $a_j$  are the heat invariants. We note that when  $n$  is odd, the powers of  $t$  are singular and hence the expansion may be viewed as a singularity expansion in which the terms become more regular. When  $n$  is even, the terms with  $-n/2 + j < 0$  are singular but the rest are smooth and hence are not residual. Rather one may view the expansion as a Taylor expansion at  $t = 0$  of  $t^{n/2}\text{Tr}e^{t\Delta_g}$ . But just like the singular terms, the coefficients are spectral invariants given by integrals of curvature invariants.

The first four heat invariants in the boundaryless case are given by [T, T2]

$$(2.2) \quad \begin{aligned} a_0 &= \text{Vol}(M) = \int d\text{Vol}_M \\ a_1 &= \frac{1}{6} \int S d\text{Vol}_M \\ a_2 &= \frac{1}{360} \int \{2|R|^2 - 2|\text{Ric}|^2 + 5S^2\} d\text{Vol}_M \\ a_3 &= \frac{1}{6!} \int \left\{ -\frac{1}{9} |\nabla R|^2 - \frac{26}{63} |\nabla \text{Ric}|^2 - \frac{143}{63} |\nabla S|^2 \right. \\ &\quad \left. - \frac{8}{21} R_{kl}^{ij} R_{rs}^{kl} R_{ijkl}^{rs} - \frac{8}{63} R^{rs} R_r^{jkl} R_{sjkl} + \frac{2}{3} S |R|^2 \right. \\ &\quad \left. - \frac{20}{63} R^{ik} R_{j\ell} R_{ijkl} - \frac{4}{7} R_j^i R_k^j R_i^k - \frac{2}{3} S |\text{Ric}|^2 + \frac{5}{9} S^3 \right\} d\text{Vol}_M. \end{aligned}$$

Here,  $S$  is the scalar curvature,  $\text{Ric}$  is the Ricci tensor and  $R$  is the Riemann tensor. In general the heat invariants are integrals of curvature polynomials of various weights in the metric. We refer to [Ber] for background.

The heat invariants are complicated and it is difficult to detect meaningful patterns in the curvature polynomials. Nevertheless, they have been successfully used to obtain inverse spectral results.

- Spheres: Tanno [T2] used  $a_0, a_1, a_2, a_3$  to prove that the round metric  $g_0$  on  $S^n$  for  $n \leq 6$  is determined among all Riemannian manifolds by its spectrum, i.e. any isospectral metric  $g$  is necessarily isometric to  $g_0$ . He also used  $a_3$  [T2] to prove that canonical spheres are locally spectrally determined



(hence spectrally rigid) in all dimensions. Patodi proved that round spheres are determined by the spectra  $\text{Spec}^0(M, g)$  and  $\text{Spec}^1(M, g)$  on zero and 1 forms.

- Complex projective space: Let  $(M, g, J)$  be a compact Kähler manifold and let  $(CP^n(H), g_0, H_0)$  be a complex  $n$ -dimensional projective space with the Fubini-Study metric of constant holomorphic sectional curvature  $H$ . Tanno [T2] proves that if the complex dimension  $n \leq 6$  and if

$$\text{Spec}(M, g, J) = \text{Spec}(CP^n(H), g_0, J_0)$$

, then  $(M, g, J)$  is holomorphically isometric to  $(CP^n(H), g_0, J_0)$ . He also proves that  $(CP^n(H), g_0, H_0)$  is locally spectrally determined in all dimensions [T3].

- Flat manifolds: Patodi [Pa] and Tanno [T, T2] used the heat invariants to prove in dimension  $\leq 5$  that if  $(M, g)$  is isospectral to a flat manifold, then it is flat. More precisely, they showed that if  $a_j = 0$  for  $j \geq 1$ , and if  $n \leq 5$  then  $(M, g)$  is flat. The result is sharp, as Patodi (loc. cit.) showed that  $a_j = 0$  for  $j \geq 1$  for the product of a 3-dimensional sphere with a 3-dimensional space of constant negative curvature. In fact, Tanno showed that if  $a_2 = a_3 = 0$ , then  $(M, g)$  is either  $E^6/\Gamma_1$ , where  $\Gamma_1$  is some discontinuous group of translations of the Euclidean space  $E^6$ , or (2)  $[S^3(C) \times H^3(-C)]/\Gamma_2$ , where  $S^3(C)[H^3(-C)]$  is the 3-sphere [hyperbolic 3-space] with constant curvature  $C > 0$  [ $-C < 0$ ] and  $\Gamma_2$  is some discontinuous group of isometries of  $S^3(C) \times H^3(-C)$ . Kuwabara [Ku2, Ku3] used the invariants to prove that flat manifolds are locally spectrally determined, hence spectrally rigid.

2.1.1. *The boundary case.* When  $\partial\Omega \neq \emptyset$ , the heat trace has the form

$$(2.3) \quad \text{Tr}e^{t\Delta_g} \sim t^{-n/2} \sum_{j=0}^{\infty} a_j t^{j/2}.$$

The coefficients have been calculated for a variety of boundary conditions (see [BG] and its references).

The formulae are simplest for plane domains, where the only invariant is the curvature  $\kappa$  of the boundary. Using a nicely adapted calculus of pseudodifferential operators, L. Smith obtained the first five heat kernel coefficients in the case of Dirichlet boundary conditions. They are given by :

$$(2.4) \quad \left\{ \begin{array}{l} a_0 = \text{area of } \Omega, \\ a_1 = -\sqrt{\frac{\pi}{2}} |\partial\Omega| \text{ (the length of the boundary),} \\ a_2 = \frac{1}{3} \int_{\partial\Omega} \kappa ds, \\ a_3 = \frac{\sqrt{\pi}}{64} \int_{\partial\Omega} \kappa^2 ds, \\ a_4 = \frac{4}{315} \int_{\partial\Omega} \kappa^3 ds, \\ a_5 = \frac{37\sqrt{\pi}}{213} \int_{\partial\Omega} \kappa^4 ds - \frac{\sqrt{\pi}}{210} \int_{\partial\Omega} (\kappa')^2 ds, \\ . \end{array} \right.$$

Here,  $ds$  is arclength and  $\kappa'$  is the derivative with respect to arclength. Certain useful patterns in the heat coefficients were used by R. B. Melrose to prove a compactness result (later improved by Osgood-Phillips-Sarnak). We refer to [Me] for the details. In higher dimensions, one still has

$$(2.5) \quad a_0 = C_n \text{Vol}_n(\Omega), \quad a_1 = C'_n \text{Vol}_{n-1}(\partial\Omega).$$

There exist a few inverse inverse spectral results using heat invariants:

- Euclidean balls in all dimensions are spectrally determined among simply connected bounded Euclidean domains by their Dirichlet or Neumann spectra. This follows from (2.5) and from the fact that isoperimetric hypersurfaces in  $\mathbb{R}^n$  are spheres.
- The exterior of the unit ball  $B_3 \subset \mathbb{R}^3$  in dimension 3 is uniquely determined among exterior domains of simply connected compact obstacles by its resonance poles. [HZ].

**2.2. Domains and metrics with the same heat invariants.** It was soon realized that heat invariants are insufficient to determine smooth metrics or domains. This is due to the fact that they are integrals of local invariants of the metrics. Pairs of non-isometric metrics with the same heat invariants can be obtained by putting two isometric bumps. The bumped spheres will not be isometric if the distances between the bumps are different, but the heat invariants will be the same. There are many variations on this well-known example. But heat invariants might be quite useful for analytic metrics and domains, and have also been used in compactness results.

### 3. Dynamics and dynamical inverse problems

We now turn to the more dynamical theory of the wave group in inverse spectral theory. The trace of the wave group expresses spectral invariants in terms of the dynamics of the geodesic flow, and often ‘reduces’ inverse Laplace spectral problems to inverse problems in dynamics. We therefore begin by recalling the relevant dynamical notions and inverse problems.

**3.1. Geodesic flow on boundaryless manifolds.** We denote by

$$(T^*M, \sum_j dx_j \wedge d\xi_j)$$

the cotangent bundle of  $M$  equipped with its natural symplectic form. Given a metric  $g$ , we define the metric Hamiltonian in a standard notation by

$$(3.1) \quad H(x, \xi) = |\xi| := \sqrt{\sum_{ij=1}^{n+1} g^{ij}(x)\xi_i\xi_j}$$

and define the energy surface to be the unit sphere bundle  $S_g^*M = \{(x, \xi) : |\xi|_g = 1\}$ . The spectral theorists geodesic flow is the Hamiltonian flow

$$(3.2) \quad G^t = \exp t\Xi_H : T^*M \setminus \{0\} \rightarrow T^*M \setminus \{0\}, \quad \Xi_H = \text{the Hamiltonian vector field of } H.$$

It is homogeneous of degree 1 with respect to the dilation  $(x, \xi) \rightarrow (x, r\xi)$ ,  $r > 0$ , so nothing is lost by restricting  $G^t$  to  $S_g^*M$ . We also denote its generator by  $\Xi$ , the Hamiltonian and metric being understood.

The periodic orbits of  $G^t$  of period  $T$  of the geodesic flow are the fixed points of  $G^T$  on  $S^*M$ . Equivalently they are the critical points of the length functional on the free loop space of  $M$  of length  $T$ .

**3.2. Billiard flow and billiard map on domains with boundary.** To define the geodesic or billiard flow  $G^t$  on a domain  $\Omega$  with boundary  $\partial\Omega$ , we need to specify what happens when a geodesic intersects the boundary. The definition is dictated by the propagation of singularities theorem for solutions of the wave equation (due to R. B. Melrose and J. Sjöstrand [MS]), and is therefore not purely geometric or dynamical. A billiard trajectory is defined to be the path along which a singularity of a solution of the wave equation propagates. We give a quick, informal review of the flow; for more details, the reader might consult [MS], [PS]. For simplicity we assume that there are no points of infinite order tangency.

We denote by  $S^*\Omega$  the unit tangent vectors to the interior of  $\Omega$  and by  $S_{in}^*\partial\Omega$  the manifold with boundary of inward pointing unit tangent vectors to  $\Omega$  with footpoints in  $\partial\Omega$ . The boundary consists of unit vectors tangent to  $\partial\Omega$ . The billiard flow  $G^t$  is a flow on  $S^*\Omega \cup S_{in}^*\partial\Omega$ , defined as follows: When an interior geodesic of  $\Omega$  intersects the boundary  $\partial\Omega$  transversally, it is reflected by the usual Snell law of equal angles. Such trajectories are called (transversal) reflecting rays. The complications occur when a geodesic intersects the boundary tangentially in  $S^*\partial\Omega$ .

Convex domains are simpler than non-convex domains, since interior rays cannot intersect the boundary tangentially. Dynamical studies of billiards (see e.g. [MF]) often restrict to convex domains. It should be noted that geodesics of  $\partial\Omega$  with the induced metric are important billiard trajectories. They are limits of ‘creeping rays’, i.e. rays with many small links (interior segments) which stay close to  $\partial\Omega$ . In particular, the boundary of a convex plane domain is a closed billiard trajectory. In higher dimensions, closed geodesics on the boundary which are limits of interior creeping rays are closed billiard trajectories. (Are all closed geodesics on the boundary limits in this sense?)

On a non-convex domain, a general billiard trajectory is divided into segments which are either geodesic segments in the interior or geodesic segments of the boundary. Geodesic segments of the boundary only occur where the boundary is convex. Intuitively, the boundary segments are limits of creeping rays along the convex parts of the boundary. Trajectories enter and exit the boundary at inflection points. In particular, the boundary of a non-convex plane domain is not a closed billiard trajectory. If a trajectory intersects the boundary tangentially at a non-inflection point of a non-convex domain, it goes straight past the point of intersection into the interior.

When  $(\Omega, g)$  is non-trapping (i.e. if there is no geodesic ray which remains forever in the interior), the set  $S_{in}^*\partial\Omega$  behaves like a global cross section to the billiard flow. It is then natural to reduce the dimension by defining the *billiard ball map*  $\beta : B^*(\partial\Omega) \rightarrow B^*(\partial\Omega)$ , where  $B^*(\partial\Omega)$  is the ball bundle of the boundary. We first identify  $S_{in}^*\partial\Omega \simeq B^*(\partial\Omega)$  by adding to a tangent (co)vector  $\eta \in B_q^*\partial\Omega$  of length  $< 1$  a multiple  $c\nu_q$  of the inward point unit normal  $\nu_q$  to form a covector in  $S_{\partial\Omega}^{in}\Omega$ . The image  $\beta(q, v)$  is then defined to be the tangential part of the first intersection of  $G^t(q, \eta + c\nu_q)$  with  $\partial\Omega$ . The billiard map is symplectic with respect to the natural symplectic form on  $B^*(\partial\Omega)$ .

An equivalent description of the billiard map of a plane domain is as follows. Let  $q \in \partial\Omega$  and let  $\phi \in (0, \pi)$ . The point  $(q, \phi)$  corresponds to an inward pointing unit vector making an angle  $\phi$  with the tangent line, with  $\phi = 0$  corresponding to a fixed orientation (say counter-clockwise). The billiard map is then  $\beta(q, \phi) = (q', \phi')$  where  $(q', \phi')$  are the parameters of the reflected ray at the first point of intersection with the boundary. The map  $\beta$  is then area preserving with respect to  $\sin \phi ds \wedge d\phi$  (see e.g. [MF]).

**3.3. Closed orbits and their Poincaré maps.** Closed orbits (or periodic orbits)  $\gamma$  of flows are orbits of points  $(x, \xi) \in T^*M$  satisfying  $G^T(x, \xi) = (x, \xi)$  for some  $T \neq 0$  (the period). They project to closed geodesics on the Riemannian manifold or domain.

We recall the definition of the nonlinear Poincaré map  $\mathcal{P}_\gamma$ : in  $S^*M$  one forms a symplectic transversal  $S_\gamma$  to  $\gamma$  at some point  $m_0$ . One then defines the first return map, or nonlinear Poincaré map,

$$\mathcal{P}_\gamma(\zeta) : S_\gamma \rightarrow S_\gamma$$

by setting  $\mathcal{P}_\gamma(\zeta) = G^{T(\zeta)}(\zeta)$ , where  $T(\zeta)$  is the first return time of the trajectory to  $S_\gamma$ . This map is well-defined and symplectic from a small neighborhood of  $\gamma(0) = m_0$  to a larger neighborhood. By definition, the linear Poincaré map is its derivative,  $P_\gamma = d\mathcal{P}_\gamma(m_0)$ .

Closed geodesics are classified by the spectral properties of the symplectic linear map  $P_\gamma$ . Its eigenvalues come in 4-tuples  $\lambda, \bar{\lambda}, \lambda^{-1}, \bar{\lambda}^{-1}$ . A closed geodesic  $\gamma$  is called:

- *non-degenerate* if  $\det(I - P_\gamma) \neq 0$ ;
- *elliptic* if all of its eigenvalues are of modulus one and not equal to  $\pm 1$ , in which case they come in complex conjugate pairs  $e^{i\pm\alpha_j}$ .
- *hyperbolic* if all of its eigenvalues are real, in which case they come in inverse pairs  $\lambda_j \lambda_j^{-1}$
- *loxodromic* or *complex hyperbolic* in the case where the 4-tuple consists of distinct eigenvalues as above.

There are other possibilities (parabolic) in the degenerate case. In the case of Euclidean domains, or more generally domains where there is a unique geodesic between each pair of boundary points, one can specify a billiard trajectory by its successive points of contact  $q_0, q_1, q_2, \dots$  with the boundary. The *n-link periodic reflecting rays* are the trajectories where  $q_n = q_0$  for some  $n > 1$ . The point  $q_0, \dots, q_n$  is then a critical point of the length functional

$$\mathcal{L}(q_0, \dots, q_n) = \sum_{i=0}^{n-1} |q_{i+1} - q_i|$$

on  $(\partial\Omega)^n$ .

Among the periodic orbits, bouncing ball orbits often have special applications in inverse spectral theory. By a *bouncing ball* orbit  $\gamma$  one means a periodic 2-link reflecting ray, i.e.  $q_0 = q_2$ . Convex domains always have at least two bouncing ball orbits, one of which is its diameter. There exist non-convex domains without any bouncing ball orbits. For geometric aspects of bouncing ball orbits, we refer to [Gh].

The projection to  $\Omega$  consists of a segment  $\overline{q_0 q_1}$  which is orthogonal to the boundary at both endpoints, i.e.  $\overline{q_0 q_1}$  is an extremal diameter. The period is of

course twice the length of the segment, which we denote by  $L$ . We write  $q_0 = A, q_1 = B$ . We orient the domain so that the links of  $\gamma$  are vertical and so that the midpoint is at the origin of  $\mathbb{R}^2$ . The top, resp. bottom, of the boundary  $\partial\Omega$  is then the graph of a function  $y = f_+(x)$  resp.  $y = f_-(x)$  over the  $x$ -axis. It is clear that the wave invariants depend only on the Taylor coefficients of  $f_{\pm}$  at  $A, B$ .

We denote by  $R_A$ , resp.  $R_B$  the radius of curvature of the boundary at the endpoints  $A$ , resp.  $B$  of an extremal diameter. The bouncing ball orbit is elliptic if  $L < \min\{R_A, R_B\}$  or if  $\max\{R_A, R_B\} < L < R_A + R_B$ , and is hyperbolic if  $L > R_A + R_B$  or if  $\min\{R_A, R_B\} < L < \max\{R_A, R_B\}$  (see [KT]). If  $AB$  is a local minimum diameter then  $L < R_A + R_B$  while if it is a local maximum diameter then  $L > R_A + R_B$ . So a (non-degenerate) local maximum diameter must be hyperbolic; a local minimum diameter is elliptic if it satisfies the additional inequalities above.

When  $\gamma$  is elliptic, the eigenvalues of  $P_\gamma$  are of the form  $\{e^{\pm i\alpha}\}$  while in the hyperbolic case they are of the form  $\{e^{\pm\lambda}\}$ . The explicit formulae for them are:

$$(3.3) \quad \begin{aligned} \cos \alpha/2 &= \sqrt{\left(1 - \frac{L}{R_A}\right)\left(1 - \frac{L}{R_B}\right)} \quad (\text{elliptic case}), \\ \cosh \lambda/2 &= \sqrt{\left(1 - \frac{L}{R_A}\right)\left(1 - \frac{L}{R_B}\right)} \quad (\text{hyperbolic case}). \end{aligned}$$

We note that  $f''_{\pm}(0) = \frac{1}{R_{\pm}} = \kappa_{\pm}$ , where  $\kappa_{\pm}$  denotes the curvature at ( $A = -, B = +$ ).

**3.4. The length spectrum and the marked length spectrum.** The *length spectrum* of a boundaryless manifold  $(M, g)$  is the discrete set

$$(3.4) \quad Lsp(M, g) = \{L_{\gamma_1} < L_{\gamma_2} < \dots\}$$

of lengths of closed geodesics  $\gamma_j$ . In the boundary case, the length spectrum  $Lsp(\Omega)$  is the set of lengths of closed billiard trajectories in the sense of §3.2 and is no longer discrete, but rather has points of accumulation at lengths of trajectories which have intervals along the boundary. In the case of convex plane domains, e.g., the length spectrum is the union of the lengths of periodic reflecting rays and multiples of  $|\partial\Omega|$ . According to the standard terminology,  $Lsp(M, g)$  is the set of distinct lengths, not including multiplicities, and one refers to the the length spectrum repeated according to multiplicity as the *extended length spectrum*.

In the notation for  $Lsp(M, g)$  we wrote  $L_{\gamma_j}$  as if the closed geodesics of this length were isolated. But in many examples (e.g. spheres or flat tori), the geodesics come in families, and the associated length  $T$  is the common length of closed geodesics in the family. In place of closed geodesics, one has components of the fixed point sets of  $G^T$  at this time. The fixed point sets could be quite messy, so it is also common to assume that they are clean, i.e. that the fixed point sets are manifolds, and that their tangent spaces are fixed point sets of  $dG^T$ . It is equivalent that the length functional is Bott-Morse on the free loop space.

The set of lengths is unformatted in the sense that one does not know which lengths in the list correspond to which closed geodesics. A formatted notion in which lengths are assigned to topologically distinct types of closed geodesics is the *marked length spectrum*  $ML_g$ . On a manifold without boundary, it assigns to each free homotopy class of closed loops the length of the shortest closed geodesic in its class. On a convex plane domain with boundary, topologically distinct closed geodesics correspond to different rotation numbers  $\frac{m}{n} = \frac{\text{winding number}}{\text{number of reflections}}$ . The

marked length spectrum then associates to each rational rotation number the *maximal length* of the closed geodesics having  $n$  reflection points and winding number  $m$ . We refer to §8.4 (based on [S1, S2]) for further discussion and applications to inverse spectral theory.

**3.5. Birkhoff normal forms.** Birkhoff normal forms are approximations to Hamiltonians (or symplectic maps) near equilibria by completely integrable Hamiltonians (or symplectic maps). We now briefly consider the simplest kind of equilibrium occurring for geodesic flows, a closed geodesic  $\gamma$ .

Let us first consider the Birkhoff normal form of the metric Hamiltonian (3.1) near a non-degenerate elliptic closed geodesic  $\gamma$ . To put  $H$  into normal form is to conjugate it (approximately) to a function of locally defined action variables  $(\sigma, I_1, \dots, I_n)$  on the model space  $T^*(S^1 \times \mathbb{R}^n)$ , where  $\sigma$  is the momentum coordinate in  $T^*S^1$  and  $I_j$  are transversal action variables which depend on the type (elliptic, hyperbolic, loxodromic) of  $\gamma$ . More precisely, the normal form algorithm defines a sequence of canonical transformations  $\chi_M$  at  $\gamma$  which conjugate  $H$  to the normal forms

$$(3.5) \quad \chi_M^* H \equiv \sigma + \frac{1}{L} \sum_{j=1}^n \alpha_j I_j + \frac{p_1(I_1, \dots, I_n)}{\sigma} + \dots + \frac{p_M(I_1, \dots, I_n)}{\sigma^M} \quad \text{mod} \quad O_{M+1}^1$$

where  $p_k$  is homogeneous of order  $k+1$  in  $I_1, \dots, I_n$ , and where  $O_{M+1}^1$  is the space of germs of functions homogeneous of degree 1 which vanish to order  $M+1$  along  $\gamma$ . Note that all the terms in (3.5) are homogeneous of degree 1 in  $(\sigma, I_1, \dots, I_n)$ , and that the order of vanishing at  $|I| = 0$  equals one plus the order of decay in  $\sigma$ . The coefficients of the monomials in the  $p_j(I_1, \dots, I_n)$  are known as the classical Birkhoff normal form invariants.

The algorithm for putting a symplectic map (or Hamiltonian) into Birkhoff normal form around a fixed point (or equilibrium point) can be found in many places (see e.g. [AKN]). For geodesic flows, we need Birkhoff normal forms for homogeneous Hamiltonians around periodic orbits. An algorithm for putting a metric Hamiltonian into Birkhoff normal form around a closed geodesic is described in [G] and also in the appendix to [Z3], among other places. To carry out the algorithm to infinite order, one needs to assume that no eigenvalues of  $\gamma$  are roots of unity, which is of course a stronger condition than non-degeneracy.

There are also a Birkhoff normal forms for the geodesic flow and for the Poincaré map. The normal form is simpler for  $\mathcal{P}_\gamma$  since we have eliminated the directions along  $\gamma$  and therefore may express  $\mathcal{P}_\gamma$  just in terms of local action-angle variables  $(I, \phi)$ . The normal form is as follows:

$$(3.6) \quad \mathcal{P}(I, \phi) = (I, \phi + \nabla_I G_M(I)), \quad \text{mod} \quad O_{M+1}^1,$$

where  $G_M(I)$  is a polynomial of degree  $M$  in the  $I$  variables. Thus, to order  $M+1$ , the Poincaré map leaves invariant the level sets of the actions (ellipses, hyperbolas etc. according to the type of  $\gamma$ ) and ‘rotates’ the angle along them.

The well-known question arises whether the full (infinite series) Birkhoff normal form converges and whether the formal symplectic map conjugating the Hamiltonian to its normal form converges. The latter is sometimes called the Birkhoff transformation. Clearly, if the Birkhoff normal form diverges, then so must the Birkhoff transformation. According to a recent article of Perez-Marco [PM], there are no known examples of analytic Hamiltonians having divergent Birkhoff normal forms,

and we refer to that article for further background and history on the problem. If both the normal form and the transformation converge, the Hamiltonian must be integrable in a neighborhood of the orbit. The Birkhoff normal form is then the expression of the Hamiltonian in local action-angle variables. In the generic non-resonant case, it was proved by H. Ito that the Birkhoff transformation converges [I1] (see also [I2] for the resonant case).

For non-integrable systems, the Birkhoff normal form (3.5) and the conjugating map around elliptic orbit  $\gamma$  are only approximations in small neighborhoods of  $\gamma$ , which shrink as  $M$  increases. The hyperbolic case is simpler. In the case of hyperbolic orbits of analytic symplectic maps in two degrees of freedom, it was proved by J. Moser [Mo] that the Birkhoff normal form and transformation do converge. Convergence is more complicated in higher dimension, and we refer to Banyaga-de La Llave-Wayne [BLW], Perez-Marco [PM] and to Rouleux [R] for recent results. In Rouleux, results of [BLW] are used to prove the existence of a local smooth symplectic conjugacy  $\kappa^*H = q(I)$  of the metric Hamiltonian  $H$  near a hyperbolic fixed point (or orbits) to a smooth normal form  $q(I)$ , obtained by Borel summation of the formal Birkhoff normal form of  $H$ . (In the elliptic case, one only has a conjugacy  $\kappa^*H = q(I) + r$  up to a remainder  $r$  which vanishes to infinite order at the orbit.) It follows that two metric Hamiltonians with the same Birkhoff normal form at respective hyperbolic closed geodesics have locally symplectically equivalent geodesic flows near those orbits. (The author thanks M. Rouleux for corroborating this point).

**3.6. Livsic cohomology.** The cohomology problem asks whether a function (cocycle)  $F \in C^\infty(S^*M)$  satisfying  $\int_\gamma F ds = 0$  for every closed geodesic of the metric  $g$  is necessarily a co-boundary,  $F = \Xi(f)$  where  $\Xi$  is the generator of the geodesic flow  $G^t$  and  $f$  is a function with some degree of regularity (see [L1] and other works of de la Llave and others for regularity results).

It is relevant to the inverse length spectral problem for the following reason, first observed by Guillemin-Kazhdan [GK]: Under a deformation  $g_\epsilon$  of a metric  $g = g_0$  preserving the extended  $Lsp(M, g)$  (including multiplicities), one has

$$(3.7) \quad \int_\gamma \dot{g} ds = 0, \quad \forall \gamma.$$

When the cohomology is trivial, one can therefore write  $\dot{g} = \Xi(f)$  for some  $f$  with the given regularity. Then one can study the harmonic analysis on  $S^*M$  to see whether the symmetric 2-tensor  $\dot{g}$  can be expressed as  $\Xi(f)$ . The answer is no for negatively curved surfaces, and that gave the rigidity result of [GK]. Since then their result has been improved (see [CS] for the most general result on rigidity for negatively curved manifolds), but the general strategy retains some potential for other kinds of metrics.

It is known that the Livsic cohomology problem is always solvable for an Anosov flow on a closed manifold. We refer to [L1, Wa] for recent results. One might ask for the analogous result for hyperbolic billiard flows on bounded domains. It is difficult to formulate the analogous regularity result since the dynamics are not smooth. For instance, the billiards are not even defined at the corners of a domain with piecewise smooth boundary and with concave boundary faces. One studies billiards on the domain by puncturing out the measure zero set of orbits which ever run into the corners. The resulting phase space is then not a closed manifold and

there would be complicated issues about regularity of solutions as one approached the punctured set. Perhaps the simplest setting would be that of Sinai billiards, i.e. the exterior of a convex obstacle in a compact manifold (e.g. the exterior of a disc in a two-dimensional torus). The billiards are hyperbolic and there are no glancing orbits or corner orbits to puncture out. But we are not aware of any work on the Livsic equation in this setting.

The Livsic equation can be studied for non-hyperbolic flows, but there do not seem to exist many studies of it other than manifolds all of whose geodesics are closed (in which case a simple Fourier analysis suffices). One study in the near-integrable case is the article [L2] of de la Llave. One does not expect the cohomology to be trivial in general settings, but the results on this equation might have interesting implications for the length spectral deformation problem. As will be explained in §4.3, in certain cases such as integrable systems, the relevant homological equation is not the Livsic equation but an analogue where one integrates over non-degenerate critical manifolds of geodesics.

**3.7. Dynamical inverse problems.** As mentioned above, inverse spectral theory often proceeds by showing that certain dynamical invariants are invariants of the Laplace spectrum. This gives rise to inverse dynamical problems. In this section, we briefly survey some of the problems and results. A recent survey of rigidity and conjugacy problems has been written by C. Croke [C3]. We are grateful to V. Baladi, G. Besson, G. Courtois, C. Gordon, R. de La Llave, and K. F. Siburg for advice on this section.

One of the important problems in dynamics is the following inverse spectral problem in dynamics.

**PROBLEM 3.1.** *For which  $(M, g)$  does the marked length spectrum of  $(M, g)$  determine  $(M, g)$  up to isometry?*

In the positive direction, V. Bangert [B] proved that the marked length spectrum of a flat two-torus determines the flat metric up to isometry. J. P. Otal [O1] and C. Croke [C] (see also [CFF, C3]) independently proved that the marked length spectrum determines surfaces of negative curvature; for more general results see [C3]. U. Hamenstädt [H] proved that a locally symmetric manifold is determined by its marked length spectrum. An example due to F. Bonahon shows that metric structures more general than Riemannian metrics are not always determined by their marked length spectra.

A related problem concerns the conjugacy rigidity of Riemannian manifolds. The geodesic flows  $G_j^t$  of Riemannian manifolds  $(M_j, g_j)$  are called  $C^k$  conjugate if there exists a time-preserving  $C^k$  conjugacy between them, i.e. a  $C^k$  homeomorphism  $\chi : S_{g_1}^* M_1 \rightarrow S_{g_2}^* M_2$  satisfying

$$(3.8) \quad \chi \circ G_1^t \circ \chi^{-1} = G_2^t.$$

If  $\chi$  extends to a homogeneous symplectic diffeomorphism of the cotangent bundles, the flows are called symplectically conjugate. Two billiard maps are conjugate if there exists a symplectic diffeomorphism  $\chi : B^* \Omega_1 \rightarrow B^* \Omega_2$  such that

$$(3.9) \quad \chi \circ \beta_1 \circ \chi^{-1} = \beta_2.$$

**PROBLEM 3.2.** *When does existence of a  $C^k$  conjugacy between two geodesic flows (for a given  $k$ ) imply isometry of the metrics? Does symplectic equivalence of billiard maps of convex domains imply their isometry?*



In some cases, equality of marked length spectra implies  $C^0$  conjugacy of geodesic flows. It was proved independently by J. P. Otal [O1] and by C. Croke [C] that negatively curved surfaces with the same marked length spectrum have  $C^0$ -conjugate geodesic flows, and that such surfaces must be isometric. The corresponding statement in higher dimensions still appears to be open in general. It is stated in [C3] that compact flat manifolds are  $C^\infty$  conjugacy rigid. In [H2], Hamenstädt constructed a time-preserving  $C^0$  conjugacy between negatively curved manifolds with the same marked length spectrum. When this conjugacy is  $C^1$  the work of Besson, Courtois and Gallot [BCG] proves that the manifolds must be isometric. The method of [H] avoids this regularity issue. The conjugacy problem for certain nilmanifolds was studied by C. Gordon, D. Schueth and Y. Mao in [GorM, GMS]. They proved for special classes of 2-step nilmanifolds that  $C^0$  conjugacy of the geodesic flow implies isometry. These manifolds occur in a non-trivial isospectral deformation and have the same marked length spectrum, so these examples also show that equality of marked length spectra does not necessarily imply  $C^0$  conjugacy of the geodesic flows.

There is at least one known example where symplectic conjugacy does not imply isometry, namely for surfaces all of whose geodesics are closed. The geodesic flow of such a Zoll surface was shown by Weinstein to be symplectically equivalent to that of the standard 2-sphere. Aside from related examples, all of the other results known to the author (see [C3]) are rigidity results showing that conjugacy implies isometry. The relations between marked length spectral equivalence, conjugacy of geodesic flows and isometry do not seem to have been studied in other settings.

One could pose local versions of the symplectic conjugacy problem, where there should be much less rigidity. A local version (on the level of germs around closed orbits) of symplectic conjugacy is symplectic conjugacy between Poincaré maps:

$$(3.10) \quad \chi : S_{\gamma_1} \subset S_{g_1}^* M_1 \rightarrow S_{\gamma_2} \subset S_{g_2}^* M_2, \quad \chi \mathcal{P}_{\gamma_1} \chi^{-1} = \mathcal{P}_{\gamma_2}.$$

One could ask:

**PROBLEM 3.3.** *When does local symplectic conjugacy of Poincaré maps at a closed geodesic (or local symplectic conjugacy of geodesic flows) imply local isometry?*

As a special case of the local equivalence problem, suppose that  $(M, g)$  is a hyperbolic manifold, and let  $\gamma$  be a closed geodesic. Let  $g'$  be a second real analytic metric on  $M$ . Suppose that there exists a closed geodesic for  $g'$  for which the Poincaré maps  $\mathcal{P}_\gamma$  are symplectically conjugate. Must  $g'$  be a hyperbolic metric?

Even less rigid is the Birkhoff normal form, since in general there only exists a formal (power series) canonical transformation conjugating the germs of the metrics.

**PROBLEM 3.4.** *To what extent can the germ of a metric  $g$  at  $\gamma$  be determined by its Birkhoff normal form at  $\gamma$ ?*

The sceptic might suspect that when the formal conjugacy does not converge, the Birkhoff normal form gives little beyond a list of numerical invariants, similar to and no simpler than the wave invariants. When it does converge, there might be too little rigidity to determine very much about the metric.

For instance, it was observed by Y. Colin de Verdière [CdV] the Birkhoff normal form of the billiard map around a bouncing ball orbit of a domain with the symmetries of an ellipse determines all of the Taylor coefficients of the domain at

the endpoints of the orbit. Hence a real analytic domain with two symmetries is determined by the Birkhoff normal form of the Poincaré map at a bouncing ball orbit.

If the domain has only one symmetry, then the calculation of [CdV] shows that the Birkhoff normal form at a bouncing ball orbit does not contain enough information to determine the Taylor coefficients at both the top and bottom of the domain. This simple, 2-dimensional example shows how weak an invariant the Birkhoff normal form is.

3.7.1. *Zeta functions of geodesic flows.* We add a final inverse problem which to our knowledge has not been studied before. Let us consider two zeta functions associated to the geodesic flow. The first is the Ruelle zeta function defined by

$$(3.11) \quad \log Z_{(M,g)}(s) = \sum_{\gamma \in \mathcal{P}} \sum_{k=1}^{\infty} \frac{e^{-skL(\gamma)}}{k|\det(I - P_{\gamma}^k)|},$$

Here, we are assuming that the metric is bumpy so that the fixed point sets of the geodesic flow consist of isolated, non-degenerate closed orbits. We denote by  $\mathcal{P} = \{\gamma\}$  the set of primitive periodic orbits of the geodesic flow  $\Phi^t$  and by  $P_{\gamma}$  the linear Poincaré map of  $\gamma$ . As will be seen below, it is a spectral invariant as long as the (extended) length spectrum is simple (multiplicity free) or if the metric has no conjugate points.

The zeta function  $\log Z_{(M,g)}(s)$  arises as the Laplace transform of the so-called flat trace of the unitary Koopman operator

$$W_t : L^2(S^*M) \rightarrow L^2(S^*M), \quad W_t f(x, \xi) := f(G^t(x, \xi)).$$

This operator does not possess a distribution trace in the usual sense, but the integral of its diagonal part  $\int_{S^*M} W(t, \omega, \omega) dV(\omega)$  is well-defined in the sense of pushing forward and pulling back distributions. In this sense, it was observed by V. Guillemin [G2] that

$$(3.12) \quad \text{Tr} W(t) = C_n \text{Vol}(M, g) \delta^{(n)}(t) + \sum_{\gamma \in \mathcal{P}} \sum_{k=1}^{\infty} \frac{\delta(t - L_{\gamma})}{|\det(I - P_{\gamma}^k)|}.$$

The Laplace transform may be viewed as the flat trace of the resolvent of  $\Xi$ , so that

$$(3.13) \quad \log Z_{(M,g)}(s) = \text{Tr}(\Xi - s)^{-1}.$$

In a formal sense, the Ruelle zeta function is a trace of a spectral function of the geodesic flow and it is natural to wonder how much of the spectrum of  $W_t$  can be determined from the traces. In particular, ergodicity, weak mixing and mixing are spectral properties of geodesic flows, i.e. can be read off from the spectrum of  $W_t$ . For instance, ergodicity is equivalent to the statement that the multiplicity of the eigenvalue 1 for  $W_t$  equals one.

But the trace is not a distribution trace on  $L^2$  in the sense that if  $\rho \in C_0^{\infty}$ ,  $\int_{\mathbb{R}} \rho(t) W_t dt$  is usually not a trace class operator on  $L^2(S^*M)$ . Hence, it is far from clear that one can read off spectral properties of  $W_t$  from these traces. At this time of writing, almost nothing seems to be known about zeta functions except in the case of hyperbolic flows, in which case one knows many properties of the spectrum of the geodesic flow, e.g. that it is mixing. We refer to papers of V. Baladi (see e.g. [Ba] for her survey) for background on spectral interpretations of zeta functions and to C. Liverani [BKL] et. al. for some possible future extensions. C. Deninger

and others (see e.g. [DS]) have recently studied the flat trace in non-hyperbolic dynamical settings.

It might be interesting to develop the theory of zeta functions for broader classes of flows. Indeed we are interested in the inverse problem of determining the dynamical properties from the trace.

A related zeta function is defined for  $\Re s$  large by

$$(3.14) \quad \log \zeta_{(M,g)}(s) = \sum_{\gamma \in \mathcal{P}} \sum_{k=1}^{\infty} \frac{e^{-skL(\gamma)}}{k |\det(I - P_{\gamma}^k)|^{1/2}},$$

It is a Laplace spectral invariant of metrics without conjugate points, even if (as with hyperbolic quotients) the extended length spectrum has multiplicity.

**PROBLEM 3.5.** *What are the analytical properties of these zeta functions when the geodesic flow has some given dynamical signature. Can one determine a priori from  $Z_{(M,g)}(s)$  whether the geodesic flow is ergodic, weak mixing or mixing?*

The first step is to see how far the zeta function can be analytically continued, and whether one can determine the multiplicity of the pole at  $s = 0$ . We then ask if the multiplicity has a spectral interpretation as the dimension of the invariant  $L^2$ -functions under the flow.

We observe that the zeta functions  $\zeta_{(M,g)}(s)$  are Laplace spectral invariant for metrics with multiplicity free length spectra as well as for those without conjugate points. Hence this dynamical inverse result would immediately imply that one can determine spectral properties of the geodesic flow from the Laplace spectrum.

### 4. Wave invariants

We now return to the Laplace spectrum and introduce the wave invariants, which are a more discriminating set of spectral invariants than heat invariants. They arise from the trace of the wave group  $U(t) = e^{it\sqrt{\Delta}}$  of  $(M, g)$ . One forms the (distribution) trace

$$(4.1) \quad \text{Tr}U(t) = \sum_{\lambda_j \in \text{Sp}(\sqrt{\Delta})} e^{it\lambda_j}.$$

It is a tempered distribution on  $\mathbb{R}$ . We denote its singular support (the complement of the set where it is a smooth function) by  $\text{Sing Supp Tr}U(t)$ .

The first result on the wave trace is the Poisson relation on a manifold without boundary,

$$(4.2) \quad \text{Sing Supp Tr}U(t) \subset \text{Lsp}(M, g),$$

proved by Y. Colin de Verdière [CdV2, CdV3], Chazarain [Ch2], and Duistermaat-Guillemin [DG] (following non-rigorous work of Balian-Bloch [BB2] and Gutzwiller [Gutz]). The generalization to manifolds with boundary was proved by Anderson-Melrose [AM] and Guillemin-Melrose [GM]. As above, we denote the length of a closed geodesic  $\gamma$  by  $L_{\gamma}$ . For each  $L = L_{\gamma} \in \text{Lsp}(M, g)$  there are at least two closed geodesics of that length, namely  $\gamma$  and  $\gamma^{-1}$  (its time reversal). The singularities due to these lengths are identical so one often considers the even part of  $\text{Tr}U(t)$  i.e.  $\text{Tr}E(t)$  where  $E(t) = \cos(t\sqrt{\Delta})$ .

We emphasize that (4.2) is only known to be a containment relation. As will be seen below, cancellations could take place if a length  $L \in Lsp(M, g)$  is multiple, so that  $TrU(t)$  might be smooth at  $L \in Lsp(M, g)$ .

PROBLEM 4.1. *Is  $Lsp(M, g)$  a spectral invariant? Are there any examples where  $TrU(t)$  is  $C^\infty$  at  $t = L_\gamma \in Lsp(M, g)$ ?*

No example of this seems to be known. Y. Colin de Verdière has pointed out that it is even unknown whether  $TrU(t)$  could be smooth on all of  $\mathbb{R} \setminus \{0\}$ . Moreover,  $Lsp(M, g)$  does not include information about multiplicities of lengths. Sunada-type isospectral pairs always have multiple length spectra, and for many (presumably, generic) examples, the length spectra have different multiplicities. The reason is that the geodesics on the base manifold of a Sunada quadruple can split in different ways in the covers. Other examples of isospectral pairs (both in the sense of eigenvalue and length spectra) with different multiplicities of lengths have been constructed by R. Gornet, C. Gordon and others.

The indeterminacy of multiplicities raises a natural problem. Let us recall that the topological entropy of the geodesic flow of  $(M, g)$  is the exponential growth rate of the length spectrum:

$$(4.3) \quad h_{top} := \liminf_{L \rightarrow \infty} \log \#\{\gamma : L_\gamma \leq L\}.$$

Here, lengths are counted with multiplicity. In the case where geodesics come in families, we count components of the fixed point sets.

PROBLEM 4.2. *Is  $h_{top}$  a spectral invariant?*

The question is only non-trivial when multiplicities in the length spectrum grow as fast as the length spectrum, as occurs for compact hyperbolic manifolds in  $\dim \geq 3$  and for arithmetic hyperbolic quotients in dimension 2. As observed by Besson-Courtois-Gallot [BCG], an affirmative answer would show that that hyperbolic metrics are spectrally determined among other negatively curved metrics, since they are the unique minimizers of  $h_{top}$ . A closely related problem is:

PROBLEM 4.3. *(see [BCG]) Suppose that  $M$  with  $\dim M \geq 3$  possesses a hyperbolic metric, and let  $\mathcal{M}_-$  denote the class of negatively curved metrics on  $M$ . Is the hyperbolic metric  $g$  determined by its spectrum among metrics in  $\mathcal{M}_-$ ? I.e. can there exist another non-isometric metric in this class which is isospectral to the hyperbolic metric?*

4.0.2. *Singular support versus analytic singular support.* While discussing the Poisson relation, we pose the following question concerning the analytic Poisson relation:

PROBLEM 4.4. *Can one tell from  $Spec(\Delta)$  if a metric (or the underlying manifold or domain) is real analytic?*

The idea is to calculate the analytic wave front set of  $TrU(t)$ , i.e.

$$WF_a\left(\sum_{\lambda_j \in Sp(\sqrt{\Delta})} e^{it\lambda_j}\right).$$

The analytic wave front set is the complement of the set where the trace is real analytic. When  $(M, g)$  is a  $C^\omega$  (real analytic) Riemannian manifold, the analytic wave front set  $WF_a TrU(t)$  is also the set  $Lsp(M, g)$ , that is,  $TrU(t)$  is a real

analytic function outside of this discrete set. If  $(M, g)$  is a  $C^\infty$  but not a  $C^\omega$  Riemannian manifold, it is plausible that  $WF_a TrU(t)$  could contain an interval or be all of  $\mathbb{R}$ . Thus simply the discreteness of  $WF_a TrU(t)$  would say that  $(M, g)$  is real analytic.

**4.1. Singularity expansions.** Much more is true than the Poisson relation:  $TrU(t)$  has a singularity expansion at each  $L \in Lsp(M, g)$ :

$$(4.4) \quad TrU(t) \equiv e_0(t) + \sum_{L \in Lsp(M, g)} e_L(t) \text{ mod } C^\infty,$$

where  $e_0, e_L$  are Lagrangian distributions with singularities at just one point, i.e.  $singsupp e_0 = \{0\}$ ,  $singsupp e_L = \{L\}$ . When the length functional on the loop space of  $M$  is a Bott-Morse functional, the terms have complete asymptotic expansions. In the Morse case (i.e. bumpy metrics), the expansions take the form

$$(4.5) \quad e_0(t) = a_{0,-n}(t + i0)^{-n} + a_{0,-n+1}(t + i0)^{-n+1} + \dots$$

$$(4.6) \quad \begin{aligned} e_L(t) = & a_{L,-1}(t - L + i0)^{-1} + a_{L,0} \log(t - (L + i0)) \\ & + a_{L,1}(t - L + i0) \log(t - (L + i0)) + \dots, \end{aligned}$$

where  $\dots$  refers to homogeneous terms of ever higher integral degrees ([DG]). The wave coefficients  $a_{0,k}$  at  $t = 0$  are essentially the same as the singular heat coefficients, hence are given by integrals over  $M$  of  $\int_M P_j(R, \nabla R, \dots) \text{dvol}$  of homogeneous curvature polynomials. The wave invariants for  $t \neq 0$  have the form:

$$(4.7) \quad a_{L,j} = \sum_{\gamma: L_\gamma=L} a_{\gamma,j},$$

where  $a_{\gamma,j}$  involves on the germ of the metric along  $\gamma$ . Here,  $\{\gamma\}$  runs over the set of closed geodesics, and where  $L_\gamma, L_\gamma^\#, m_\gamma$ , resp.  $P_\gamma$  are the length, primitive length, Maslov index and linear Poincaré map of  $\gamma$ . (The primitive length of a closed orbit is the least non-zero period of the orbit, i.e the length once around). For instance, the principal wave invariant at  $t = L$  in the case of a non-degenerate closed geodesic is given by

$$(4.8) \quad a_{L,-1} = \sum_{\gamma: L_\gamma=L} \frac{e^{\frac{i\pi}{4} m_\gamma} L_\gamma^\#}{|\det(I - P_\gamma)|^{\frac{1}{2}}}.$$

The same formula for the leading singularity is valid for periodic reflecting rays of compact smooth Riemannian domains with boundary and with Neumann boundary conditions, while in the Dirichlet case the numerator must be multiplied by  $(-1)^r$  where  $r$  is the number of reflection points (see [GM, PS]).

The wave invariants for  $t \neq 0$  are both less global and more global than the heat invariants. First, they are more global in that they are not integrals of local invariants, but involve the semi-global first return map  $\mathcal{P}_\gamma$ . One could imagine different local geometries producing the same first return map. Second, they are less global because they are determined by the germ of the metric at  $\gamma$  and are unchanged if the metric is changed outside  $\gamma$ .

Thus, associated to any closed geodesic  $\gamma$  of  $(M, g)$  is the sequence  $\{a_{\gamma^r, j}\}$  of wave invariants of  $\gamma$  and of its iterates  $\gamma^r$ . These invariants depend only on the germ of the metric at  $\gamma$ . The principal question of this survey may be stated as follows:

**PROBLEM 4.5.** *How much of the local geometry of the metric  $g$  at  $\gamma$  is contained in the wave invariants  $\{a_{\gamma r, j}\}$ ? Can the germ of the metric  $g$  at  $\gamma$  be determined from the wave invariants? At least, can the symplectic equivalence class of its germ be determined?*

As will be discussed in the next section §6, one of the principal results on this problem is the theorem (due to V. Guillemin [G, G2], with some input and generalizations by the author [Z3, Z4]) that the classical Birkhoff normal form of the metric (or the Poincaré map  $\mathcal{P}_\gamma$ ) at  $\gamma$  is determined by the wave trace invariants. On a global level:

**PROBLEM 4.6.** *How much of the global geometry  $(M, g)$  is contained in the entire set of wave invariants  $\{a_{\gamma r, j}\}$ ?*

As these questions suggest, one may divide the potential use of wave invariants into two classes: (i) those which use all of the closed geodesics, and (ii) those which involve one or a few closed geodesics. Obviously, (i) is more powerful if one can combine information from all of the geodesics, since it adds precisely the global feature which is lost by studying just one closed geodesic. But it seems very difficult in general to combine information coming from different geodesics.

To the author's knowledge, the global problem of combining wave invariants of all closed geodesics has only been led to successful results on isospectral deformations. We will briefly survey the methods and results in the next section §4.3. Otherwise, the main results use only one or two closed geodesics, and this cannot possibly succeed unless the metrics or domains are assumed real analytic. We will survey the results in the analytic case in §7.5.

Another dichotomy in the use of wave invariants is whether one uses only 'principal term' information at each geodesic, i.e. the invariant  $a_{\gamma, -1}$  (4.8), or whether one uses all of the terms. Before the latter is possible, one needs to calculate the lower order terms to some degree. We will discuss the possible calculations in detail in later sections.

**4.2. Inverse spectral results using wave invariants.** We now consider results which use wave invariants to specify a class of metrics or domains which are spectrally determined in the class or which admit no isospectral deformations. We will discuss the proofs of some of the results in later sections to illustrate the methods.

The positive results based on wave invariant analysis are as follows.

- Negatively curved compact manifolds are spectrally rigid [GK, C].
- Simple real analytic surfaces of revolution of 'simple type' (with one critical distance from the axis) are spectrally determined within the class of such surfaces [Z6]. Any other surface which is isospectral to a simple surface of revolution must be  $C^0$ -integrable [Z7, S2]. Smooth surfaces of revolution with a mirror symmetry through the  $x - y$  plane are spectrally determined among metrics of this kind [BH].
- Simply connected analytic plane domains with two symmetry axes (i.e. with the symmetries of an ellipse) and with a bouncing ball orbit of fixed length  $L$  are spectrally determined within this class ([Z1, ISZ, GM2], see also [CdV] for an earlier result proving spectral rigidity of domains in this class). A closely related result is that convex analytic domains with two symmetry axes are spectrally determined within this class [Z1]. The shortest orbit

is necessarily a bouncing ball orbit and of course its length is a spectral invariant [**Gh**].

- Simply connected analytic plane domains with one symmetry, and with a bouncing ball orbit whose orientation of a fixed length  $L$  which is reversed by the symmetry, are spectrally determined within this class [**Z5**]. (This implies the preceding result, but we state it separately since it is a new result based on different methods).
- There is a spectrally determined class of convex plane domains (ellipses?) for which each element is spectrally determined among all convex plane domains [**MM**]. (Other isolated spectrally determined examples have recently been given in [**W**]).
- The *mean minimal action* of a convex billiard table is invariant under isospectral deformations [**S1**] (see §8.4 for further discussion).
- The exterior of a two-component analytic obstacle with two symmetries around a bouncing ball orbit between the components is determined by its resonances (poles of its scattering matrix) among other such exterior domains [**Z10**]. The proof is essentially the same as in the interior case, once some known facts on resonance poles (explained to the author by M. Zworski) are added.

4.2.1. *Domains and metrics with the same wave invariants.* A Penrose mushroom type example due to Michael Lifshits (see [**Me**], [**Rau**] for pictures and background) shows that wave invariants are not sufficient to discriminate between all pairs of smooth billiard tables. Indeed, Lifshitz constructs (many) pairs of smooth domains  $(\Omega_1, \Omega_2)$  which have the same length spectra and the same wave invariants at corresponding pairs of closed billiard orbits  $\gamma_j$  of  $\Omega_j$  ( $j = 1, 2$ ). It follows that the Poincaré maps  $\mathcal{P}_{\gamma_j}$  have the same Birkhoff normal forms as well.

The idea is to exploit the complete integrability of the billiard flow on an ellipse, i.e. the fact that it is foliated by caustics. Caustics are curves in the domain with the property that any billiard trajectory which starts off tangent to the caustic will remain tangent to it, as with the confocal ellipses and hyperbolae of an ellipse. The billiard trajectories fall into two families separated by the bouncing ball orbit between the foci. One family consists of trajectories which remain tangent to confocal ellipses (which degenerate to the segment between the foci). The second family consists of trajectories which are tangent to confocal hyperbolae that pass through the segment between the foci. Now divide the major axis into two parts, the segment between the foci and the other two segments. Remove the segment between the foci and replace it with any simple curve with the foci as endpoints, e.g. a ‘tongue’ below the segment. Next, remove the outer segments of the axis between the foci and replace them with any ‘bumps’ below the segment. The trajectories which start in the outer bumps never intersect the segment between the foci and therefore never go into the tongue. Similarly trajectories which come into the elliptical part from the tongue pass through the segment between the foci and never go into the outer bumps.

It follows that the closed billiard orbits fall into two families: those which never intersect the segment between the foci and therefore bounce back and forth between the bumps; or those which do intersect this segment and never go into the outer bumps.

To obtain non-isometric domains with the same wave invariants it suffices to reverse the relative orientations of the two families, either by reflecting the segment between the foci at the center (i.e. mirror reversing the tongue) or equivalently by reversing the outer bumps. Since the reversal involution is an isometry on each ‘half’ of the domain, it does not change the wave invariants for each half. But there is no ‘interaction’ between the halves (i.e. no closed geodesic intersecting both halves), so the involution preserves all wave invariants. We refer to [Me, Rau] for pictures.

The question may occur whether such a domain and its reversal are isospectral or not. This seems dubious, but we are not aware of a proof of it. As mentioned above, equality of wave invariants implies equality of Birkhoff normal form invariants at corresponding pairs. Are the billiard maps of each pair symplectically conjugate? Again dubious, but again we don’t know a proof.

**4.3. Isospectral deformations and spectral rigidity.** One of the first uses of wave invariants and dynamics was the proof by Guillemin-Kazhdan [GK] that negatively curved surfaces are spectrally rigid. In the negatively curved case, the Maslov indices are always zero and no cancellation takes place in the wave trace formula as one sums over closed geodesics of the same length. Hence,  $Lsp(M, g)$  is a spectral invariant of negatively curved manifolds. An isospectral deformation therefore preserves the length spectrum.

If  $\gamma$  is an isolated, non-degenerate closed geodesic of  $g$ , then for any deformation  $g_t$  of  $g$ ,  $\gamma$  deforms smoothly as a closed geodesic  $\gamma_t$  of  $g_t$  and one may define its variation

$$(4.9) \quad \dot{L}_\gamma = \left. \frac{d}{dt} \right|_{t=0} L_{\gamma_t}.$$

It is not hard to compute that

$$(4.10) \quad \dot{L}_\gamma = \int_\gamma \dot{g} ds,$$

where  $\dot{g}$  is viewed as a quadratic function on  $TM$  and  $\gamma$  is viewed as the curve  $(\gamma(s), \gamma'(s))$  in  $TM$ .

It follows that whenever the closed geodesics are non-degenerate and of multiplicity one in the length spectrum, we have

$$(4.11) \quad \int_\gamma \dot{g} ds = 0, \quad \forall \gamma.$$

Thus, the Livsic cohomology problem enters. Since  $Lsp(M, g)$  is generically simple, this equation holds for generic isospectral deformations. We note that only the principal level of the wave invariants was used here.

In the case of negatively curved surfaces, the geodesic flow is Anosov and it is known that the cohomology is trivial, i.e. that  $\dot{g} = \Xi(f)$  for some smooth  $f$  (the regularity problem is a separate issue, but for the sake of brevity we will not consider it). The next step is to study harmonic analysis on the unit sphere bundle  $S^*M$  to determine if there actually can exist  $f$  when  $\dot{g}$  is a quadratic form. For surfaces of negative curvature, it was proved in [GK] that there cannot exist a smooth solution, and hence there exist no isospectral deformations of negatively curved surfaces. After a series of partial results by Guillemin-Kazhdan, Min-Oo and others, the result was extended to higher dimensions by Croke-Sharafutdinov [CS]. In [SU], Sharafutdinov-Uhlmann prove infinitesimal spectral rigidity for the



more general class of closed two-dimensional manifold without focal points whose geodesic flow is of Anosov type.

Isospectral deformations also lead to the inverse marked length spectral problem and hence (in some cases) to the  $C^0$  conjugacy rigidity problem for the geodesic flows. An isospectral deformation preserves lengths of closed geodesics in the generic case where the multiplicities all equal one (or more generally where no length is cancelled in the wave trace formula), and therefore marks the length spectrum, i.e. gives a correspondence between closed geodesics and lengths, by the length spectrum of the initial metric. More precisely, as one deforms the metric or domain, each one-parameter family of closed geodesics  $\gamma_\epsilon$  stays in a fixed free homotopy class of the fundamental group of  $M$ . Therefore, an isospectral deformation gives rise to a one-parameter family of metrics or domains with the same marked length spectrum. As discussed above, the marked length spectrum determines the metric for surfaces of negative curvature (Croke, Otal), which gave a different proof of the result of [GK].

There do not seem to exist many studies of the Livsic cohomology equation for non-hyperbolic systems. It was proved by Kuwabara using heat invariants that flat tori are spectrally rigid. It might be interesting to review the result in terms of the Livsic equation, and possibly generalizing them to a broader class of completely integrable systems. As will be discussed in §8.1, closed geodesics of integrable systems (such as surfaces of revolution or Liouville tori) come in families on invariant tori  $\mathcal{T}$ , and the analogue of (4.11) would say that

$$(4.12) \quad \int_{\mathcal{T}} \dot{g} dx = 0$$

for each invariant torus of the geodesic flow. This is an analogue of the Livsic cohomology equation when geodesics come in non-degenerate critical manifolds. In the case of integrable systems, it could be studied using Fourier analysis relative to the foliation by invariant tori of a completely integrable system. However, it is weaker than the Livsic equation. At least for some Riemannian surfaces, isospectral deformations of metrics with integrable geodesic flows must be through metrics whose geodesic flows are  $C^0$  integrable (see §8.1), and the study of the Livsic type equation (4.12) might combine to form a more powerful tool. The simplest case to consider is whether simple surfaces of revolution on  $S^2$  are spectrally rigid (see §8.1 for background).

To give an extreme example of an open isospectral deformation problem, consider the case of Zoll manifolds. It is simple to see that isospectral deformations of Zoll manifolds must be Zoll [Z8], but it is an open problem whether any non-trivial isospectral deformations exist even for  $M = S^2$ . One of the main problems is that all principal symbol level spectral invariants of Zoll manifolds are the same, so one has to dig further into the wave trace expansions to find obstructions to isospectral deformability.

4.3.1. *Isospectral deformations of domains.* A few words on the analogues for Euclidean domains. To the author's knowledge, there are no spectral rigidity results even for Euclidean plane domains except in the case of a disc. It is not even known whether ellipses are spectrally rigid. The length spectrum of a bounded domain is a spectral invariant for both Dirichlet and Neumann boundary conditions, so an isospectral deformation must preserve the length spectrum and even the marked

length spectrum. We will discuss some results of K.-F. Siburg on the marked length spectrum problem in §8.4.

Since the length spectrum is generically an isospectral invariant, it is natural to consider the analogue of the Guillemin-Kazhdan rigidity result for domains with hyperbolic billiards. The billiard flow is hyperbolic, so presumably the Livsic equation can be solved (although the author is not aware of a specific reference). It is not clear how the harmonic analysis would change from the case of negatively curved surfaces.

There is another approach which reduces the wave group to the boundary. Suppose that  $\Omega_t$  is a smooth one parameter family of smooth compact plane domains such that the Dirichlet (resp. Neumann) spectrum  $Spec(\Delta_D^{\Omega(0)})$  (resp.  $Spec(\Delta_N^{\Omega(0)})$ ) is constant in  $t$ . Since the area and length of the boundary are spectral invariants, both must be fixed under the deformation. With no loss of generality, we assume that the variation is generated by a normal vector field  $\rho\nu$ , where  $\nu$  is the outward unit normal to  $\Omega(0)$  and where  $\rho \in C^\infty(\partial\Omega(0))$ .

The variations of the eigenvalues are given by Hadamard's variational formulae [O1]:

$$(4.13) \quad \dot{\lambda}_j(0) = \begin{cases} \int_{\partial\Omega(0)} \rho |\partial_\nu \phi_j(0)|_{\partial\Omega}^2 dA, & \text{Dirichlet} \\ \int_{\partial\Omega} \rho |\phi_j(0)|_{\partial\Omega(0)}^2 dA, & \text{Neumann.} \end{cases}$$

Hence, the infinitesimal deformation condition is that the right hand sides are zero for all  $j$ . To normalize the problem, we assume with no loss of generality that the deformation is volume preserving, which implies that

$$(4.14) \quad \int_{\partial\Omega} \rho dA = 0.$$

Any such  $\rho$  defines a volume preserving deformation of  $\Omega$ .

Thus, the infinitesimal deformation is orthogonal to all boundary traces of eigenfunctions:

$$(4.15) \quad \dot{\lambda}_j(0) = 0 \forall j, \iff \begin{cases} \int_{\partial\Omega(0)} \rho |\partial_\nu \phi_j(0)|_{\partial\Omega}^2 dA = 0, & \text{Dirichlet} \\ \int_{\partial\Omega} \rho |\phi_j(0)|_{\partial\Omega(0)}^2 dA = 0, & \text{Neumann.} \end{cases}$$

We may rewrite these conditions in terms of the boundary values of the wave kernel:

$$(4.16) \quad E^b(t, x, x) := \begin{cases} \partial_{\nu_x} \partial_{\nu_y} U(t, x, x)|_{x \in \partial\Omega}, & \text{Dirichlet} \\ U(t, x, x)|_{x \in \partial\Omega}, & \text{Neumann} \end{cases}$$

as saying that

$$(4.17) \quad \int_{\partial\Omega} E^b(t, x, x) \rho(x) dA(x) = 0, \forall t.$$

Using the calculations in [O1], one can obtain expansions for (4.17) at  $t = 0$  similar to a heat kernel expansion in terms of integrals of  $\rho$  against polynomials in the (extrinsic) curvature invariants of the boundary. The singularities at  $t \neq 0$  in turn give integrals over closed orbits of the billiard map. In the case of an ellipse, one can directly study the boundary traces of the eigenfunctions, and it appears that  $\int_\Gamma \rho ds = 0$  for every caustic  $\Gamma$  for the billiard map. It would be interesting to explore this further.

### 5. Formulae for wave invariants

When studying the inverse spectral problem for individual pairs of metrics rather than isospectral deformations, it is difficult in general to bring in the global dynamics of the geodesic flow. Nothing better is known than to concentrate on a single closed geodesic. So to make progress on the inverse spectral problem, it is crucial to have useful formulae for the wave invariants  $a_{\gamma^r, j}$  associated to a closed geodesic  $\gamma$  and its iterates  $\gamma^r$ . The purpose of this section is to survey the known methods of calculation and the formulae which they bring.

There are several potential approaches:

- (i) Construct a microlocal parametrix for  $e^{it\sqrt{\Delta}}$  at  $\gamma$  and apply a stationary phase method to calculate the wave invariants (cf. [Z9]).
- (ii) Construct a Birkhoff normal form for  $e^{it\sqrt{\Delta}}$  at  $\gamma$  and relate normal form invariants to wave invariants (cf. [G, Z1, Z3, Z4, Z9]).
- (iii) Construct a Birkhoff normal form for the monodromy operator  $\mathcal{M}$  on the microlocal solution space  $\ker_{m_0}(\Delta - \lambda^2)$  at a fixed initial point  $m_0$  of  $\gamma$  and relate it to the wave invariants (cf. [SjZ, ISZ]).
- (iv) For bounded domains, apply the Balian-Bloch or Calderon projector methods (cf. [Z5]§7).

We now display the formulae and give a brief discussion of the methods. In the following sections, we describe the methods in much more detail.

**5.1. The parametrix method.** In boundaryless manifolds one can construct a Hadamard parametrix for  $\cos t\sqrt{\Delta}(x, x)$  for small times. This operator solves the initial value problem

$$(5.1) \quad \begin{cases} (\frac{\partial}{\partial t} - \Delta)u = 0 \\ u|_{t=0} = f \end{cases} \quad \frac{\partial}{\partial t}u|_{t=0} = 0 \quad ,$$

and has the form of the real part of the oscillatory integral

$$(5.2) \quad \int_0^\infty e^{i\theta(r^2-t^2)} \sum_{j=0}^\infty U_j(x, y) \theta_{reg}^{\frac{n-1}{2}-j} d\theta \quad \text{mod } C^\infty$$

where the Hadamard-Riesz coefficients  $U_j$  are determined inductively by the transport equations

$$(5.3) \quad \begin{aligned} \frac{\Theta'}{2\Theta}U_0 + \frac{\partial U_0}{\partial r} &= 0 \\ 4ir(x, y) \left\{ \left( \frac{k+1}{r(x, y)} + \frac{\Theta'}{2\Theta} \right) U_{k+1} + \frac{\partial U_{k+1}}{\partial r} \right\} &= \Delta_y W U_k. \end{aligned}$$

The solutions are given by:

$$(5.4) \quad \begin{aligned} U_0(x, y) &= \Theta^{-1/2}(x, y) \\ U_{j+1}(x, y) &= \Theta^{-1/2}(x, y) \int_0^1 s^j \Theta(x, x_s)^{1/2} \Delta_2 U_j(x, x_s) ds \end{aligned}$$

where  $x_s$  is the geodesic from  $x$  to  $y$  parametrized proportionately to arc-length, where  $\Theta(x, y)$  denotes the volume density in normal coordinates centered at  $x$ , and where  $\Delta_2$  operates in the second variable.

The simplest case is where the metric is without conjugate points, in which case the parametrix is global in time on the universal cover. It may then be projected

to  $M$  in the usual way by summing over the deck transformation group. One then takes the trace and computes the coefficients by a stationary phase method.

This was first done by Donnelly [D] for negatively curved surfaces and then for all manifolds without conjugate points by the author in [Z9]. To get an impression of the complexity of the result, here is the formula for the subprincipal wave invariant (the coefficient of the logarithmic singularity):

$$\begin{aligned}
 a_{\gamma o} &= \frac{1}{|\det(I - P_\gamma)|^{\frac{1}{2}}} \{ C_{n,o,o,o}^o \int_\gamma \Theta_\gamma^{\frac{1}{2}}(\sigma) \{ \Theta_\gamma^{-\frac{1}{2}} \text{Hess}(f_\gamma)_\sigma^{-1}(J_\gamma)((0,0), (\sigma,0)) \\
 &+ \text{Hess}(f_\gamma)_\sigma^{-1}(\Theta_\gamma^{-\frac{1}{2}}((0,0), (\sigma,0)) d\sigma \\
 &+ C_{n,o,o,1}^o \int_0^{L_\gamma} \int_0^{L_\gamma} \Theta_\gamma^{\frac{1}{2}}(s\sigma_1) \{ \text{Hess}(f_\gamma)_\sigma^{-1} \}^2 (g_\gamma \Theta_\gamma^{-\frac{1}{2}} J_{[\gamma]})|_{\nu=0} ds \\
 &+ C_{n,o,o,1}^o \int_0^{L_\gamma} \int_0^{L_\gamma} \Theta_\gamma^{\frac{1}{2}}(s\sigma_1) \{ \text{Hess}(f_\gamma)_\sigma^{-1} \}^3 (f_\gamma) \Theta_\gamma^{-\frac{1}{2}}|_{\nu=0} ds \\
 &+ C_{n,o,1}^o \int_0^{L_\gamma} \int_0^{L_\gamma} \Theta_\gamma^{\frac{1}{2}}(\sigma_1) (\Delta_2 \Theta^{-\frac{1}{2}})((\sigma_1,0), (\sigma_2,0)) \cdot d\sigma_1 d\sigma_2 \} .
 \end{aligned}$$

Here,  $\Theta_\gamma(x) = \Theta(x, \gamma x)$ ,  $J_\gamma$  is a similar quantity using geodesic polar coordinates,  $r = r(x, y)$  denote the distance function of the universal cover,  $\gamma \in \Gamma$ , the fundamental group,  $f_\gamma(x) = r(x, \gamma x)^2$  denotes the displacement function. See [Z9] for details. The higher wave invariants are of course exponentially more complicated.

In the case of bounded domains, the parametrix approach is even messier. A microlocal parametrix for  $\cos t\sqrt{\Delta}$  near a transversal reflecting ray was constructed by J. Chazarain [Ch] (see also [GM, PS] for more details ). In principle it could be used in the same way as a microlocal parametrix in the boundaryless case, but in the boundary case there are almost always conjugate points and additionally the transport equations and phase are more complicated. Applying the method of stationary phase to all orders to such a parametrix descends into a jungle of formulae. To our knowledge, no concrete inverse spectral results have been proved using the parametrix method.

**5.2. Wave invariants and quantum Birkhoff normal forms.** Evidently, some guiding principle is needed to civilize the wilderness of formulae. Two such principles have emerged, at least in the case of bounded domains. The first is the method of quantum Birkhoff normal forms. As for classical Birkhoff normal forms, one conjugates the Laplacian to a model normal form on a model space. The wave invariants are then expressed in terms of the coefficients of the normal form. The method is described in detail in the next section. For the moment, we are interested in the formula it gives for the wave invariants.

To state the result we will need some notation. The wave invariants at a closed geodesic  $\gamma$  are invariants of the germ of the metric at  $\gamma$ . For simplicity, we assume that the geodesic is isolated and non-degenerate ( $\det(I - P_\gamma) \neq 0$ ). We assume  $\dim M = n + 1$  and introduce Fermi normal coordinates  $(s, y)$  along  $\gamma$ . We denote the corresponding metric coefficients by  $g_{ij}$ , and refer to the coordinate vector fields  $\frac{\partial}{\partial s}, \frac{\partial}{\partial y_j}$  ( $j = 1, \dots, n$ ) and their real linear combinations as Fermi normal vector fields along  $\gamma$ . We denote the Riemannian connection, resp. curvature tensor, by  $\nabla$  resp.  $R$  and refer to contractions of tensor products of the  $\nabla^m R$ 's with the Fermi normal vector fields as *Fermi curvature polynomials*. Such polynomials will

be called *invariant* if they are invariant under the action of  $O(n)$  in the normal spaces.

We consider the Jacobi equation  $Y'' + R(Y, \gamma')\gamma' = 0$  along  $\gamma$ . The space of complex Jacobi fields along  $\gamma$  is denoted  $\mathcal{J}_\gamma$ . The linear Poincaré map is the monodromy map  $Y(t) \rightarrow Y(t + L_\gamma)$  on this space. We refer to its eigenvectors  $Y_j, \bar{Y}_j$  as Jacobi eigenvectors. We denote by  $y_{jk}$  their components relative to the Fermi normal vector fields.

We then define *Fermi–Jacobi polynomials* to be invariant contractions of tensor products of  $\nabla^m R$ 's against  $\frac{\partial}{\partial s}$  and against the Jacobi eigenvectors, with coefficients given by invariant polynomials in the components  $y_{jk}$ . We will also use this term repeated integrals over  $\gamma$  of such polynomials. Finally, FJ polynomials whose coefficients are given by polynomials in the Floquet invariants

$$\beta_j = (1 - e^{i\alpha_j})^{-1}$$

are *Fermi–Jacobi–Floquet* polynomials. Here, we assume for simplicity that the closed geodesic is elliptic and that the eigenvalues of its Poincaré map are  $\{e^{i\alpha_j}\}$ . (The Floquet invariants in the general non-degenerate case are analogous; see [Z4] for details).

We define the ‘weight’ of the data going into a Fermi–Jacobi–Floquet polynomial in terms of its scaling behavior under  $g \rightarrow \epsilon^2 g$ . Thus, the variables  $g_{ij}, D_{s,y}^m g_{ij}$  (with  $m := (m_1, \dots, m_{n+1})$ ),  $L := L_\gamma, \alpha_j, y_{ij}, \dot{y}_{ij}$  have the following weights:

$$\text{wgt}(D_{s,y}^m g_{ij}) = -|m|,$$

$\text{wgt}(L) = 1, \text{wgt}(\alpha_j) = 0, \text{wgt}(y_{ij}) = \frac{1}{2}, \text{wgt}(\dot{y}_{ij}) = -\frac{1}{2}$ . A polynomial in this data is homogeneous of weight  $s$  if all its monomials have weight  $s$  under this scaling.

The general result is [Z4]:

**THEOREM 5.1.** *Let  $\gamma$  be a strongly non-degenerate closed geodesic. Then  $a_{\gamma k} = \int_\gamma I_{\gamma;k}(s; g) ds$  where:*

- (i)  $I_{\gamma;k}(s; g)$  is a homogeneous Fermi–Jacobi–Floquet polynomial of weight  $-k - 1$  in the data  $\{y_{ij}, \dot{y}_{ij}, D_{s,y}^m g\}$  with  $m = (m_1, \dots, m_{n+1})$  satisfying  $|m| \leq 2k + 4$  ;
- (ii) The degree of  $I_{\gamma;k}$  in the Jacobi field components is at most  $6k + 6$ ;
- (iii) At most  $2k + 1$  indefinite integrations over  $\gamma$  occur in  $I_{\gamma;k}$ ;
- (iv) The degree of  $I_{\gamma;k}$  in the Floquet invariants  $\beta_j$  is at most  $k + 2$ .

The formula is simplest in dimension 2, where there is only one Floquet invariant  $\beta$ . We use the notation  $\tau$  for the scalar curvature,  $\tau_\nu$  for its unit normal derivative along  $\gamma$ ,  $\tau_{\nu\nu}$  for the Hessian  $\text{Hess}(\tau)(\nu, \nu)$ . We denote by  $Y$  the unique normalized Jacobi eigenvector along  $\gamma$  and by  $\dot{Y}$  its time-derivative.

The subprincipal wave invariant  $a_{\gamma 0}$  is then given by:

$$(5.5) \quad a_{\gamma 0} = \frac{a_{\gamma,-1}}{L^\#} [B_{\gamma 0;4}(2\beta^2 - \beta - \frac{3}{4}) + B_{\gamma 0;0}]$$

where:

- (a)  $a_{\gamma,-1}$  is the principal wave invariant (4.8);
- (b)  $L^\#$  is the primitive length of  $\gamma$ ;  $\sigma$  is its Morse index;  $P_\gamma$  is its Poincaré map;
- (c)  $B_{\gamma 0;j}$  has the form:

$$B_{\gamma 0;j} = \frac{1}{L^\#} \int_0^{L^\#} [a |\dot{Y}|^4 + b_1 \tau |\dot{Y} \cdot Y|^2 + b_2 \tau \text{Re}(\bar{Y} \dot{Y})^2 + c \tau^2 |Y|^4 + d \tau_{\nu\nu} |Y|^4 + e \delta_{j0} \tau] ds$$

$$\begin{aligned}
 & + \frac{1}{L^\#} \sum_{0 \leq m, n \leq 3; m+n=3} C_{1;mn} \frac{\sin((n-m)\alpha)}{|(1 - e^{i(m-n)\alpha})|^2} \left| \int_0^{L^\#} \tau_\nu(s) \bar{Y}^m \cdot Y^n(s) ds \right|^2 \\
 & + \frac{1}{L^\#} \sum_{0 \leq m, n \leq 3; m+n=3} C_{2;mn} \operatorname{Im} \left\{ \int_0^{L^\#} \tau_\nu(s) \bar{Y}^m \cdot Y^n(s) \left[ \int_0^s \tau_\nu(t) \bar{Y}^n \cdot Y^m(t) dt \right] ds \right\}
 \end{aligned}$$

for various universal coefficients.

The formula is very complicated, and it is only the second term in the expansion! That is why one hopes to directly use the information that these coefficients determine the Birkhoff normal form.

In the case of surfaces of revolution, it is possible to simplify the formulae to the point where one can determine the metric along the invariant geodesics. The point is to recover the Taylor expansion of the curvature and its normal derivatives along invariant geodesics. In [Z2], a variant of this method was used to show that ‘simple’ surfaces analytic of revolution (those with just one invariant geodesic) are determined by their wave invariants, hence are spectrally determined among such surfaces. The variant was to use the existence of a global Birkhoff normal form for the metric and Laplacian, which greatly simplified the calculations.

The wave invariants can also be calculated by the method of Birkhoff normal forms in the boundary case, but the method becomes appreciably more difficult. We therefore use a different approach which is inspired by the Balian-Bloch approach to the Poisson formula.

**5.3. Balian-Bloch approach in the boundary case.** We now consider wave invariants associated to periodic reflecting rays of a bounded plane domain  $\Omega$ . For simplicity and because they are often useful in applications, we consider wave invariants at a 2-link periodic reflecting ray  $\gamma$  of length  $rL_\gamma$  (a ‘bouncing ball orbit’). We refer to the §3.3 for the basic definitions.

We introduce some further notation: we write  $y = f_{\sigma(j)}(x_j)$  where  $\sigma(2j+1) = -, \sigma(2j) = +$ . The length functional in Cartesian coordinates for a given assignment  $\sigma$  of signs is given by

$$(5.6) \quad \mathcal{L}_\sigma(x_1, \dots, x_{2r}) = \sum_{j=1}^{2r-1} \sqrt{(x_{j+1} - x_j)^2 + (f_{\sigma(j+1)}(x_{j+1}) - f_{\sigma(j)}(x_j))^2}.$$

We will need formulae for the entries of its Hessian  $H_{2r}$  in Cartesian coordinates at the critical point corresponding to the  $r$ th repetition of the bouncing ball orbit with a given orientation. Thus,  $x_j = 0$  for all  $j$ . We will assume, with no essential loss of generality, that  $q(x_{odd}) = A, q(x_{ev}) = B$ . We put:  $a = 1 - Lf''_+(0), b = 1 - Lf''_-(0)$ .

The Hessian is given in either angular or Cartesian coordinates by:

$$(5.7) \quad H_{2r} = \frac{1}{L} \left\{ \begin{array}{ccccc} a & 1 & 0 & \dots & 1 \\ 1 & b & 1 & \dots & 0 \\ 0 & 1 & a & 1 & 0 \\ 0 & 0 & 1 & b & 1 \dots \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & b \end{array} \right\},$$

where  $a = 2(1 - \frac{L}{R_A})$ ,  $b = 2(1 - \frac{L}{R_B})$ . A well known formula relates the determinant of this Hessian to that of the Poincaré map:

$$(5.8) \quad \det(I - P_{\gamma^r}) = L^{2r} \det H_{2r}; \quad (\gamma \text{ 2-link.})$$

We denote the matrix elements of the inverse Hessian  $H_{2r}^{-1}$  at the bouncing ball orbit by  $h^{ij}$ . To be precise, the bouncing ball orbit has two possible orientations, one (which we denote by + which starts at the top graph  $y = f_+$  and proceeds to the bottom  $y = f_-$  and the other - which in the reverse order. We then have two (closely related) length functions and Hessians. We denote the matrix elements of their inverses by  $h_{\pm}^{ij}$ . One has  $h_{\pm}^{pq} = h_{\pm}^{p-1, q-1}$ .

We now state the formulae for the wave invariants.

**THEOREM 5.2.** [Z5] *Let  $\gamma$  be a 2-link reflecting ray of length  $L_{\gamma}$ , and let  $\gamma^r$ , resp.  $\gamma^{-r}$  be the  $r$ th iterate of  $\gamma$ , resp.  $\gamma^{-1}$ . Then there exist polynomials*

$$p_{2,r,j}(\xi_1, \dots, \xi_{2j}; \eta_1, \dots, \eta_{2j}),$$

which are homogeneous of degree  $-j$  under the dilation  $f \rightarrow \lambda f$ , which are invariant under the substitutions  $\xi_j \iff \eta_j$  and under  $f(x) \rightarrow f(-x)$ , and which have degree  $j + 1$  in the Floquet data  $e^{i\alpha r}$ , such that

$$a_{\gamma^r, j} = p_{2,r,j}(f_-^{(2)}(0), f_-^{(3)}(0), \dots, f_-^{(2j+2)}(0); f_+^{(2)}(0), f_+^{(3)}(0), \dots, f_+^{(2j+2)}(0)).$$

The leading order term in derivatives of  $f_{\pm}$  has the form

$$\begin{aligned} & r\{(h_+^{22})^j \sum_{q=1; q \equiv 0}^{2r} h_+^{2q} + (h_+^{22})^{j-2} \sum_{q; q \equiv 0}^{2r} (h_+^{2q})^3\} f_+^{(2j-1)}(0) f_+^{(3)}(0) \\ & + r\{(h_+^{11})^j \sum_{q; q \equiv 1}^{2r} h_-^{1q} + (h_+^{11})^{j-2} \sum_{q; q \equiv 1}^{2r} (h_+^{1q})^3\} f_-^{(2j-1)}(0) f_-^{(3)}(0) \\ & - r\{(h_+^{22})^{j-1} h_+^{11} \sum_{q=1; q \equiv 1}^{2r} h_{\pm}^{2q} + (h_+^{22})^{j-2} \sum_{q=1; q \equiv 1}^{2r} (h_+^{2q})^3\} f_+^{(2j-1)}(0) f_-^{(3)}(0) \\ & - r\{(h_+^{11})^{j-1} h_+^{22} \sum_{q=1; q \equiv 0}^{2r} h_{\pm}^{1q} + (h_+^{11})^{j-2} \sum_{q=1; q \equiv 0}^{2r} (h_+^{1q})^3\} f_-^{(2j-1)}(0) f_+^{(3)}(0) \\ & + R_{2r}(j^{2j-2} f_+(0), j^{2j-2} f_-(0)), \end{aligned}$$

where the remainder  $R_{2r}(j^{2j-2} f_+(0), j^{2j-2} f_-(0))$  is a polynomial in the designated jet of  $f_{\pm}$ .

In the case where the domain has a symmetry interchanging the top and bottom, so that  $f_{\pm}$  are mirror images of the graph of  $y = f(x)$ . The length functions for  $\gamma^r$  and  $\gamma^{-r}$  (and hence their Hessians) are equal, so we may drop the subscripts  $\pm$ . The formula simplifies as follows:

**COROLLARY 5.3.** *Suppose that  $\gamma$  (as above) is invariant under an isometric involution  $\sigma$ . Then, modulo the error term  $R_{2r}(j^{2j-2}f(0))$ , we have:*

$$la_{\gamma^r, j-1} = r\{2(h^{11})^j f^{(2j)}(0) + \{2(h^{11})^j \frac{1}{2 - 2 \cos \alpha/2} + (h^{11})^{j-2} \sum_{q=1}^{2r} (h^{1q})^3\} f^{(3)}(0) f^{(2j-1)}(0)\}.$$

Here, we sum over repeated indices.

The inverse spectral problem is then reduced to analyzing Hessian coefficients. It is not evident to the author how this reduction of the inverse spectral problem would be visible using Birkhoff normal forms.

5.3.1. *Melrose-Marvizi invariants.* In [MM], Melrose-Marvizi introduce further spectral invariants of a convex domain which could be interpreted as quantum normal form invariants, namely the normal form of  $\Delta$  around the closed geodesic  $\partial\Omega$ . The calculation in [MM] involves Melrose’s normal form for glancing hypersurfaces and may be viewed as a normal form construction, but one might also follow Lazutkin’s construction of whispering gallery quasi-modes to put  $\Delta$  into normal form as a function of a so-called Airy operator. This would presumably give a new way to calculate the invariants. J. Toth and the author have partial results in this direction (unpublished), and from discussions with G. Popov it appears that they are known to others.

The first two Melrose-Marvizi integrals have the form [MM]:

$$I_1 = -2 \int_{\partial\Omega} \kappa^{2/3} ds, \quad 1080I_2 = \int_{\partial\Omega} (9\kappa^{4/3} + 8\kappa^{-8/3}\dot{\kappa}^2) ds.$$

**6. Calculation of wave invariants I: Birkhoff normal forms**

We now outline the construction of the Birkhoff normal form of  $\Delta$  in both the boundaryless and boundary cases, and indicate how it is used to calculate the wave invariants. Our goals are to explain how the calculations in Theorem 5.1 are done.

Let  $\gamma$  be a non-degenerate closed geodesic on an  $n$ -dimensional Riemannian manifold, and at first let us assume it to be elliptic. During the 70’s, various authors (Babic, Lazutkin, Ralston, Guillemin-Weinstein) constructed a series of quasi-modes and quasi-eigenvalues associated to  $\gamma$ . Roughly speaking, the results showed that for each transversal quantum number  $q \in \mathbb{Z}^{n-1}$ , there was an approximate eigenvalue of the form

$$(6.1) \quad \lambda_{kq} \equiv r_{kq} + \frac{p_1(q)}{r_{kq}} + \frac{p_2(q)}{r_{kq}^2} + \dots$$

where

$$r_{kq} = \frac{1}{L} (2\pi k + \sum_{j=1}^n (q_j + \frac{1}{2}) \alpha_j)$$

where the coefficients are polynomials of specified degrees and parities. We refer to [BB] for a clear exposition and for details.



A natural question is whether the wave invariants  $a_{\gamma_j}$  of (4.6) can be determined from the quasi-eigenvalues (6.1), i.e. from the polynomials  $p_j(q)$ , which could be computed explicitly from the quasimode constructions described in [BB]. This question was difficult to answer or even to formulate precisely until the article of V. Guillemin [G] on quantum Birkhoff normal forms for the Laplacian around elliptic closed orbits. This article does not refer to quasi-modes or quasi-eigenvalues but it effectively proves that the wave invariants may indeed be expressed in terms of the polynomials  $p_j(q)$  and conversely that these polynomials are spectral invariants. A somewhat different proof of this result, which constructed quantum Birkhoff normal forms by adapting the algorithm in [BB] for constructing quasi-modes, was given by the author for elliptic orbits in [Z2], and for general orbits in [Z3].

The results of [G] are stated in the terms of quantum Birkhoff normal forms. A number of expositions now exist which explain this notion in detail (see for instance [Z9] in addition to the original articles), so we will only briefly review the notion. To put  $\Delta$  into normal form, is first to conjugate it into a distinguished maximal abelian algebra  $\mathcal{A}$  of pseudodifferential operators on a model space, the cylinder  $S^1_L \times \mathbb{R}^n$ , where  $S^1_L$  is the circle of length  $L$ . The algebra is generated by the tangential operator  $D_s := \frac{\partial}{i\partial s}$  on  $S^1_L$  together with the transverse action operators. The nature of these action operators depends on  $\gamma$ . When  $\gamma$  is elliptic, the action operators are harmonic oscillators

$$I_j = I_j(y, D_y) := \frac{1}{2}(D_{y_j}^2 + y_j^2),$$

while in the real hyperbolic case they have the form

$$I_j = y_j D_{y_j} + D_{y_j} y_j.$$

When  $\gamma$  is non-degenerate, they involve some mixture of these operators (and also complex hyperbolic actions) according to the spectral decomposition of  $P_\gamma$ . For notational simplicity we restrict to the elliptic case and put

$$H_\alpha := \frac{1}{2} \sum_{k=1}^n \alpha_k I_k$$

where  $e^{\pm i\alpha_k}$  are the eigenvalues of the Poincare map  $P_\gamma$ .

To put  $\Delta$  into normal form is to conjugate it to the model space and algebra as a function of  $D_s$  and the action operators. The conjugation is only defined in a neighborhood of  $\gamma$  in  $T^*M - 0$ , i.e. one constructs a microlocally elliptic Fourier Integral operator  $W$  from the conic neighborhood of  $\mathbb{R}^+ \gamma$  in  $T^*N_\gamma - 0$  to a conic neighborhood of  $T^*_+ S^1_L$  in  $T^*(S^1_L \times \mathbb{R}^n)$  such that:

(6.2)

$$W \sqrt{\Delta_\psi} W^{-1} \equiv \mathcal{D} + \frac{\tilde{p}_1(\hat{I}_1, \dots, \hat{I}_n)}{LD} + \frac{\tilde{p}_2(\hat{I}_1, \dots, \hat{I}_n)}{(LD)^2} + \dots + \frac{\tilde{p}_{k+1}(\hat{I}_1, \dots, \hat{I}_n)}{(LD)^{k+1}} + \dots$$

where the numerators  $p_j(\hat{I}_1, \dots, \hat{I}_n), \tilde{p}_j(\hat{I}_1, \dots, \hat{I}_n)$  are polynomials of degree  $j+1$  in the variables  $\hat{I}_1, \dots, \hat{I}_n$ , where  $W^{-1}$  denotes a microlocal inverse to  $W$ . Here,  $\mathcal{D} = D_s + \frac{1}{L} H_\alpha$ . The  $k$ th remainder term lies in the space  $\oplus_{j=0}^{k+2} O_{2(k+2-j)} \Psi^{1-j}$ , where  $\Psi^s$  denotes the pseudo-differential operators on the model space of order  $s$  and where  $O_{2(k+2-j)}$  denotes the operators whose symbols vanish to the order  $2(k+2-j)$  along  $\gamma$ . We observe that (6.2) is an operator version of (6.1).

The inverse result of [G] (see also [Z3, Z4]) is:

**Theorem** *Let  $\gamma$  be a non-degenerate closed geodesic. Then the quantum Birkhoff normal form around  $\gamma$  is a spectral invariant; in particular the classical Birkhoff normal form is a spectral invariant.*

In other words, one can determine the polynomials  $p_j(\hat{I}_1, \dots, \hat{I}_n)$  from the wave trace invariants of  $\Delta$  at  $\gamma$ . When  $\gamma$  is a non-elliptic (e.g. hyperbolic) geodesics, the action operators have continuous spectra and there are no approximate eigenvalue expansions as in (6.1). Thus the quantum Birkhoff normal form approach is conceptually clearer.

We give a brief review of how wave invariants and normal form invariants for the Laplacian are calculated by the algorithm in [Z2, Z3, Z4, Z5] in the case of an elliptic closed geodesic. The algorithm is similar in the general non-degenerate case. We then consider the algorithm for calculating the normal form of the monodromy operator in [ISZ].

**6.1. Manifolds without boundary.** Our algorithm for conjugating  $\Delta$  to normal form along a closed geodesic  $\gamma$  is inspired by the constructions due to Lazutkin [L] and Babich-Buldirev [BB] for constructing quasimodes associated to elliptic closed orbits. The same conjugation method works in the general non-degenerate case.

Since we have recently written an exposition of the method [Z9], we only include only a few formal aspects of the calculations, in the hope they provide sufficient explanation of how the wave invariants are calculated. In particular, we wish to emphasize:

- The *semiclassical normal form*, and how it arises from a semi-classical scaling of the Laplacian along  $\gamma$ . Our method differs from the others [G, ISZ] by working entirely on the quantum level and inductively on the Taylor expansion of the metric around  $\gamma$ . The homogeneous normal form (6.2) is obtained from the semiclassical normal form by (roughly speaking) replacing the large parameter by  $|D_s|$  along  $\gamma$ .
- The homological equations and the obstructions to their solvability along a closed orbit. The obstructions determine the normal form.

**6.2. Quasi-mode heuristics.** We wish to conjugate the Laplacian to a normal form on a model space, namely the normal bundle  $N_\gamma$  of  $\gamma$ , or equivalently, the cylinder  $S_L^1 \times \mathbb{R}^n$ , where as above  $S_L^1 = \mathbb{R}/L\mathbb{Z}$ . On the phase space level, the model is  $T^*(S_L^1 \times \mathbb{R}^n)$  or more precisely the cone  $|\eta| \leq \epsilon\sigma, |y| \leq \epsilon$  in the natural symplectic coordinates  $(s, \sigma, y, \eta)$  corresponding to Fermi normal coordinates along  $\gamma$ .

In the construction of quasi-modes associated to  $\gamma$ , one uses the WKB ansatz

$$\Phi_{kq}(s, \sqrt{r_{kq}}y) = e^{ir_{kq}s} U_{kq}(s, \sqrt{r_{kq}}y, r_{kq}^{-1}),$$

where  $r_{kq} = \frac{2\pi}{L}(k + (q + \frac{1}{2})\alpha)$  and where

$$U_{kq}(s, \sqrt{r_{kq}}y, r_{kq}^{-1}) \sim \sum_{j=0}^{\infty} r_{kq}^{-\frac{j}{2}} U_q^{\frac{j}{2}}(s, \sqrt{r_{kq}}y, r_{kq}^{-1}).$$

One then solves asymptotically the eigenvalue

$$\Delta_y e^{i r_{kq} s} U_{kq}(s, \sqrt{r_{kq}} y, r_{kq}^{-1}) \sim \lambda_{kq} e^{i r_{kq} s} U_{kq}(s, \sqrt{r_{kq}} y, r_{kq}^{-1}),$$

satisfying the periodicity condition of being well defined on the cylinder. The intertwining operator  $W_\gamma$  to normal form may be thought of as the operator taking the model eigenfunctions to the quasimodes,

$$W_\gamma \phi_{kq}(s, y) = \Phi_{kq}(s, \sqrt{r_{kq}} y).$$

We now change our point of view and concentrate on the construction of the intertwining operator.

**6.3. Semiclassically scaled Laplacian.** In view of the form of the quasi-eigenvalue problem,

$$\Delta_u e^{\frac{i}{hL} s} U(s, h^{-\frac{1}{2}} u, h) = \lambda(h) e^{\frac{i}{hL} s} U(s, h^{-\frac{1}{2}} u, h),$$

it is natural to rescale the Laplacian before conjugating it to normal form. In fact, this rescaling is all we retain from the quasi-mode construction. We therefore introduce the unitary operators  $T_h$  and  $M_h$  on  $\mathcal{H}_T$  or equivalently on the 1/2-density version  $L_T^2(\mathbb{R}^1 \times \mathbb{R}^n, \Omega_{1/2})$  given by

$$T_h(f(s, u) | ds |^{1/2} | du |^{1/2}) := h^{-n/2} f(s, h^{-\frac{1}{2}} u) | ds |^{1/2} | du |^{1/2}$$

$$M_h(f(s, u) | ds |^{1/2} | du |^{1/2}) := e^{\frac{i}{hL} s} f(s, y) | ds |^{1/2} | du |^{1/2}.$$

**Definition** *The rescaling of an operator  $A_u = a(s, D_s, u, D_u)$  of the adapted model is the operator*

$$A_h := T_h^* M_h^* A T_h M_h.$$

We now rescale the Laplacian in Fermi normal coordinates. It is convenient to first conjugate to the unitarily equivalent 1/2-density Laplacian

$$\Delta_{1/2} := J^{1/2} \Delta J^{-1/2},$$

which can be written in the form:

$$\begin{aligned} -\Delta_{1/2} &= J^{-1/2} \partial_s g^{oo} J \partial_s J^{-1/2} + \sum_{ij=1}^n J^{-1/2} \partial_{u_i} g^{ij} J \partial_{u_j} J^{-1/2} \\ &\equiv g^{oo} \partial_s^2 + \Gamma^o \partial_s + \sum_{ij=1}^n g^{ij} \partial_{u_i} \partial_{u_j} + \sum_{i=1}^n \Gamma^i \partial_{u_i} + \sigma_o. \end{aligned}$$

Here,  $\partial_x := \frac{\partial}{\partial x}$ , and  $J = J(s, u) = \sqrt{g}$  is the volume density in these coordinates.

We then have:

$$-M_h^* \Delta M_h = -(hL)^{-2} g^{oo} + 2i(hL)^{-1} g^{oo} \partial_s + i(hL)^{-1} \Gamma^o + \Delta$$

Conjugation with  $T_h$  then gives

$$\begin{aligned} -\Delta_h &= -(hL)^{-2} g_{[h]}^{oo} + 2i(hL)^{-1} g_{[h]}^{oo} \partial_s + i(hL)^{-1} \Gamma_{[h]}^o \\ &\quad + h^{-1} (\sum_{ij=1}^n g_{[h]}^{ij} \partial_{u_i} \partial_{u_j}) + h^{-\frac{1}{2}} (\sum_{i=1}^n \Gamma_{[h]}^i \partial_{u_i}) + (\sigma)_{[h]}, \end{aligned}$$

the subscript  $[h]$  indicating to dilate the coefficients of the operator in the form,  $f_h(s, u) := f(s, h^{\frac{1}{2}} u)$ .

Expanding the coefficients in Taylor series at  $h = 0$ , we obtain the asymptotic expansion

$$\Delta_h \sim \sum_{m=0}^{\infty} h^{(-2+m/2)} \mathcal{L}_{2-m/2}$$

where  $\mathcal{L}_2 = L^{-2}$ ,  $\mathcal{L}_{3/2} = 0$  and where

$$\mathcal{L}_1 = 2L^{-1} \left[ i \frac{\partial}{\partial s} + \frac{1}{2} \left\{ \sum_{j=1}^n \partial_{u_j}^2 - \sum_{ij=1}^n K_{ij}(s) u_i u_j \right\} \right].$$

**6.4. Conjugation of scaled Laplacian to semi-classical normal form.**

We now conjugate the semi-classically scaled Laplacian to normal form. The conjugating operators are  $s$ -dependent operators acting on the transverse space to  $\gamma$  (they are sometimes called tangential semi-classical Fourier integral operators).

The first step is to conjugate the principal term into quadratic normal form on the transverse space. Since it is quadratic, there exists a metaplectic conjugation  $\mu(s)$  depending on  $s$  on the transverse space which conjugates it to the operator  $\mathcal{D}$  of (6.2). That is, for each  $s \in [0, L]$  there exists an operator  $\mu(s)$  in the metaplectic representation of the metaplectic group  $ML(n, \mathbb{R}) \rightarrow Sp(n, \mathbb{R})$  which acts on the transverse  $\mathbb{R}^n$  of the model space. For background on metaplectic operators and details on the conjugation we refer to [Z3, Z4]. This conjugates the other operators  $\mathcal{L}_j$  to new operators  $\mathcal{D}_j$ .

Once the principal term is in normal form, we continue by conjugating the lower order terms to functions of the actions by using perturbation theory, i.e. formal series in  $h$ . Thus we wish to put the formal series  $\Delta_h$  into a *semi-classical normal form* by an infinite sequence of conjugations, each one putting one new order in  $h$  into normal form. We carry out the procedure to two orders to illustrate the main points.

The first step is to construct  $Q_{\frac{1}{2}}(s, x, D_x)$  such that

$$e^{-ih^{\frac{1}{2}}Q_{\frac{1}{2}}} \mathcal{D}_h e^{ih^{\frac{1}{2}}Q_{\frac{1}{2}}} |_{\circ} = [-h^{-2} + 2h^{-1}\mathcal{D} + \mathcal{D}_{\circ}^{\frac{1}{2}} + \dots] |_{\circ}$$

where

(6.3)  $|_{\circ}$  denotes the restriction of the operator to functions of  $y$ ,

and where the dots  $\dots$  indicate higher powers in  $h$ . Introduction of the  $|_{\circ}$  operator is motivated by the construction of quasi-modes, since the normal form is only being applied to the amplitude (a function of  $y$ ). A more conceptual explanation (suggested by the work of Iantchenko-Sjöstrand-Zworski [ISZ]) is that the correct Hilbert space on which to define the operators is the microlocal solution space of  $\Delta - \lambda^2$  along  $\gamma$ . These are essentially the space of quasi-modes along the open arc  $(0, L)$  of  $\gamma$ . Thus,  $|_{\circ}$  restricts the operator to the microlocal solution space.

Expanding the exponential, we find that the operator  $Q_{\frac{1}{2}}$  then must satisfy the *homological equation*

$$\{[L^{-1}\mathcal{D}, Q_{\frac{1}{2}}] + \mathcal{D}_{\frac{1}{2}}\} |_{\circ} = 0.$$

One may solve this equation explicitly by further conjugating  $\mathcal{D}$  to  $D_s$ . When  $\mathcal{D} = D_s$ , the homological equation becomes

$$\{[L^{-1}D_s, Q_{\frac{1}{2}}] + \mathcal{D}_{\frac{1}{2}}\} |_{\circ} = 0,$$

that is,

$$L^{-1}\partial_s Q_{\frac{1}{2}}|_o = -i\{\mathcal{D}_{\frac{1}{2}}\}|_o.$$

Here,  $\partial_s A$  is the Weyl operator whose complete symbol is the  $s$ -derivative of that of  $A$ . To solve this equation we rewrite it in terms of complete Weyl symbols. We will use the notation  $A(s, x, \xi)$  for the complete Weyl symbol of the operator  $A(s, x, D_x)$ . We then arrive at the homological equation

$$L^{-1}\partial_s \tilde{Q}_{\frac{1}{2}}(s, x, \xi) = -i\mathcal{D}_{\frac{1}{2}}|_o(s, x, \xi)$$

We solve with the Weyl symbol

$$\tilde{Q}_{\frac{1}{2}}(s, x, \xi) = \tilde{Q}_{\frac{1}{2}}(0, x, \xi) + L \int_0^s -i\mathcal{D}_{\frac{1}{2}}|_o(u, x, \xi) du$$

where  $\tilde{Q}_{\frac{1}{2}}(0, x, \xi)$  is determined by the consistency condition

$$\tilde{Q}_{\frac{1}{2}}(L, x, \xi) - \tilde{Q}_{\frac{1}{2}}(0, x, \xi) = L \int_0^L -i\mathcal{D}_{\frac{1}{2}}|_o(u, x, \xi) du.$$

To solve the equation, we invoke the fact (which is not obvious) that  $\mathcal{D}_{\frac{1}{2}}|_o(u, x, \xi)$  is a polynomial of degree 3 in  $(x, \xi)$ . We also switch to complex coordinates  $z_j = x_j + i\xi_j$  and  $\bar{z}_j = x_j - i\xi_j$  in which the action of  $r_\alpha(L)$  is diagonal. The homological equation becomes

$$\tilde{Q}_{\frac{1}{2}}(0, e^{i\alpha}z, e^{-i\alpha}\bar{z}) - \tilde{Q}_{\frac{1}{2}}(0, z, \bar{z}) = L \int_0^L -i\mathcal{D}_{\frac{1}{2}}|_o(u, z, \bar{z}) du.$$

We put:

$$\tilde{Q}_{\frac{1}{2}}(s, z, \bar{z}) = \sum_{|m|+|n|\leq 3} q_{\frac{1}{2};mn}(s) z^m \bar{z}^n$$

and

$$\mathcal{D}_{\frac{1}{2}}|_o(s, z, \bar{z}) du = \sum_{|m|+|n|\leq 3} d_{\frac{1}{2};mn}(s) z^m \bar{z}^n$$

then the homological equation becomes

$$\sum_{|m|+|n|\leq 3} (1 - e^{(m-n)\alpha}) q_{\frac{1}{2};mn}(0) z^m \bar{z}^n = -iL^2 \sum_{|m|+|n|\leq 3} \bar{d}_{\frac{1}{2};mn} z^m \bar{z}^n.$$

Since there are no terms with  $m = n$  in this (odd-index) equation, and since the  $\alpha_j$ 's are independent of  $\pi$  over  $\mathbb{Z}$ , there is no obstruction to the solution. Thus we can simply eliminate the term  $\mathcal{D}_{\frac{1}{2}}$ .

6.4.1. *The normal form coefficients.* We now consider the second step, where the first obstruction occurs. We thus seek a pseudodifferential operator  $\tilde{Q}_1(s, x, D_x)$  and a quadratic polynomial  $f_o(I_1, \dots, I_n)$  in the action operators so that

$$e^{-ih\tilde{Q}_1} \mathcal{D}_{\frac{1}{2}} e^{ih\tilde{Q}_1} = h^{-2}L^{-2} + h^{-1}L^{-1}D_s + h^{-\frac{1}{2}}\mathcal{D}_{\frac{1}{2}} + \mathcal{D}_o^1(s, D_s, x, D_x) + \dots$$

with

$$\mathcal{D}_o^1(s, D_s, x, D_x)|_o = f_o(I_1, \dots, I_n).$$

As usual, the dots signify terms of higher order in  $h$ . The homological equation is then

$$\{[D_s, \tilde{Q}_1] + \mathcal{D}_o^{\frac{1}{2}}\}|_o = f_o(I_1, \dots, I_n),$$

or equivalently

$$\partial_s \tilde{Q}_1|_o = \{-\mathcal{D}_o^{\frac{1}{2}} + f_o(I_1, \dots, I_n)\}|_o.$$

We rewrite the equation in terms of the complete Weyl symbols, to obtain

$$L^{-1} \partial_s \tilde{Q}_1(s, z, \bar{z}) = -i \{ \mathcal{D}_o^{\frac{1}{2}} |_o(s, z, \bar{z}) - f_o(|z_1|^2, \dots, |z_n|^2) \}$$

or equivalently

$$\tilde{Q}_1(s, z, \bar{z}) = \tilde{Q}_1(0, z, \bar{z}) - iL \int_0^s [\mathcal{D}_o^{\frac{1}{2}} |_o(u, z, \bar{z}) - f_o(|z_1|^2, \dots, |z_n|^2)] du,$$

or again,

$$\tilde{Q}_1(0, e^{i\alpha} z, e^{-i\alpha} \bar{z}) - \tilde{Q}_1(0, z, \bar{z}) = -iL \{ \int_0^L \mathcal{D}_o^{\frac{1}{2}} |_o(u, z, \bar{z}) du - L f_o(|z_1|^2, \dots, |z_n|^2) \}.$$

By construction,  $\mathcal{D}_o^{\frac{1}{2}} |_o(u, z, \bar{z})$  is a polynomial of degree 4, so if we put

$$\tilde{Q}_1(s, z, \bar{z}) = \sum_{|m|+|n|\leq 4} q_{1;mn}(s) z^m \bar{z}^n, \quad f_o(|z_1|^2, \dots, |z_n|^2) = \sum_{|k|\leq 2} c_{ok} |z|^{2k}$$

and

$$\mathcal{D}_o^{\frac{1}{2}} |_o(s, z, \bar{z}) du := \sum_{|m|+|n|\leq 4} d_{o;mn}^{\frac{1}{2}}(s) z^m \bar{z}^n, \quad d_{o;mn}^{\frac{1}{2}} := \frac{1}{L} \int_0^L d_{o;mn}^{\frac{1}{2}}(s) ds,$$

we can solve for the off-diagonal coefficients,

$$q_{1;mn}(0) = -iL^2 (1 - e^{i(m-n)\alpha})^{-1} d_{o;mn}^{\frac{1}{2}}.$$

We also must set the diagonal coefficients equal to zero, and this determines the  $c_{ok}$  coefficients:

$$c_{ok} = d_{1;kk}^{\frac{1}{2}}.$$

These coefficients are the quantum normal form coefficients. Since we used only algebraic operations on the rescaled Laplacian, it is clear that the coefficients can be calculated in terms of the data described in Theorem 5.1. We refer to [Z4, Z9] for further details.

**6.5. Bounded domains.** There are two serious differences in the normal form for the Laplacian around a bouncing ball orbit of a bounded domain. First, we need to straighten out the domain to define a model space. Secondly, we need the conjugation to normal form to incorporate the boundary conditions.

6.5.1. *The model domain  $\Omega_o$ .* The configuration space of the model is the infinite strip  $\Omega_o$ . We denote the coordinate on  $[0, L]$  by  $s$  and that on  $\mathbb{R}$  by  $y$ , with dual cotangent coordinates  $\sigma, \eta$  on  $T^*[0, L] \times T^*\mathbb{R}$ . As in the boundaryless case, we would like to view the model space as the normal bundle of the orbit. In the case of a bouncing ball orbit in the boundary case, the normal bundle and exponential map are ill-defined at the reflection points but Lazutkin has constructed a nice replacement for them. Namely, he constructs a formal power series map from  $\Omega_\epsilon$  to  $\Omega_0$ ,

$$\Phi : \Omega_\epsilon \rightarrow U_\epsilon, \quad \Phi(s, y) = (\bar{s}, \bar{y})$$

of the form

$$\bar{s} = s + \sum_{m=2}^\infty \kappa_{mm}(s) y^m$$

$$\bar{y} = \sum_{p=1}^\infty \psi_{pp}(s) y^p$$

with real valued analytic coefficients in  $s$  extending analytically to a neighborhood of  $[0, L]$  in  $\mathbb{C}$  and satisfying the boundary conditions:

$$\{\bar{s} = 0\} \cup \{\bar{s} = L\} = \Phi(\partial\Omega_\epsilon).$$

LEMMA 6.1. *Suppose that  $\overline{AB}$  is a bouncing ball orbit. Then there exists a transversal power series map  $\Phi : (\Omega_\epsilon, \partial\Omega_\epsilon) \rightarrow (\Omega_o, \partial\Omega_o)$  which straightens the domain and puts  $\bar{\Delta}$  in the form:*

$$\bar{\Delta} \sim D_{\bar{s}}^2 + B(\bar{s}, \bar{y})(\bar{y}^2 D_{\bar{s}}^2 + D_{\bar{y}}^2) + \Gamma'_s D_s + \Gamma'_y D_y \quad \text{elliptic case}$$

$$\bar{\Delta} \sim D_{\bar{s}}^2 + B(\bar{s}, \bar{y})(D_{\bar{y}}^2 - \bar{y}^2 D_{\bar{s}}^2) + \Gamma'_s D_s + \Gamma'_y D_y \quad \text{hyperbolic case}$$

in the sense that the left and right sides agree to infinite order at  $y = 0$ . Here,  $B(s, y)$  is a transversal power series.

By choosing  $\Phi$  carefully, one can arrange that the Laplacian  $(\Phi^{*-1} \Delta \Phi^*)_{\frac{1}{2}}$  in the transformed coordinates has the principal term  $D_{\bar{s}}^2 + (a_{22}(s)y^2 D_{\bar{s}}^2 + e_{11}(s)\bar{y} D_{\bar{y}} + b_{00}(s)D_{\bar{y}}^2$ . We can further put the principal part into the normal form

$$(6.4) \quad D_{\bar{s}}^2 + \frac{1}{2} \dot{b}_{00}(s)[\bar{y}^2 D_{\bar{s}}^2 + D_{\bar{y}}^2] \quad \text{elliptic case}$$

$$D_{\bar{s}}^2 + \frac{1}{2} \dot{b}_{00}(s)[D_{\bar{y}}^2 - \bar{y}^2 D_{\bar{s}}^2] \quad \text{hyperbolic case}$$

**6.6. Semi-classical scaling of the Laplacian.** Having straightened the domain, and hence having transferred the information about the boundary into the metric, we now follow the approach in the boundaryless case by scaling the Laplacian. Following [L, Z2], we use the notation  $N = h^{-1}$ .

We define operators  $T_N, M_N$  on the model space  $L^2(\Omega_o)$  by

$$(6.5) \quad \begin{aligned} &\bullet \quad T_N f(\bar{s}, \bar{y}) := N f(\bar{s}, N\bar{y}) \\ &\bullet \quad M_N f(\bar{s}, \bar{y}) := e^{iN^2 \bar{s}} f(\bar{s}, \bar{y}) \end{aligned}$$

DEFINITION 6.2. *The semiclassically scaled Laplacian is the operator on  $\Omega_o$  defined by*

$$\bar{\Delta}_N = M_N^* T_N^* \bar{\Delta} T_N M_N.$$

In the straightened form, the rescaled Weyl symbol of  $\bar{\Delta}$  has the form:

$$(6.6) \quad \sigma_{\bar{\Delta}_N}^w \sim (\sigma + N)^2 + B(s, \frac{1}{N}y)(\pm y^2 N^{-2}(\sigma + N^2)^2 + N^2 \eta^2) + K(s, \frac{1}{N}y).$$

**6.7. Conjugation to semiclassical normal form.** As in the boundaryless case, we begin by conjugating the principal term (which again is quadratic)

$$(6.7) \quad \sigma_{\bar{\Delta}_N}^w = N^4 + 2N^2[D_{\bar{s}} + \dot{b}_{00}(s)\hat{I}] \text{ mod } N$$

into quadratic normal form by an  $s$ -dependent metaplectic conjugation on the transverse space to the bouncing ball orbit. Here,  $\hat{I}$  denotes the quantum action operator:  $\hat{I}^e = \frac{1}{2}(D_{\bar{y}}^2 + y^2)$  in the elliptic case and  $\hat{I}^h = \frac{1}{2}(D_{\bar{y}}^2 - y^2)$  in the hyperbolic case.

In [Z2], we prove that there exists an  $SL(2, \mathbb{R})$ -valued function  $a_\alpha(\bar{s})$  so that

$$\mu(a_\alpha)^*[D_{\bar{s}} + b_{00}(\bar{s})\hat{I}]\mu(a_\alpha) = D_s + \frac{\alpha}{L}\hat{I}, \quad \mu(a_\alpha)(0) = \mu(a_\alpha)(L) = Id,$$

where  $\alpha = \int_0^L b_{00}(s)ds$ . We then conjugate the resulting operator

$$(6.8) \quad \mathcal{R}_N := \mu(a_\alpha)^* \bar{\Delta}_N \mu(a_\alpha)$$

to the semiclassical normal form

$$(6.9) \quad F_N(\hat{I})^2 \sim N^4 + N^2 \frac{\alpha \hat{I}}{L} + p_1(\hat{I}) + N^{-2} p_2(\hat{I}) + \dots$$

by conjugations as in the boundaryless case, but now additionally preserving the boundary conditions.

We assume for simplicity that the bouncing ball orbit is elliptic, but the same argument and result hold in the hyperbolic case. We again use the notation of (6.3): we denote by  $A|_o$  the restriction of  $A$  to functions of  $y$  only. In the boundary case, we need to construct complex valued Weyl symbols  $P_{j/2}, Q_{j/2}$  so that iterated composition with  $e^{N^{-j}(P+iQ)_{j/2}^w}$  will successively remove the lower order terms in  $\mathcal{R}_N$  after restriction by  $|_o$  and so that the boundary condition is satisfied. It turns out that the boundary condition on the conjugating operator involves only the odd part of the real part and the even part of the imaginary part:

$$(6.10) \quad P_{j/2}^o(0, \bar{y}, \bar{\eta}) = P_{j/2}^o(L, \bar{y}, \bar{\eta}) = 0, \quad Q_{j/2}^e(0, \bar{y}, \bar{\eta}) = Q_{j/2}^e(L, \bar{y}, \bar{\eta}) = 0.$$

We emphasize that there is no condition on the odd part  $Q_{j/2}^o$  of the imaginary part or the even part  $P_{j/2}^e$  of the real part.

6.7.1. *A Sturm-Liouville homological equation.* As before, we see what happens in the first and second steps. We first find  $P_{\frac{1}{2}}(\bar{s}, \bar{y}, D_{\bar{y}}), Q_{\frac{1}{2}}(\bar{s}, \bar{y}, D_{\bar{y}})$  so that the boundary conditions are satisfied and so that

$$(6.11) \quad e^{-N^{-1}(P+iQ)_{\frac{1}{2}}} \mathcal{R}_N e^{N^{-1}(P+iQ)_{\frac{1}{2}}}|_o = \{N^4 + N^2[D_s + \frac{\alpha}{L}\hat{I}] + \dots\}|_o.$$

Expanding the exponential, we get the *homological equation*:

$$(6.12) \quad \{[D_s + \frac{\alpha}{L}\hat{I}, (P+iQ)_{\frac{1}{2}}] + \mathcal{R}_{\frac{1}{2}}\}|_o = 0.$$

Taking the complete symbol of both sides we get the symbolic homological equation:

$$(6.13) \quad i\{\sigma + \frac{\alpha}{L}I, P_{\frac{1}{2}} + iQ_{\frac{1}{2}}\} + \mathcal{R}_{\frac{1}{2}}|_o = 0.$$

The equation may be rewritten in the form:

$$(6.14) \quad \partial_{\bar{s}}(P+iQ)_{\frac{1}{2}}(\bar{s}, r_{\alpha}(\bar{s})(\bar{y}, \bar{\eta})) = -i\mathcal{R}_{\frac{1}{2}}|_o,$$

whose solution is given by

$$(6.15) \quad (P+iQ)_{\frac{1}{2}}(\bar{s}, r_{\alpha}(\bar{s})(\bar{y}, \bar{\eta})) = (P+iQ)_{\frac{1}{2}}(0) - i \int_0^{\bar{s}} \mathcal{R}_{\frac{1}{2}}(u, \bar{y}, \bar{\eta}) du.$$

We thus need to determine  $(P+iQ)_{\frac{1}{2}}(0)$  so that the boundary conditions at  $\bar{s} = 0$  and  $\bar{s} = L$  are satisfied. Thus, we must solve the system

$$(6.16) \quad \begin{cases} \partial_{\bar{s}} P_{\frac{1}{2}}^e + \frac{\alpha}{L} \{I, P_{\frac{1}{2}}^e\} = \Im \mathcal{R}_{\frac{1}{2}}^e|_o & \partial_{\bar{s}} P_{\frac{1}{2}}^o + \frac{\alpha}{L} \{I, P_{\frac{1}{2}}^o\} = \Im \mathcal{R}_{\frac{1}{2}}^o|_o \\ \partial_{\bar{s}} Q_{\frac{1}{2}}^e + \frac{\alpha}{L} \{I, Q_{\frac{1}{2}}^e\} = -\Re \mathcal{R}_{\frac{1}{2}}^e|_o & \partial_{\bar{s}} Q_{\frac{1}{2}}^o + \frac{\alpha}{L} \{I, Q_{\frac{1}{2}}^o\} = -\Re \mathcal{R}_{\frac{1}{2}}^o|_o \end{cases}$$

with the boundary conditions

$$(6.17) \quad P_{\frac{1}{2}}^o(0) = P_{\frac{1}{2}}^o(L) = 0, \quad Q_{\frac{1}{2}}^e(0) = Q_{\frac{1}{2}}^e(L) = 0.$$

As in the boundaryless case,  $\mathcal{R}_{\frac{1}{2}}^{\pm}(u, \bar{y}, \bar{\eta}) = \bar{y} \circ (ar^{-1}(u))I$  is a polynomial of degree 3 in  $(\bar{y}, \bar{\eta})$  in which every term is of odd degree in  $(\bar{y}, \bar{\eta})$ . It follows that  $P_{\frac{1}{2}}, Q_{\frac{1}{2}}$  are also odd polynomials of degree 3. Assuming we are in the elliptic case,



we change coordinates to the complex cotangent variables  $z = \bar{y} + i\bar{\eta}$ ,  $\bar{z} = \bar{y} - i\bar{\eta}$  and write

$$(6.18) \quad \begin{cases} P_{\frac{1}{2}}^e(\bar{s}, z, \bar{z}) = \sum_{m,n:m+n \leq 3} p_{\frac{1}{2}mn}^e(\bar{s})(z^m \bar{z}^n + \bar{z}^m z^n), \\ P_{\frac{1}{2}}^o(\bar{s}, z, \bar{z}) = \sum_{m,n:m+n \leq 3} p_{\frac{1}{2}mn}^o(\bar{s})(z^m \bar{z}^n - \bar{z}^m z^n), \\ Q_{\frac{1}{2}}^e(\bar{s}, z, \bar{z}) = \sum_{m,n:m+n \leq 3} q_{\frac{1}{2}mn}^e(\bar{s})(z^m \bar{z}^n + \bar{z}^m z^n), \\ Q_{\frac{1}{2}}^o(\bar{s}, z, \bar{z}) = \sum_{m,n:m+n \leq 3} q_{\frac{1}{2}mn}^o(\bar{s})(z^m \bar{z}^n - \bar{z}^m z^n). \end{cases}$$

Then (6.16) may be rewritten in terms of these coordinates. In the elliptic case, we have:

$$(6.19) \quad \begin{cases} \frac{d}{ds} p_{\frac{1}{2}mn}^e(s) + i \frac{\alpha}{L} (m-n) p_{\frac{1}{2}mn}^o(s) = \Im \mathcal{R}_{\frac{1}{2}}^e|_o \\ \frac{d}{ds} p_{\frac{1}{2}mn}^o(s) + i \frac{\alpha}{L} (m-n) p_{\frac{1}{2}mn}^e(s) = \Im \mathcal{R}_{\frac{1}{2}}^o|_o \\ \frac{d}{ds} q_{\frac{1}{2}mn}^e(s) + i \frac{\alpha}{L} (m-n) q_{\frac{1}{2}mn}^o(s) = -\Re \mathcal{R}_{\frac{1}{2}}^e|_o \\ \frac{d}{ds} q_{\frac{1}{2}mn}^o(s) + i \frac{\alpha}{L} (m-n) q_{\frac{1}{2}mn}^e(s) = -\Re \mathcal{R}_{\frac{1}{2}}^o|_o \end{cases}$$

Since  $m \neq n$ , we can eliminate  $p_{\frac{1}{2}mn}^e, q_{\frac{1}{2}mn}^o$  and reduce the homological equation to (uncoupled) second order Sturm-Liouville boundary value problems

$$(6.20) \quad \begin{cases} -\frac{d^2}{ds^2} p_{\frac{1}{2}mn}^o(s) - [\frac{\alpha}{L}(m-n)]^2 p_{\frac{1}{2}mn}^o(s) = \frac{L}{\alpha(m-n)} \frac{d}{ds} \Im \mathcal{R}_{\frac{1}{2}}^e|_o + \Im \mathcal{R}_{\frac{1}{2}}^e|_o \\ -\frac{d^2}{ds^2} q_{\frac{1}{2}mn}^e(s) - [\frac{\alpha}{L}(m-n)]^2 q_{\frac{1}{2}mn}^e(s) = -\frac{L}{\alpha(m-n)} \frac{d}{ds} \Re \mathcal{R}_{\frac{1}{2}}^e|_o + \Re \mathcal{R}_{\frac{1}{2}}^e|_o \\ p_{\frac{1}{2}mn}^o(0) = 0, p_{\frac{1}{2}mn}^o(L) = 0; \quad q_{\frac{1}{2}mn}^e(0) = 0, q_{\frac{1}{2}mn}^e(L) = 0 \end{cases}$$

for the independent variable  $p_{\frac{1}{2}mn}^o, q_{\frac{1}{2}mn}^e(s)$ . The boundary value problem (6.20) is always solvable unless 0 is an eigenvalue of the operator  $D_{\bar{s}}^2 - [\frac{\alpha}{L}(m-n)]^2$  (elliptic case). The eigenfunction would have to have the form  $\sin(\frac{\alpha}{L}(m-n)\bar{s})$ , hence a sufficient condition for solvability is that  $\alpha/\pi \notin \mathbb{Q}$ . The analogous boundary problem in the hyperbolic case is always solvable.

Thus, the conjugation eliminated the sub-principal term  $\mathcal{R}_{\frac{1}{2}}$  and put our operator in normal form up to order  $N$ . The exponents  $P_{\frac{1}{2}}, Q_{\frac{1}{2}}$  are odd polynomial differential operators of degree 3 with smooth coefficients defined in a neighborhood of  $[0, L]$ .

6.7.2. *The normal form coefficients emerge.* We carry the process forward one more step because, as in the boundaryless case, the even steps produce the non-trivial normal form coefficients. In outline, the conjugated operator in the second step,  $\mathcal{R}_{\frac{1}{2}}$ , is further conjugated with an exponential of the form  $e^{N^{-2}(P+iQ)_1}$ . The homological equation is again a Sturm-Liouville boundary problem, with the new feature that diagonal terms with  $m = n$  do occur in (6.19). In the elliptic case, a diagonal term is a function of  $|z|^2$ , hence is even under the involution  $z \rightarrow \bar{z}$ . We

write the diagonal terms in the form:

$$(6.21) \quad \begin{cases} P_j^d(\bar{s}, |z|^2) = \sum_{m:m \leq j} p_{jm}^d(\bar{s})|z|^{2m} \\ Q_j^d(\bar{s}, |z|^2) = \sum_{m:m \leq j} q_{jm}^d(\bar{s})|z|^{2m} \end{cases}$$

It is impossible to solve the inhomogeneous boundary problem (6.19) for the diagonal terms as in the odd case with zero boundary conditions. To obtain a solvable system, we must add new terms  $f_j(|z|^2)$  to the right side:

$$(6.22) \quad \begin{cases} \frac{d}{ds} p_{jm}^d(s) = \Im \mathcal{R}_{2-j}^d|_o - \Im f_j(|z|^2) \\ \frac{d}{ds} q_{jm}^d(s) = -\Re \mathcal{R}_{2-j}^d|_o + \Re f_j(|z|^2) \\ p_{jm}^d(0) = p_{jm}^d(L) = 0 \\ q_{jm}^d(0) = q_{jm}^d(L) = 0. \end{cases}$$

The right hand sides are determined by the condition that equations with

$$(6.23) \quad \begin{cases} p_{jm}^d(s) = \int_0^s \{-\Im \mathcal{R}_{2-j}^{j,d}|_o(u, |z|^2) + \Re f_j(|z|^2)\} \\ q_{jm}^d(s) = \int_0^s \{\Re \mathcal{R}_{2-j}^{j,d}|_o(u, |z|^2) + \Im f_j(|z|^2)\} du, \end{cases}$$

satisfies the boundary conditions. This forces

$$(6.24) \quad f_j(|z|^2) = -\frac{1}{L} \int_0^L \mathcal{R}_{2-j}^{j,d}|_o(u, |z|^2) du.$$

In the case  $j = 1$ , we have then conjugated  $\mathcal{R}_N^+$  to  $N^4 + N^2\mathcal{R} + Op^w(f_1(I)) + O(N^{-1})$ . The coefficients of the polynomial  $f_1(I)$  are the first Birkhoff invariants. We note that  $Op^w(f_1(I))$  is a function of  $\hat{I}$ , so we have conjugated to normal form to fourth order. We now repeat the process to complete the normal form.

**6.8. Monodromy operator.** We now outline a different approach to normal forms due to Iantchenko-Sjöstrand-Zworski [SjZ, ISZ]. (See also the appendix). Their goal is to construct the normal form of a so-called *quantum monodromy operator* rather than of the Laplacian or wave group. The normal form of the monodromy operator is apparently the same as the semi-classical normal form of the wave group that we constructed in §6.4 and §6.7. However, it is presented in a conceptually clearer way which works equally well in the boundary or boundaryless case. On the other hand, the method presented above also gives an algorithm for calculating the normal form, which is at present lacking in the approach of [ISZ].

The new concepts introduced in [SjZ, ISZ] following earlier one-dimensional definitions in works of Helffer-Sjöstrand and Colin de Verdière-Parisse) are the following:

- The monodromy operator  $\mathcal{M}(\lambda) : \ker_{m_0}(\sqrt{\Delta} - \lambda) \rightarrow \ker_{m_0}(\sqrt{\Delta} - \lambda)$  acting on the space  $\ker_{m_0}(\sqrt{\Delta} - \lambda)$  of microlocal solutions of the equation  $(\sqrt{\Delta} - \lambda)u(h) = O(\lambda^{-\infty})$  near  $m_0$ , where  $m_0$  is an arbitrarily chosen base point of the given geodesic  $\gamma$ .
- The *flux norm* on the microlocal solution space  $\ker_{m_0}(\sqrt{\Delta} - \lambda)$ , with respect to which  $\mathcal{M}(\lambda)$  is unitary;

- The Grushin reduction of the wave operator to the monodromy operator and the expression of the wave trace as a trace of the monodromy operator.

The microlocal solution space is essentially the space of quasi-modes along arcs of  $\gamma$ , so the monodromy approach is not in essence so different from the approach we outline above. We give a brief exposition of the key ideas, which are discussed in more detail in the appendix by Sjöstrand-Zworski.

Assuming for simplicity that  $\gamma$  projects to an embedded curve in  $M$ , we again work on the model space  $S^1 \times \mathbb{R}^{n-1}$  given by the normal bundle  $N_\gamma$  to  $M$ . We then consider the universal cover  $\mathbb{R} \times \mathbb{R}^{n-1}$ , fix the base point  $m = (0, 0)$ , and consider microlocal solutions along the geodesic  $\mathbb{R} \times \{0\}$ . In the simply connected space  $\mathbb{R} \times \mathbb{R}^{n-1}$ , there is no obstruction to constructing a global quasimode along  $\mathbb{R} \times \{0\}$ . So we may define

$$(6.25) \quad \mathcal{M}(\lambda) : \ker_0(\sqrt{\Delta} - \lambda) \rightarrow \ker_L(\sqrt{\Delta} - \lambda)$$

where  $L = L_\gamma$  by

$$(6.26) \quad \mathcal{M}(\lambda)[u_\lambda]_0 = [u_\lambda]_{(L,0)}$$

where  $[u_\lambda]_x$  is the germ of the quasimode at the point  $x$ . We obtain a quasi-mode when  $(I - \mathcal{M}(\lambda))[u_\lambda]_0 = 0$ .

To make contact with the wave group, we quote the following observation of Sjöstrand-Zworski (see the appendix):

PROPOSITION 6.3. *We have:*

$$\mathcal{M}(\lambda) = \exp(-iL(\sqrt{\Delta} - \lambda)) : \ker_0(\sqrt{\Delta} - \lambda) \rightarrow \ker_L(\sqrt{\Delta} - \lambda).$$

**Sketch of Proof** (See the appendix for more details) It is obvious that  $\exp(-iL(\sqrt{\Delta} - \lambda))$  takes microlocal solutions into themselves since it commutes with  $\sqrt{\Delta}$ . The rest of the proof is to verify that the complete symbol of the quasimode  $\exp(-iL(\sqrt{\Delta} - \lambda))u_\lambda(s, y)$  at  $(0, 0)$  is the same as that of  $u_\lambda(s, y)$  at  $s = L$ . QED

6.8.1. *Flux norm.* We now describe the *flux norm*  $\langle \cdot, \cdot \rangle$  on  $\ker_m(\sqrt{\Delta} - \lambda)$ . It is motivated by the properties of the probability current density  $\mathbf{j}$  of a solution  $\psi(x, t)$  of the Schrödinger equation in quantum mechanics (see e.g. [LL], §19). It is defined by

$$\mathbf{j} = \frac{\hbar}{i}(\psi \nabla \bar{\psi} - \bar{\psi} \nabla \psi).$$

The integral  $\int_S \mathbf{j} \cdot dA$  over a surface  $S$  measures the probability that the particle described by  $\psi$  will cross the surface  $S$  in a unit of time. As is pointed out in [LL], if  $\psi = Ae^{\frac{i}{\hbar}S}$  then  $\mathbf{j} = |A|^2 \nabla S$ .

The flux norm on microlocal solutions is an invariantly defined version of this. We consider the microlocal solution space  $\ker_{m_0}$  for  $P = \hbar^2 \Delta - 1$  along an initial arc of a closed geodesic  $\gamma$ . We let  $Y$  denote a transversal to  $\gamma$  in  $M$ . If  $\nu_Y$  is the unit normal to  $Y$ , then

$$\|u\|_{QF}^2 = \int_Y \mathbf{j} \cdot \nu_Y \, dvol_Y, \quad u \in \ker_{m_0}(P).$$

More precisely and generally,

$$\|u\|_{QF}^2 = \frac{i}{\hbar} \langle [P, \chi]u, u \rangle,$$

where we use the semi-classical notation  $P = -h^2 \Delta - 1$  with  $h = \lambda^{-1}$ , and where  $\chi$  is a microlocal cutoff defined in the universal cover of a small tube around  $\gamma$ ,

whose complete symbol equals 0 before  $m_0$  and 1 after  $m_0$  in a somewhat smaller tube. We use the semi-classical notation to conform to the notation of [ISZ, S $\mathbf{j}$ Z] and also because their results apply to much more general semi-classical operators  $P(h)$ . For further details we refer to [ISZ, S $\mathbf{j}$ Z].

PROPOSITION 6.4. [ISZ] *The monodromy operator  $\mathcal{M}$  is a unitary operator on  $\ker_{m_0}(\sqrt{\Delta} - \lambda)$  with respect to the flux norm  $\|\cdot\|_{QF}$ .*

6.8.2. *Grushin problem.* One of the basic steps of [S $\mathbf{j}$ Z] is the proof of a trace formula which relates the trace of the wave group at  $\gamma$  to the traces involving the monodromy operator. Formally, the trace formula says roughly that

$$(6.27) \quad \text{tr} \rho(P/h) \chi(P) A = \sum_{\pm} \text{tr} \int_{\mathbb{R}} \rho(z/h) \frac{d}{dz} \log \det(I + \mathcal{M}(z \pm i0)) \chi(z) dz,$$

where  $\hat{\rho} \in C_0^\infty(L_\gamma - \epsilon, L_\gamma + \epsilon)$  (compare to (7.1)), and  $\chi$  is a cutoff to the sphere bundle. The precise statement is given in Theorem 1 of [S $\mathbf{j}$ Z]:

$$(6.28) \quad \text{tr} \rho(P/h) \chi(P) A = \frac{1}{2\pi} \sum_{-N-1}^{N+1} \text{tr} \int_{\mathbb{R}} \rho(z/h) \mathcal{M}(z, h)^k \frac{d}{dz} \mathcal{M}(z, h) \chi(z) dz + O(h^\infty).$$

This formula is analogous to one in the case of domains with boundary, where the boundary replaces the transversal, see Proposition 7.1.

To prove the formula, one sets up a microlocal Grushin problem near the closed trajectory  $\gamma$ . One begins by forming a microlocally invertible system

$$\mathcal{P}(z) = \begin{pmatrix} \frac{i}{h} P(z) & R_-(z) \\ R_+(z) & 0 \end{pmatrix} : \mathcal{D}'(M) \times \mathcal{D}'(\mathbb{R}^{n-1}) \rightarrow \mathcal{D}'(M) \times \mathcal{D}'(\mathbb{R}^{n-1}),$$

where  $R_\pm(z)$  are defined microlocally near  $\gamma$ . The microlocal inverse of  $\mathcal{P}(z)$  is given by:

$$(6.29) \quad \mathcal{E}(z) = \begin{pmatrix} E(z) & E_+(z) \\ E_-(z) & E_{-+}(z) \end{pmatrix}$$

We thus obtain the key formula

$$(6.30) \quad (\sqrt{\Delta} - \lambda)^{-1} \sim E(\lambda) - E_+(\lambda)(I - \mathcal{M})^{-1}E_-(\lambda),$$

which reduces the eigenvalue problem to the microlocal invertibility of  $(I - \mathcal{M})$ . Using this formula one proves (6.27). This equation is analogous to one which occurs in the Fredholm-Neumann reduction of the Dirichlet problem to the boundary of this type (see 7.4).

The authors of [ISZ, S $\mathbf{j}$ Z, IS $\mathbf{j}$ ] then put  $\mathcal{M}(\lambda)$  into quantum Birkhoff normal form. In view of Proposition 6.3, the normal form of the monodromy operator is just what we called the semi-classical normal form of the wave group. The operation  $|_o$  thus has been interpreted as restriction to the microlocal solution space, or as part of a Grushin reduction to the transversal. In [Z4, Z5], we proved that the semi-classical normal form was a spectral invariant by first turning into the homogeneous normal form, by proving that the homogeneous normal form was a spectral invariant and then by relating the semi-classical and homogeneous normal forms. The formula (6.27) eliminates this latter step and directly shows that the semi-classical normal form is a spectral invariant by using the Grushin reduction.

At the present time, the only application of the monodromy method to inverse spectral results for plane domains is for the case of bouncing ball orbits of analytic domains with two symmetries as in [Z2]. This does not involve any new calculations since the principal symbol of the normal form of the monodromy operator is the classical Birkhoff normal form of  $\mathcal{P}_\gamma$  and, as mentioned above, this normal form determines  $\Omega$  when it is analytic with two symmetries (as proved in [CdV]).

### 7. Calculation of wave invariants II: Balian-Bloch approach

In this section, we describe a method for calculating wave invariants of bounded domains which does not use normal forms or parametrices, but rather is based on an exact formula for the Dirichlet (or Neumann) resolvent in terms of the ‘free resolvent’. Our goal is to explain how the calculations of Theorem 5.2 and Corollary 5.3 are done.

For simplicity, we will confine ourselves to calculating wave invariants at a bouncing ball orbit of a bounded simply connected domain  $\Omega \subset \mathbb{R}^2$ . The method extends with no difficulty to higher dimensions, more general metrics and domains (in particular, non-convex domains) and more general periodic orbits, but the special case is already sufficiently rich and difficult.

The method is based on the use of a classical formula due to C. Neumann and I. Fredholm for the resolvent  $R_B^\Omega(z)$  for the boundary problem in  $\Omega$  with boundary conditions  $B$  in terms of layer potentials and the boundary integral operators they induce. This approach was first used by the physicists Balian-Bloch [BB1, BB2] in their well-known work on the Poisson relation on three dimensional Euclidean domains, and we will refer to it as the Balian-Bloch approach. The crucial advantage of this approach over the use of normal forms or monodromy operators is its computability. For the first time, one can calculate wave invariants of all orders explicitly and find out what information they contain.

**7.1. Wave invariants as semi-classical resolvent trace invariants.** First, we explain how wave trace expansions are equivalent to resolvent expansions. We fix a non-degenerate bouncing ball orbit  $\gamma$  of length  $L_\gamma$ , and let  $\hat{\rho} \in C_0^\infty(L_\gamma - \epsilon, L_\gamma + \epsilon)$  be a cutoff, equal to one on an interval  $(L_\gamma - \epsilon/2, L_\gamma + \epsilon/2)$  which contains no other lengths in  $\text{Lsp}(\Omega)$  occur in its support. We then define the regularized resolvent by

$$(7.1) \quad R_\rho^\Omega(k + i\tau) := \int_{\mathbb{R}} \rho(k - \mu)(\mu + i\tau)R_B^\Omega(\mu + i\tau)d\mu.$$

From the resolvent identity (e.g.)

$$R_B^\Omega(\mu + i\tau) = \frac{1}{\mu + i\tau} \int_0^\infty e^{i(\mu+i\tau)t} E_B^{\Omega^c}(t)dt,$$

it follows that

$$(7.2) \quad R_{\rho_B}^\Omega(k + i\tau) = \int_0^\infty \hat{\rho}(t)e^{i(k+i\tau)t} E_B^\Omega(t)dt.$$

When  $\gamma, \gamma^{-1}$  are the unique closed orbits of length  $L_\gamma$ , it follows from the Poisson relation for manifolds with boundary (see §3 and [GM, PS]) that the trace

$Tr1_{\Omega}R_{\rho}((k + i\tau))$  of the regularized resolvent on  $L^2(\Omega)$  admits a complete asymptotic expansion of the form:

$$(7.3) \quad Tr1_{\Omega}R_{\rho}(k + i\tau) \sim e^{(ik - \tau)L_{\gamma}} \sum_{j=1}^{\infty} (B_{\gamma,j} + B_{\gamma^{-1},j})k^{-j}, \quad k \rightarrow \infty$$

with coefficients  $B_{\gamma,j}, B_{\gamma^{-1},j}$  determined by the jet of  $\Omega$  at the reflection points of  $\gamma$ . The coefficients  $B_{\gamma,j}, B_{\gamma^{-1},j}$  are thus essentially the same as the wave trace coefficients at the singularity  $t = L_{\gamma}$ .

**7.2. Reduction to the boundary.** To calculate  $Tr1_{\Omega}R_{\rho}(k + i\tau)$  asymptotically, we use the Fredholm-Neumann reduction of the Dirichlet or Neumann problems in a bounded domain to the boundary. The key formula is similar to (6.30):

$$(7.4) \quad R_{\Omega}(k + i\tau) = R_0(k + i\tau) - \mathcal{D}\ell(k + i\tau)(I + N(k + i\tau))^{-1}\gamma\mathcal{S}\ell^{tr}(k + i\tau),$$

where  $R_0(k + i\tau)$  is the free resolvent  $-(\Delta_0 + (k + i\tau)^2)^{-1}$  on  $\mathbb{R}^2$ . Here,  $\gamma : H^s(\Omega) \rightarrow H^{s-1/2}(\partial\Omega)$  is the restriction to the boundary, and  $\mathcal{D}\ell(k + i\tau)$  (resp.  $\mathcal{S}\ell(k + i\tau)$ ) is the double (resp. single) layer potential is the operator from  $H^s(\partial\Omega) \rightarrow H_{loc}^{s+1/2}(\Omega)$  defined by

$$(7.5) \quad \begin{cases} \mathcal{S}\ell(k + i\tau)f(x) = \int_{\partial\Omega} G_0(k + i\tau, x, q)f(q)ds(q), \\ \mathcal{D}\ell(k + i\tau)f(x) = \int_{\partial\Omega} \frac{\partial}{\partial\nu_y} G_0(k + i\tau, x, q)f(q)ds(q), \end{cases}$$

where  $ds(q)$  is the arc-length measure on  $\partial\Omega$ , where  $\nu$  is the interior unit normal to  $\Omega$ , and where  $\partial_{\nu} = \nu \cdot \nabla$ . Also,  $\mathcal{S}\ell^{tr}$  is its transpose (from the interior to the boundary), and  $G_0(z, x, y)$  is the free Green's function, i.e. the kernel of  $R_0(z)$ . Further,

$$(7.6) \quad N(k + i\tau)f(q) = 2 \int_{\partial\Omega} \frac{\partial}{\partial\nu_y} G_0(k + i\tau, q, q')f(q')ds(q')$$

is the boundary integral operator induced by  $\mathcal{D}\ell$ . It is classical that this operator satisfies

$$(7.7) \quad \begin{aligned} (i) \quad & N(k + i\tau) \in \Psi^{-1}(\partial\Omega), \\ (ii) \quad & (I + N(k + i\tau)) : H^s(\partial\Omega) \rightarrow H^s(\partial\Omega) \text{ is an isomorphism.} \end{aligned}$$

The reader might now compare the formulae (7.4) and (6.30) to see how  $N(\lambda)$  is analogous to  $\mathcal{M}(\lambda)$ . Moreover, it is used in an analogous way to reduce the calculation of the (distribution) trace of  $R_{\Omega}(k + i\tau)$ , formula (7.4) to the boundary. In doing so, we found it [Z10] more convenient to combine the interior and exterior problems as follows: We write  $L^2(\mathbb{R}^2) = L^2(\Omega) \oplus L^2(\Omega^c)$  and let  $R_N^{\Omega}$ , resp.  $(k + i\tau), R_D^{\Omega^c}(k + i\tau)$  denote the Neumann resolvent on the exterior domain, resp. the Dirichlet resolvent on the interior domain. We then regard  $R_D^{\Omega^c}(k + i\tau) \oplus R_N^{\Omega}(k + i\tau)$  as an operator on this space. The reason for combining the interior/exterior problems is that we can cycle around the layer potentials in (7.4) when taking the trace and simplify the formula to:

$$(7.8) \quad Tr_{\mathbb{R}^2}[R_D^{\Omega^c}(k + i\tau) \oplus R_N^{\Omega}(k + i\tau) - R_0(k + i\tau)] = \frac{d}{dk} \log \det \left( I + N(k + i\tau) \right),$$

where the determinant is the usual Fredholm determinant. We refer to [Z10] for a proof of this apparently well-known formula. This gives:

PROPOSITION 7.1. *Suppose that  $L_\gamma$  is the only length in the support of  $\hat{\rho}$ . Then,*

$$\int_{\mathbb{R}} \rho(k - \lambda) \frac{d}{d\lambda} \log \det(I + N(\lambda + i\tau)) d\lambda \sim \sum_{j=0}^{\infty} B_{\gamma;j} k^{-j},$$

where  $B_{\gamma;j}$  are the wave invariants of  $\gamma$ .

**7.3. Semi-classical analysis of  $N(k + i\tau)$ .** The next step is to analyse the operator  $N(k + i\tau)$  and the geometric series expansion of  $(I + N(k + i\tau))^{-1}$ . The operator  $N(k + i\tau)$  has the singularity of a homogeneous pseudodifferential operator of order  $-1$  on the diagonal (in fact, it is of order  $-2$  in dimension 2) and that is the way it is normally described in potential theory. However, away from the diagonal, it has a WKB approximation which exhibits it as a semi-classical Fourier integral operator with phase  $d_{\partial\Omega}(q, q') = |q - q'|$  on  $\partial\Omega \times \partial\Omega$ , the boundary distance function of  $\Omega$ . Indeed, the free Green's function in dimension two is given by:

$$G_0(k + i\tau, x, y) = H_0^{(1)}((k + i\tau)|x - y|) = \int_{\mathbb{R}^2} e^{i\langle x-y, \xi \rangle} (|\xi|^2 - (k + i\tau)^2)^{-1} d\xi.$$

Here,  $H_0^{(1)}(z)$  is the Hankel function of index 0. It has the near diagonal and off-diagonal asymptotics,

$$(7.9) \quad (i) \quad H_0^{(1)}(z) \sim \begin{cases} -\frac{1}{2\pi} \ln |z| \text{ as } |z| \rightarrow 0, \\ (ii) \quad e^{i(z|x-y|-\pi/4)} \frac{1}{|z|^{1/2}} \text{ as } |z| \rightarrow \infty. \end{cases}$$

By the explicit formula we have:

$$\begin{aligned} \frac{1}{2} N(k + i\tau, q(\phi), q(\phi')) &= \partial_{\nu_y} G_0(\mu, q(\phi), q(\phi')) \\ &= - (k + i\tau) H_1^{(1)}(k + i\tau|q(\phi) - q(\phi')|) \\ &\quad \times \cos \angle(q(\phi) - q(\phi'), \nu_{q(\phi)}). \end{aligned}$$

Combining with (7.9(ii)), we see that when  $|q(\phi) - q(\phi')| \geq |k|^{1-\epsilon}$  for some  $\epsilon < 1$ ,  $N(k + i\tau, q(\phi), q(\phi'))$  is a semi-classical Fourier integral kernel whose phase is the boundary distance function. For a convex domain, the boundary distance function generates the billiard map of  $\partial\Omega$  and hence we view  $N(k + i\tau)$  as a global quantization of the billiard map. For non-convex domains, the boundary distance function additionally generates ‘ghost orbits’ which in part exit the domain, but these only present a mild complication. We refer to [HZel] for discussion of these orbits and to [Z2, HZel] for further discussion of  $N(k)$ .

We now explain how to use Proposition (7.1) to calculate the wave trace coefficients  $B_{\gamma,j}$  at a closed geodesic. When calculating these coefficients we first compose with a special kind of semiclassical cutoff operator  $\chi(x, k^{-1}D_x)$  to a neighborhood of  $\gamma$  on both sides of (7.4). Since we are not dealing with conventional Fourier integral operators, it must be proved that composition with such a cutoff does in fact microlocalize to  $\gamma$ . We will suppress this issue until the end.

At least formally, we expand  $(I + N(k + i\tau))^{-1}$  in a finite geometric series plus remainder:

$$(7.10) \quad (I + N(\lambda + i\tau))^{-1} = \sum_{M=0}^{M_0} (-1)^M N(\lambda)^M + (-1)^{M_0+1} N(\lambda)^{M_0+1} (I + N(\lambda + i\tau))^{-1}.$$

To calculate a given wave invariant, we need to show that, for each order  $k^{-J}$  in the trace expansion of Corollary (7.1), there exists  $M_0(J)$  such that

$$(7.11) \quad \begin{aligned} (i) \quad & \sum_{M=0}^{M_0} (-1)^M \text{Tr} \int_{\mathbb{R}} \rho(k - \lambda) N(\lambda)^M N'(k + i\tau) d\lambda \\ & = \sum_{j=0}^J B_{\gamma;j} k^{-j} + O(k^{-J-1}), \\ (ii) \quad & \text{Tr} \int_{\mathbb{R}} \rho(k - \lambda) N(\lambda)^{M_0+1} (I + N(\lambda + i\tau))^{-1} N'(k + i\tau) d\lambda = O(k^{-J-1}). \end{aligned}$$

We outline the method for obtaining the wave trace asymptotics at a closed geodesic  $\gamma$  from the first term. Since  $N$  has singularities on the diagonal, one cannot just apply the stationary phase method to the trace. Rather one has to regularize the operator. We do this by separating out the tangential and transversal parts of  $N$  by introducing a cutoff of the form  $\chi(k^{1-\delta}|q - q'|)$  to the diagonal, where  $\delta > 1/2$  and where  $\chi \in C_0^\infty(\mathbb{R})$  is a cutoff to a neighborhood of 0. We then put

$$(7.12) \quad N(k + i\tau) = N_0(k + i\tau) + N_1(k + i\tau), \quad \text{with}$$

$$(7.13) \quad \begin{cases} N_0(k + i\tau, q, q') = \chi(k^{1-\delta}|q - q'|) N(k + i\tau, q, q'), \\ N_1(k + i\tau, q, q') = (1 - \chi(k^{1-\delta}|q - q'|)) N(k + i\tau, q, q'). \end{cases}$$

The term  $N_1$  is a semiclassical Fourier integral kernel quantizing the billiard map, while  $N_0$  behaves like an Airy operator close to the diagonal with the singularity of a homogeneous pseudodifferential operator on the diagonal.

Now consider the powers  $N(k + i\tau)^M$  in the first term (i) of (7.11). We write

$$(7.14) \quad (N_0 + N_1)^M = \sum_{\sigma: \{1, \dots, M\} \rightarrow \{0, 1\}} N_{\sigma(1)} \circ N_{\sigma(2)} \circ \dots \circ N_{\sigma(M)}.$$

We regularize  $N^M$  by eliminating the factors of  $N_0$  from each of these terms. This is obviously not possible for the term  $N_0^M$  but it is possible for the other terms. By explicitly writing out the composition in terms of Hankel functions and using the basic identities for these special functions, we prove that  $N_0 \circ N_1 \circ \chi_0(k + i\tau, \phi_1, \phi_2)$  is a semiclassical Fourier integral operator on  $\partial\Omega$  of order  $-1$  associated to the billiard map. Thus, composition with  $N_0$  lowers the order. The remaining terms  $N_0^M$ , when composed with a cut-off to  $\gamma$ , do not contribute asymptotically to the trace.

The successive removal of the factors of  $N_0$  thus gives a semi-classical quantization of the billiard map near  $\gamma$ . We then calculate the traces of each term by the stationary phase method and obtain the result stated in (5.2). The terms displayed there with the maximum number of derivatives of the defining function of  $\partial\Omega$  come only from the  $N_1^M$  terms. Thus, although the removal of the  $N_0$  factors does change the amplitude of the  $N_1$  factors and does contribute to the wave trace, it turns out to be negligible in the inverse spectral problem. This is one of the principal virtues of the Balian-Bloch approach.



We will explain how the explicit formulae for the wave invariants are derived in the next section.

**7.3.1. Remainder estimate.** We now address some sketchy remarks to the remainder estimate, in the hopes of convincing the reader that it is plausible to expect the remainder trace to be small. The ability to insert a microlocal cutoff to  $\gamma$  is crucial here. The main obstacle to the remainder is that the norm of  $N(k + i\tau)$  fails to decrease with increasing  $\tau$  due to the Airy part associated to creeping rays. This again is a difference to the monodromy operator.

To obtain a small remainder we set the spectral parameter in  $N$  equal to  $k + i\tau \log k$ . The presence of the  $\log k$  in the imaginary part changes the wave trace expansions by  $k^{-C\tau L_\gamma}$ , but this does not hurt the expansions such the remainder estimates will be of lower order. We then estimate the remainder

$$\text{Tr} \int_{\mathbb{R}} \rho(k - \lambda) N(\lambda)^{M_0+1} (I + N(\lambda + i\tau))^{-1} \chi_\gamma N'(k + i\tau) d\lambda$$

by applying the Schwarz inequality for the Hilbert-Schmidt inner product, and using the relation  $(I + N(\lambda + i\tau))^{-1}$  to the Poisson kernel to estimate this factor. These estimates leave a trace of  $N^M N^{*M}$  for fixed  $M$ , microlocalized to  $\gamma$ . We regularize these traces as above and obtain oscillatory integrals whose phases have critical points corresponding to  $M$ -link closed circuits which being at some point  $q$ , bounce along the boundary until  $q'$  and then return to  $q$  by traversing the links in reverse order. However, the cutoffs to  $\gamma$  force the links in critical paths to point in the direction of  $\gamma$  and hence to be of length roughly  $ML_\gamma$ . The imaginary part  $i\tau \log k$  of the semiclassical parameter then contributes a damping factor of  $e^{-\tau ML_\gamma \log k}$  for each link. The links correspond to the  $N_1$  factors. Thus, for each string, we have one  $k^{-1}$  for each  $N_0$  factor and one  $e^{-\tau ML_\gamma \log k}$  for each  $N_1$  factor. For sufficiently large  $\tau$  these combine to give a factor of  $k^{-R}$  for any prescribed  $R$ .

**7.3.2.  $N(\lambda)$  versus  $\mathcal{M}(\lambda)$ .** We digress momentarily to compare  $N(\lambda)$  and the monodromy operator  $\mathcal{M}(\lambda)$ . Both operators arise as a reduction of the wave group to the boundary (or a transversal) and are quantizations of the billiard map,  $N(\lambda)$  globally and  $\mathcal{M}(\lambda)$  microlocally at a closed orbit  $\gamma$ . (An earlier reduction also occurs in [MM, P, P3] but in a rather different way. ) In applications to boundary problems, they are quite similar, as evidence by the comparisons of (6.27) with Proposition 7.1 and of (6.30) with (7.4).

But there do exist rather important differences. The operator  $N(\lambda)$  is not a standard Fourier integral operator on  $\partial\Omega$ , while  $\mathcal{M}(\lambda)$  is one. The diagonal singularities of  $N(\lambda)$  require a complicated regularization procedure. Moreover it is a global invariant of  $\partial\Omega$ , not a microlocal one at  $\gamma$ ; On the other hand,  $N(\lambda)$  is just the restriction of a canonically defined free Green's kernel to  $\partial\Omega \times \partial\Omega$ , and thus is an elementary and computable object. By comparison,  $\mathcal{M}(\lambda)$  must be constructed by some kind of parametrix method. It was precisely the complexity of microlocal parametrices for the wave group, even at periodic reflecting rays, which motivated our turning to the Balian-Bloch approach.

**7.4. Stationary phase expansion.** After regularizing the traces, we end up with oscillatory integrals in the standard sense and obtain expansions by applying stationary phase. The amplitudes and phases are canonical, since we began with canonical amplitudes and phases and since the regularization procedure is essentially the same for all domains. So most of the complexity of the expansion is due to the stationary phase method.

The key point in inverse spectral theory is to identify data in the stationary phase term of order  $k^{-j}$  which represents ‘new data’ not contained in the previous terms. Terms of the coefficient of  $k^{-j}$  which contain the maximum number of derivatives of the phase are the most important ones. Thus we face the combinatorial problem of locating such terms in the stationary phase expansion. The Feynman diagram method of assigning labelled graphs to each term in the expansion proves to be very effective for this purpose.

Consider a general oscillatory integral  $Z_k = \int_{\mathbb{R}^n} a(x)e^{ikS(x)}dx$  where  $a \in C_0^\infty(\mathbb{R}^n)$  and where  $S$  has a unique critical point in *suppa* at 0. Let us write  $H$  for the Hessian of  $S$  at 0. The stationary phase expansion takes the form:

$$Z_k = \left(\frac{2\pi}{k}\right)^{n/2} \frac{e^{i\pi \operatorname{sgn}(H)/4}}{\sqrt{|\det H|}} e^{ikS(0)} Z_k^{h\ell},$$

where  $Z_k^{h\ell} = \sum_{j=0}^\infty k^{-j} \left\{ \sum_{(\Gamma, \ell): \chi_{\Gamma'}=j} \frac{I_\ell(\Gamma)}{S(\Gamma)} \right\}.$

Here, the sum runs over the set  $\mathcal{G}_{V,I}$  of labelled graphs  $(\Gamma, \ell)$  with  $V$  closed vertices of valency  $\geq 3$  (each corresponding to the phase), with one open vertex (corresponding to the amplitude), and with  $I$  edges. Further, the graph  $\Gamma'$  is defined to be  $\Gamma$  minus the open vertex, and  $\chi_{\Gamma'} = V - I$  equals its Euler characteristic. We note that there are only finitely many graphs for each  $\chi$  because the valency condition forces  $I \geq 3/2V$ . Thus,  $V \leq 2j, I \leq 3j$ .

The function  $\ell$  ‘labels’ each end of each edge of  $\Gamma$  with an index  $j \in \{1, \dots, n\}$ . Also,  $S(\Gamma)$  denotes the order of the automorphism group of  $\Gamma$ , and  $I_\ell(\Gamma)$  denotes the ‘Feynman amplitude’ associated to  $(\Gamma, \ell)$ . By definition,  $I_\ell(\Gamma)$  is obtained by the following rule: To each edge with end labels  $j, k$  one assigns a factor of  $\frac{1}{ik} h^{jk}$  where  $H^{-1} = (h^{jk})$ . To each closed vertex one assigns a factor of  $ik \frac{\partial^\nu S(0)}{\partial x^{i_1} \dots \partial x^{i_\nu}}$  where  $\nu$  is the valency of the vertex and  $i_1 \dots, i_\nu$  at the index labels of the edge ends incident on the vertex. To the open vertex, one assigns the factor  $\frac{\partial^\nu a(0)}{\partial x^{i_1} \dots \partial x^{i_\nu}}$ , where  $\nu$  is its valence. Then  $I_\ell(\Gamma)$  is the product of all these factors. To the empty graph one assigns the amplitude 1. In summing over  $(\Gamma, \ell)$  with a fixed graph  $\Gamma$ , one sums the product of all the factors as the indices run over  $\{1, \dots, n\}$ .

7.4.1. *The data  $f_\pm^{2j}(0)$ .* An analysis of the diagrams and amplitudes shows that the  $j$ th even Taylor coefficients  $f_\pm^{(2j)}(0)$  of the boundary defining functions appear first in the  $k^{-j+1}$  term.

When the domain has one symmetry axis, which we visualize as an up/down symmetry, the terms with this data have the form

$$2rL(h^{11})^j f^{(2j)}(0) + \dots,$$

where  $\dots$  refers to terms with  $\leq 2j - 1$  derivatives.

When the domain has two symmetries, a left/right symmetry in addition to an up/down symmetry, the odd Taylor coefficients vanish and we see immediately that the even Taylor coefficients can be determined inductively from the wave trace invariants. This gives a new proof that analytic domains with two symmetries can be determined from the wave trace invariants at a bouncing ball orbit which is one of the symmetry axes.

This does not quite prove that such domains are spectrally determined among other analytic domains with two symmetries, since the length of the bouncing ball orbit must be known in order to obtain the wave invariants. This length is a spectral

invariant if the domains are additionally convex [Gh]. For non-convex domains we needed to add as an assumption that the bouncing ball symmetry axis had a fixed length  $L$  in [Z1].

7.4.2. *The data  $f_{\pm}^{(2j-1)}(\mathbf{0})$ .* When we do not assume a left/right symmetry, the odd Taylor coefficients are non-zero in general, and the problem arises whether there is sufficient information in the wave invariants to determine all of the even and odd Taylor coefficients of the boundary defining function (or curvature function) of a domain with one symmetry. We assume the axis of symmetry is a bouncing ball orbit whose orientation is reversed by the symmetry.

An analysis of the diagrams and amplitudes for the odd Taylor coefficients  $f_{\pm}^{(2j-1)}(0)$  show that they appear first in the term of order  $k^{-j+1}$ . It turns out that five diagrams contain this data, but the amplitudes of three automatically vanish. They two amplitudes have the following forms:

- (i)  $(h_{\pm}^{pp})^{j-1} h_{\pm}^{qq} h_{\pm}^{pq} f^{(2j-1)}(0) f^{(3)}(0)$ .
- (ii)  $(h_{\pm}^{pp})^{j-2} (h_{\pm}^{pq})^3 f^{(2j-1)}(0) f^{(3)}(0)$ .

To decouple them, we need to analyze the behavior of power sums of columns the Hessian matrix elements. In [Z5] we proved that cubic column sums are linearly independent from linear ones as  $r \rightarrow \infty$ .

**7.5. Positive results for analytic domains and metrics.** We now review the proof in [Z5] that the Taylor coefficients  $f^{2j-1}(0), f^{2j}(0)$  can be determined from the wave invariants  $B_{\gamma^r, j}$  as  $r$  varies over  $r = 1, 2, 3, \dots$

It suffices to separately determine the two terms

$$(7.15) \quad 2(h_{2r}^{11})^2 \left\{ f^{(2j)}(0) + \frac{1}{2-2\cos\alpha/2} f^{(3)}(0) f^{(2j-1)}(0) \right\},$$

$$\text{and } \left\{ \sum_{q=1}^{2r} (h_{2r}^{1q})^3 \right\} f^{(3)}(0) f^{(2j-1)}(0).$$

It is easy to see that the terms decouple as  $r$  varies if and only if the cubic sums  $\sum_{q=1}^{2r} (h_{2r}^{1q})^3$  are non-constant in  $r = 1, 2, 3, \dots$ . By the explicit calculation in [Z2], we have:

$$\sum_{q=1}^{2r} (h^{pq})^3 = 2^r \sum_{k_1, k_2=0}^{2r} \frac{1}{(\cosh\alpha/2 + \cos\frac{k_1\pi}{r})(\cosh\alpha/2 + \cos\frac{k_2\pi}{r})(\cos\alpha/2 + \cosh\frac{(k_1+k_2)\pi}{r})}.$$

It is obvious that the sum is strictly increasing as  $r$  varies over even integers.

This independence of the linear and cubic sums permits us to use an inductive argument when the domain has the given symmetry. From the  $j = 0$  term we determine  $f''(0)$ . Indeed,  $(1 - Lf^{(2)}(0) = \cos(h)\alpha/2$  and  $\alpha$  is a wave trace invariant. From the  $j = 2$  term we recover  $f^3(0), f^4(0)$ . The induction hypothesis is then that the Taylor polynomial of  $f$  of degree  $2j - 2$  has been recovered by the  $j - 1$ st stage. By the decoupling argument we can determine  $f^{2j}(0), f^{2j-1}(0)$ . (Strictly speaking, there is the minor annoyance of the factor of  $f^{(3)}(0)$ , which is resolved in [Z5]).

7.5.1. *Final comments on the Balian-Bloch invariants.* It would of course be desirable to remove all the symmetry assumptions, if possible. Thus one would need to recover the Taylor coefficients  $f_{\pm}^{(2j)}(0), f_{\pm}^{(2j-1)}(0)$  when  $f_+ \neq f_-$  from the wave invariants. The problem is that the induction does not work because the term  $R_{2r}(j^{2j-2} f(0))$ , which we do not need to know in the symmetric case, might have a different dependence on the Taylor coefficients of  $f_{\pm}$  than the ‘principal

term', i.e. the one with the highest derivatives. The latter is basically a sum of the coefficients of  $f_{\pm}$  while  $R_{2r}(j^{2j-2}f(0))$  could involve any symmetric polynomial in these coefficients. At the present time, we do not even know whether the set  $\{f_+^{(2)}(0), f_-^{(2)}(0)\}$  of second derivatives can be determined from the second wave coefficient.

Some hope has been provided by computer calculations of C. Hillar [Hil]. The results suggest that Hessian column power sums with different powers are linearly independent as  $r \rightarrow \infty$ . This would give quite a large supply of Taylor coefficients.

## 8. Surfaces and domains with integrable dynamics

As mentioned above, Birkhoff normal forms at a periodic orbit (on both the classical and quantum level) are approximations to the Hamiltonian by an integrable one. When the Hamiltonian is completely integrable, its dynamics and Birkhoff normal forms are very special and the metrics might be rather spectrally rigid, at least in low dimensions. This is the case with flat metrics [Ku3]. In this section, we consider some model inverse spectral problems on two-dimensional surfaces or domains with integrable dynamics.

Let us recall that the geodesic flow of an  $n$ -dimensional Riemannian manifold  $(M, g)$  is *completely integrable* if it commutes with a Hamiltonian action of  $\mathbb{R}^n$  on  $T^*M - 0$  ( $n = \dim M$ ). That is, the metric Hamiltonian  $|\xi|_g$  satisfies

$$\{|\xi|_g, p_j\} = 0 = \{p_i, p_j\}, \quad i, j = 1, \dots, n$$

where  $p_j : T^*M - 0 \rightarrow \mathbb{R}$  are homogeneous of degree one and are independent in the sense that

$$dp_1 \wedge dp_2 \wedge \dots \wedge dp_n \neq 0 \text{ on a dense open subset } U \subset T^*M.$$

The orbits of an  $\mathbb{R}^n$  action give a (usually singular) foliation of  $S_g^*M$  (the unit sphere bundle for the metric), called the Liouville foliation, by affine manifolds of the form  $\mathbb{R}^n \cdot (x, \xi) \equiv \mathbb{R}^n / \Gamma$  where  $\Gamma$  is the isotropy group at  $(x, \xi)$ . If it is a lattice of full rank, the orbit is a torus  $T^n$  of dimension  $n$ . If it has less than full rank, the orbit type is  $T^k \times \mathbb{R}^{n-k}$ . The isotropy group might have the form  $\mathbb{R}^k \times \mathbb{Z}^{n-k}$  in which case the orbit becomes a singular torus of dimension  $n - k$ . The orbits which contain periodic geodesics are sometimes called 'periodic tori', and they furnish the components  $\mathcal{T}$  in the trace formula.

In the case of bounded plane domains, complete integrability of the billiard map  $\beta$  means that  $B^*\partial\Omega$  is foliated by invariant curves. The well-known conjecture of Birkhoff is that ellipses are the only example of compact smooth Euclidean plane domains with integrable billiard map. There are more examples if curved metrics are allowed, see Popov-Topalov [PT].

**8.1. Trace formulae for integrable systems.** So far, we have mainly considered the trace of the wave group around non-degenerate periodic orbits. For integrable systems, the periodic orbits usually come in families filling out invariant tori. We now consider the appropriate notions of non-degeneracy in this context.

We will always assume that the closed geodesics come in clean families in the following sense:

**DEFINITION 8.1.** *A metric  $g$  on a compact manifold  $M$  will be said to have a simple clean length spectrum if the length function  $L_g$  on the loop space  $\text{Map}(S^1, M)$*

is a Bott-Morse function which takes distinct values at distinct components of its critical set,  $Crit(L_g)$ .

The term Bott-Morse means that each component of  $Crit(L_g)$  is a manifold, whose tangent space is the kernel of  $dL_g$ . Equivalently, each component is a clean fixed point set for  $G_g^t$ . One needs the clean (Bott-Morse) condition to get a nice wave trace expansion and one needs the simple length spectrum condition to determine geometric information from the expansion.

Under this assumption, the trace of its wave group has the form:

$$(8.1) \quad Tr e^{it\sqrt{\Delta_g}} = e_0(t) + \sum_{\mathcal{T}} e_{\mathcal{T}}(t)$$

where the singular term  $e_0(t) = C_n Vol(M, g) (t + i0)^{-n} + \dots$  at  $t = 0$  is the same as in the non-degenerate case, where  $\{\mathcal{T}\}$  runs over the critical point components of  $L_g$ , and where

$$e_{\mathcal{T}} = c_{\mathcal{T}, d_{\mathcal{T}}}(t - L_{\mathcal{T}} + i0)^{-d_{\mathcal{T}}/2} + c_{\mathcal{T}, d_{\mathcal{T}}-1}(t - L_{\mathcal{T}} + i0)^{-(d_{\mathcal{T}}/2+1)} + \dots, \quad d_{\mathcal{T}} = \dim \mathcal{T}.$$

Here,  $d_{\mathcal{T}}$  is the dimension of the symplectic cone formed by the family of closed geodesics within  $T_g^*M$  and  $L_{\mathcal{T}}$  is the common length of the closed geodesic in  $\mathcal{T}$ . For instance, in the non-degenerate case,  $\mathcal{T} = \mathbb{R}_+\gamma$  is the symplectic cone generated by  $\gamma \subset S_g^*M$  and  $d_{\mathcal{T}} = 2$ .

**8.2. Spectral determination of simple surfaces of revolution.** The simplest example of what can be done with integrable systems is given by ‘simple’ analytic surfaces of revolution. We first review the proof in [Z2] that they are spectrally determined among other simple analytic surfaces of revolution and then sketch a result observed independently by the author and G. Forni and by K.F. Siburg regarding their spectral determination among all metrics on  $S^2$ .

The precise class of metrics we consider are those metrics  $g$  on  $S^2$  which belong to the class  $\mathcal{R}^*$  of real analytic, rotationally invariant metrics on  $S^2$  with simple length spectrum in the above sense and satisfying the following ‘simplicity’ condition

- $g = dr^2 + a(r)^2 d\theta^2$
- $\exists! r_0 : a'(r_0) = 0;$
- The Poincare map  $\mathcal{P}_0$  of  $r = r_0$  is elliptic of twist type

Convex analytic surfaces of revolution are examples, but of course there are others. The unique isolated closed geodesic (at distance  $r_0$ ) is an elliptic orbit. Peanut surfaces are obviously non-simple, since they possess three isolated closed geodesics of which one is hyperbolic. The standard round metric and other Zoll metrics of revolution are non-simple since the Poincare maps are not twist maps.

The following results shows that one can solve the inverse spectral problem in the class of analytic simple surfaces of revolution.

**THEOREM 8.2.** ([Z2]) *Suppose that  $g_1, g_2$  are two real analytic metrics on  $S^2$  such that  $(S^2, g_i)$  are simple surfaces of revolution with simple length spectra. Then  $Sp(\Delta_{g_1}) = Sp(\Delta_{g_2})$  implies  $g_1 = g_2$ .*

The proof is based on quantum Birkhoff normal forms for the Laplacian  $\Delta$ . But the special feature of simple surfaces of revolution is that there exists a global

Birkhoff normal form as well as local ones around the critical closed orbit or the invariant tori. This is because  $\Delta$  is a toric integrable Laplacian in the following sense: there exist commuting first order pseudo-differential operators  $\hat{I}_1, \hat{I}_2$  such that:

- the joint spectrum is integral, i.e.  $Sp(\mathcal{I}) \subset \mathbb{Z}^2 \cap \Gamma + \{\mu\}$  where  $\Gamma$  is the cone  $I_2 \geq |I_1|$  in  $\mathbb{R}^2$ .
- The square root of  $\Delta$  is a first order polyhomogeneous function  $\sqrt{\Delta} = \hat{H}(\hat{I}_1, \hat{I}_2)$  of the action operators.

By polyhomogeneous, we mean that  $\hat{H}$  has an asymptotic expansion in homogeneous functions of the form:

$$\hat{H} \sim H_1 + H_o + H_{-1} + \dots, \quad H_j(rI) = r^j H_j(I).$$

The principal symbols  $I_j$  of the  $\hat{I}_j$ 's generate a classical Hamiltonian torus action on  $T^*S^2 - 0$ . Analysis of the normal form shows that  $H_o = 0$ .

It follows that

$$Sp(\sqrt{\Delta_g}) = \{\hat{H}(N + \mu) : N \in \mathbb{Z}^2 \cap \Gamma_o\},$$

where the eigenvalues have expansions

$$\lambda_N \sim H_1(N + \mu) + H_{-1}(N + \mu) + \dots$$

**THEOREM 8.3.** ([Z2]) *Let  $(S^2, g)$  be an analytic simple surface of revolution with simple length spectrum. Then the normal form  $\hat{H}(\xi_1, \xi_2)$  is a spectral invariant.*

The normal form and the proof are very different from the non-degenerate case in [G, Z3, Z4], although the philosophy of the proof is similar.

To complete the proof of Theorem 8.2, we need to show that  $\hat{H}$  determines a metric in  $\mathcal{R}$ . As in the bounded domain case outline above, the crucial point is to calculate the normal form invariants. It turns out to be sufficient to calculate  $H_1 = H$  and  $H_{-1}$  in terms of the metric (i.e. in terms of  $a(r)$ ) and then to invert the expressions to determine  $a(r)$ .

The method given in [Z2] for calculating  $H$  and  $H_{-1}$  was to study the spectral asymptotics of  $\sqrt{\Delta} = \hat{H}(\hat{I}_1, \hat{I}_2)$  along ‘rays of representations’ of the quantum torus action, i.e. along multiples of a given lattice point  $(n_o, k_o)$ . The lattice points  $(n_o, k_o + \frac{1}{2})$  parametrize tori  $T_{n_o, k_o}$  satisfying so-called Bohr-Sommerfeld quantization conditions, which imply that one can construct associated joint eigenfunctions  $\phi_{n_o, k_o}$  of  $(\hat{I}_1, \hat{I}_2)$  by the WKB method. This is reminiscent of the quasi-mode method mentioned above around non-degenerate closed geodesics, but there we studied the conjugation to normal form rather than the asymptotics of quasi-modes. Here, the existence of a global torus action makes the quasi-modes easier to study, and indeed the  $\{\phi_{n, k}\}$  are actually modes (eigenfunctions) of  $\Delta$  with complete asymptotic expansions along rays. By studying the eigenvalue problem as  $|(n, k)| \rightarrow \infty$  we determine the  $H_{-j}$ 's.

Once it is known that the global Birkhoff normal for  $\hat{H}$  is a spectral invariant, it follows that for each  $n_o \in \mathbb{Z}$ , the function  $\hat{H}(n_o, \hat{I}_2)$  is a known function of the variable  $I := \hat{I}_2$ . Its principal symbol  $H_{n_o}(I) := H_1(n_o, I)$  is then a known function and its inverse function

$$I_{n_o}(E) = \int_{r_-(E)}^{r_+(E)} \sqrt{E - \frac{1}{a(r)^2}} dr$$

is also known. Here,  $r_{\pm}(E)$  are the upper and lower values of the radius of the projection of the torus of energy  $E$  and angular momentum  $n_0$  to  $S^2$ . We may write the integral in the form

$$\int_{\mathbb{R}} (E - x)_+^{\frac{1}{2}} d\mu(x)$$

where  $\mu$  is the distribution function  $\mu(x) := |\{r : \frac{1}{a(r)^2} \leq x\}|$  of  $\frac{1}{a^2}$ , with  $|\cdot|$  the Lebesgue measure. This Abel transform is invertible and hence

$$d\mu(x) = \sum_{r: \frac{1}{a(r)^2} = x} \left| \frac{d}{dr} \frac{1}{a(r)^2} \right|^{-1} dx$$

and therefore

$$J(x) := \sum_{r: a(r) = x} \frac{1}{|a'(r)|}$$

are spectral invariants. By the simplicity assumption on  $a$ , there are just two solutions of  $a(r) = x$ ; the smaller will be written  $r_-(x)$  and the larger,  $r_+(x)$ . Thus, the function

$$J(x) = \frac{1}{|a'(r_-(x))|} + \frac{1}{|a'(r_+(x))|}$$

is a spectral invariant.

By studying  $H_{-1}$ , we find in a somewhat similar way that

$$K(x) = |a'(r_-(x))| + |a'(r_+(x))|$$

is a spectral invariant. It follows that we can determine  $a'(r_+(x))$  and  $a'(r_-(x))$ .

Since both metrics  $g_1$  and  $g_2$  are assumed to belong to  $\mathcal{R}^*$ , they are determined by their respective functions  $a_j(r)$ . We conclude that  $a_1 = a_2$  and hence  $g_1 = g_2$ .

**8.3. Unconditional spectral determination.** We now ask whether we can remove the assumption that  $g_2 \in \mathcal{R}^*$  in Theorem 8.2? The question is, if  $\text{Spec}\Delta_g = \text{Spec}\Delta_h$  and  $g \in \mathcal{R}^*$ , then is  $h \in \mathcal{R}^*$ ? An affirmative answer would give a large class of metrics which are spectrally determined. To the author's knowledge, the only metric on  $S^2$  known to be spectrally determined is the canonical round one.

We cannot answer this question, but we will give a partial result suggesting that it is true. The following theorem was stated in [Z7] and was worked out in a conversation with G. Forni in 1997. A similar but somewhat stronger conclusion was drawn by K. F. Siburg (Theorem 5.2 of [S2]), under the stronger hypothesis of an isospectral (or length-spectral) deformation. After sketching our proof, we will also sketch his, which has other applications.

**THEOREM 8.4.** *Let  $g \in \mathcal{R}^*$  and let  $h$  be any metric on  $S^2$  with simple clean length spectrum for which  $\text{Spec}\Delta_g = \text{Spec}\Delta_h$ . Then  $h$  has the following properties:*

- (i) *It has just one isolated non-degenerate closed geodesic  $\gamma_h$  (up to orientation); all other closed geodesics come in one-parameter families lying on invariant tori in  $S_h^*S^2$ ;*
- (ii) *The Birkhoff normal form of  $G_h^t$  at  $\gamma_h$  is identical to that of  $G_g^t$  at its unique non-degenerate closed orbit. Hence it is convergent.*
- (iii) *the geodesic flow  $G_h^t$  of  $h$  is  $C^0$ -integrable. That is,  $S_h^*S^2$  has a  $C^0$ -foliation by 2-tori invariant under  $G_h^t$ .*

If we knew in (ii) that the Birkhoff transformation conjugating  $G_h^t$  to its Birkhoff normal form was convergent, then it would follow that  $G_h^t$  is completely integrable with global action-angle variables, and that it would commute with a Hamiltonian torus action. We conjecture that this is the case. Statement (iii) shows that it is at least integrable in the  $C^0$  sense. At the present time, metrics on  $S^2$  whose geodesic flows commute with Hamiltonian torus actions have not been classified. For a result which classifies such metrics on the torus (they must be flat), see [LS].

8.3.1. *Wave invariants for  $g \in \mathcal{R}^*$ .* The proof is based on a study of the wave trace formula in this setting.

PROPOSITION 8.5. *Suppose that  $g \in \mathcal{R}^*$ . Then the trace of its wave group has the form:*

$$Tr e^{it\sqrt{\Delta_g}} = e_0(t) + e_{\gamma_g}(t) + \sum_{\mathcal{T}} e_{\mathcal{T}}(t)$$

where

$$e_0(t) = C_n \text{area}(M, g)(t + i0)^{-n} + \dots$$

is singular only at  $t = 0$ , where

$$e_{\gamma_g}(t) = c_{\gamma}(t - L_{\gamma} + i0)^{-1} + a_{\gamma_0} \log(t - L_{\gamma} + i0) + \dots$$

and where

$$e_{\mathcal{T}} = c_{\mathcal{T}}(t - L_{\mathcal{T}} + i0)^{-3/2} + \dots$$

Suppose now that  $h$  is any other metric with  $\text{Spec } \Delta_h = \text{Spec } \Delta_g$  and with simple length spectrum. Then the wave trace of  $h$  has precisely the same singularities as the wave trace of  $g$ . Since there is only one singularity of the order  $(t + i0)^{-1}$ , there can exist only one non-degenerate closed geodesic, proving (i). By Guillemin’s inverse result, the Birkhoff normal form of the metric and Poincaré map for  $\gamma_h$  is the same as for  $\gamma_g$ . In particular, the Poincaré map  $\mathcal{P}_h$  of  $\gamma_h$  is elliptic of twist type. From the fact that all other critical components have the singularity of a three-dimensional cone, it follows that in  $S_h^*S^2$  the other closed geodesics come in one-parameter families. They sweep out a surface foliated by circles, which can only be a two dimensional torus, proving (ii)

It is the third statement (iii) which requires a new idea. So far we only know that periodic orbits lie on invariant tori, but we do not know what lies between these tori. Aubry-Mather theory will now close the gaps.

We recall that Aubry-Mather theory is concerned with an area-preserving diffeomorphism  $\phi$  of an annulus  $A = S^1 \times (a, b)$  [KH, MF, Ka, M, S1, S2]. Let  $\tilde{\phi}$  denote a lift to  $\mathbb{R} \times (a, b)$  with  $\tilde{\phi}(x + 1, y) = \tilde{\phi}(x, y) + (1, 0)$ . The map  $\phi$  is called a *monotone twist mapping* if it preserves the orientation of  $A$ , if it preserves the boundary components and if the lift  $\tilde{\phi}(x_0, y_0) = (x_1, y_1)$  satisfies:

- The twist condition:  $\frac{\partial x_1}{\partial y_0} > 0$ ;
- The exactness condition:  $y_1 dx_1 - y_0 dx_0 = dh(x_0, x_1)$ .

If  $a, b$  are finite,  $\tilde{\phi}$  extends continuously to the boundary as a pair of ‘rotations’:

$$\tilde{\phi}(x, a) = (x + \omega_-, a), ; \quad \tilde{\phi}(x, b) = (x + \omega_+, b).$$

The interval  $(\omega_-, \omega_+)$  is called the *twist interval* of  $\phi$ .



Let  $\{(x_i, y_i)\}$  be an orbit of  $\tilde{\phi}$ . Its *rotation number* is defined to be

$$\lim_{|i| \rightarrow \infty} \frac{x_i - x_0}{i}.$$

A curve  $C \subset A$  is called an *invariant circle* if it is an invariant set which is homeomorphic to the circle and which separates boundary components. According to Birkhoff’s invariant circle theorem, an invariant circle is a Lipschitz graph over the factor  $S^1$  of  $A$ . Any invariant circle has a well-defined rotation number (the common rotation number of orbits in the circle) and the rotation number belongs to the twist interval.

An orbit  $\{(x_i, y_i)\}$  is determined by the sequence  $\{x_i\}$  of its  $x$ -coordinates. It is called *minimal* if every finite segment is action-minimizing with fixed endpoints:

$$\sum_{i=k}^{n-1} h(x_i, x_{i+1}) \leq \sum_{i=k}^{n-1} h(\xi_i, \xi_{i+1}), \quad \forall (\xi_k, \dots, \xi_n) \text{ with } \xi_k = x_k, \xi_n = x_n.$$

The corresponding orbit  $(x_i, y_i)$  is called a *minimal orbit*. The Aubrey-Mather theorem states (cf. [Ka], Theorem 1):

*A monotone twist map possesses minimal orbits for each rotation number  $\omega \in (\omega_-, \omega_+)$  in its twist interval. Every minimal orbit lies on a Lipschitz graph over the  $x$ -axis. For each rational rotation number  $\omega = \frac{p}{q}$ , there exists a periodic minimal orbit of rotation number  $\frac{p}{q}$ . When  $\omega$  is irrational, there exists either an invariant circle with rotation number  $\omega$ , or an invariant Cantor set  $E$ .*

The theorem also describes three possible orbit types in both the rational or irrational case.

We now return to our problem on simple surfaces of revolution. We fix local transverse discs (Poincaré sections)  $S_g$ , resp.  $S_h$  to the geodesic flows  $G_g^t$ , resp.  $G_h^t$  at the orbits  $\gamma_g$ , resp.  $\gamma_h$ . Concretely, the transversals can be taken to be small variations of  $\gamma'_g$ , resp.  $\gamma'_h$  moved up and down a small orthogonal geodesic arc to  $\gamma'_g$ , resp.  $\gamma'_h$ . Since  $\gamma_g$  is non-degenerate elliptic, its Poincaré map  $\mathcal{P}_g$  defines area-preserving map of the symplectic disc  $S_g$  with a non-degenerate elliptic fixed point corresponding to  $\gamma_g$ . Since the Birkhoff normal forms of  $\mathcal{P}_g$  and  $\mathcal{P}_h$  are the same,  $\mathcal{P}_h$  defines the same kind of map of  $S_h$ . To obtain a twist map of an annulus, we puncture out the fixed point of  $\mathcal{P}_g$ , resp.  $\mathcal{P}_h$ . We define the rotation angle  $\omega_0$  corresponding to this orbit by continuity from nearby orbits, which can be read off from the Birkhoff normal form. Indeed, the Poincaré maps have the form

$$\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} \cos 2\pi\theta & -\sin 2\pi\theta \\ \sin 2\pi\theta & \cos 2\pi\theta \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + O(|x|^2 + |y|^2),$$

with  $\theta \sim \omega_0 + \beta(x^2 + y^2) + \dots$ , as  $x^2 + y^2 \rightarrow 0$ . For further discussion of area-preserving maps around elliptic fixed points and twist maps, see [S2].

The foliation of  $S^*S_g \setminus \gamma_h$  by 2-tori intersects  $S_g$  in a foliation by invariant circles converging to the fixed point. Circles with rational rotation numbers contain only periodic orbits, and conversely all periodic orbits belong to invariant circles with rational rotation numbers.

In the case of  $h$ , we know that once  $\gamma_g$  is punctured out, all periodic orbits come in two-tori which project to invariant curves in  $S_h$  which are diffeomorphic to circles. The Birkhoff normal form shows that they are invariant circles in the

sense that they are also homotopically non-trivial in the punctured  $S_h$  (indeed, the action-angle variables are essentially polar coordinates on this disc).

Let  $C \subset S_h$  be an invariant circle for  $\mathcal{P}_h$ . Set:

- $I(C)$  = area enclosed by  $C$ .
- $C_I$  = invariant circle enclosing an area  $I$ . (It is clearly unique).
- $\omega_I$  = rotation number of  $\mathcal{P}_H|_{C_I}$ .

Since  $\mathcal{P}_h$  is a twist map, the rotation number  $\omega_I$  is a monotone increasing function of  $I$ .

Let  $I_+$  denote the area of  $S_h$  with respect to the symplectic form. By shrinking  $S_h$  we may assume that the boundary of  $S_h$  is a periodic circle  $C_{I_+}$ .

Since the origin is a fixed point of  $\mathcal{P}_h$  and since the rotation number is increasing, it is clear that the set of rotation numbers lie in the interval  $[0, \omega_{I_+}]$ . It follows by the Aubry-Mather theorem that every rational number  $p/q \in [0, \omega_{I_+}]$  is the rotation number of a periodic circle  $C_I \subset S_h$ .

LEMMA 8.6. *Let  $\alpha \in [0, \omega_{I_+}]$ . Then there exists an invariant circle for  $\mathcal{P}_h$  of rotation number  $\alpha$ .*

**Proof** This follows from Corollary 6.1 of [Ka], which shows that as soon as Birkhoff periodic points of all rational rotation numbers are constructed, then there exist orbits of each irrational rotation number which are dense in an invariant circle. Alternatively one could let  $p_n/q_n \rightarrow \alpha$  and let  $C_{p_n/q_n}$  be the corresponding periodic circles. Each  $C_{p_n/q_n}$  is a Lipschitz circle and the sequence of these circles tends monotonically to a limit circle. It is Lipschitz and its rotation number is  $\alpha$ .

To complete the proof, we make the observation:

LEMMA 8.7.  *$S_h$  is foliated in the  $C^0$  sense by invariant circles for  $\mathcal{P}_h$ .*

**Proof:** If not there exists an annulus  $A \subset S_h$  with boundary consisting of two invariant circles and containing no invariant circles in its interior. But by the Aubrey-Mather theorem, there must exist a periodic point in  $A$ . Since the periodic points come in circles, there must exist a periodic circle, contradicting the non-existence of invariant circles in  $A$ . QED

This completes the proof of Theorem 8.4.

8.3.2. *Isospectral class of an ellipse.* Instead of surfaces in  $\mathcal{R}^*$ , one can apply this reasoning to the Dirichlet (or Neumann) problem for an ellipse  $E_{a,b} = \{(x, y) : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\}$ . It is well-known that ellipses have integrable billiards.

The ellipse has three distinguished periodic billiard orbits:

- The bouncing ball orbit along the minor axis, which is a non-degenerate elliptic orbit;
- The bouncing ball orbit along the major axis, which is a non-degenerate hyperbolic orbit;
- Its boundary.

All other periodic orbits come in one-parameter families. The existence of a hyperbolic orbit means that the billiard flow is very different from geodesic flows of metrics in  $\mathcal{R}^*$ .

The wave trace formula shows that any domain  $\Omega$  with  $Spec(\Omega) = Spec(E_{a,b})$  has precisely one isolated elliptic orbit, one isolated hyperbolic orbit. The accumulation points in the length spectrum must be multiples of the perimeter of the domain (a spectral invariant), so the boundary must be a closed geodesic as well.

One can apply the twist map theory either to the boundary orbit or to the unique non-degenerate elliptic orbit with isolated length in the length spectrum. The argument above shows that there exists a  $C^0$  foliation by invariant circles at least near these two orbits.

**8.4. Marked length spectral rigidity of domains.** We now review a result due to K.F. Siburg [S1, S2] on isospectral deformations of integrable systems which has interesting applications to metrics in  $\mathcal{R}^*$  and to bounded plane domains.

The key invariant is the *mean minimal action*

$$(8.2) \quad \alpha : [\omega_-, \omega_+] \rightarrow \mathbb{R},$$

of a twist map  $\phi$ , which associates to a rotation number  $\omega$  in the ‘twist interval’ the mean action

$$(8.3) \quad \alpha(\omega) = - \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{i=-N}^N h(x_i, x_{i+1})$$

of a minimal orbit  $(q_i, \eta_i)$  of  $\phi$  of rotation number  $\omega$ . It is a strictly convex function which is differentiable at all irrational numbers. If  $\omega = p/q$ , then  $\alpha$  is differentiable at  $\omega$  if and only if there exists an invariant circle of rotation number  $p/q$  consisting entirely of periodic minimal orbits. If a monotone twist map possesses an invariant circle of rotation number  $\omega$ , then every orbit on the circle is minimal ([MF], Theorem 17.4). In the case of a bounded plane domain,  $h(q, q') = -|q - q'|$ .

It is observed by K. F. Siburg [S1] (Theorem 4.1) that the marked length spectrum is essentially the same invariant as the mean minimal action. The mean minimal action is therefore an isospectral deformation invariant. He used this to give a proof that there cannot exist isospectral deformations within  $\mathcal{R}^*$  (or within more general classes of deformations, see [S2], Theorem 4.5). Indeed, the mean minimal actions  $\alpha$  would all be the same. Hence if that  $\alpha_0$  of the original surface is differentiable at all rationals, so are they all. But this implies that they all have invariant circles of rational rotation numbers. By taking limits, one obtains invariant circles of all rotation numbers. (We remark that the trace formula already shows that there existed invariant circles of all rational rotation numbers).

Siburg further connects the mean minimal to the Melrose-Marvizi invariants: Let  $\alpha^* : [-1, 1] \rightarrow \mathbb{R}$  denote its convex conjugate of  $\alpha$ . Then Siburg shows (loc. cit. p. 300) that the Melrose-Marvizi invariants are algebraically equivalent to the Taylor coefficients of  $(\alpha^*)^{2/3}$  at  $-1$ .

It appears that the only explicitly known mean minimal action is that of the disc  $D$ , where it is given by  $\alpha(\omega) = \frac{-1}{\pi} \sin \pi \omega$  ([S1]);  $\alpha$  is only smooth when  $\Omega = D$  ([S1], Theorem 4.6) It is likely that it is computable in the case of an ellipse (perhaps in terms of elliptic functions). Perhaps it can be proved that ellipses are the unique domains with these particular minimal action functions (which could be simpler than the long outstanding problem of proving that they are the unique integrable billiard systems). If so, this would prove that ellipses are spectrally rigid.

**PROBLEM 8.1.** *Is the map from curvature functions  $\kappa$  of convex plane domains to the mean minimal action  $\alpha$  of the associated convex domain injective or finitely many to one? at least near ellipses or under some additional analyticity or discrete symmetry condition?*

## References

- [AM] K. G. Andersson and R. B. Melrose, The propagation of singularities along gliding rays. *Invent. Math.* 41 (1977), no. 3, 197–232.
- [AKN] V. I. Arnol'd, V.V. Kozlov, and A. I. Neushtadt, *Mathematical aspects of classical and celestial mechanics. Dynamical systems, III*, pp. vii–xiv and 1–291, Encyclopaedia Math. Sci., 3, Springer, Berlin, 1993.
- [BB] V.M.Babic, V.S. Buldyrev: *Short-Wavelength Diffraction Theory*, Springer Series on Wave Phenomena 4, Springer-Verlag, New York (1991).
- [Ba] V. Baladi, Periodic orbits and dynamical spectra. *Ergodic Theory Dynam. Systems* 18 (1998), no. 2, 255–292.
- [BB1] R. Balian and C. Bloch, Distribution of eigenfrequencies for the wave equation in a finite domain I: three-dimensional problem with smooth boundary surface, *Ann. Phys.* 60 (1970), 401–447.
- [BB2] R. Balian and C. Bloch, Distribution of eigenfrequencies for the wave equation in a finite domain. III. Eigenfrequency density oscillations. *Ann. Physics* 69 (1972), 76–160.
- [B] V. Bangert, Geodesic rays, Busemann functions and monotone twist maps. *Calc. Var. Partial Differential Equations* 2 (1994), no. 1, 49–63.
- [BLW] A. Banyaga, R. de la Llave, and C. E. Wayne, Cohomology equations near hyperbolic points and geometric versions of Sternberg linearization theorem. *J. Geom. Anal.* 6 (1996), no. 4, 613–649 (1997).
- [BGR] C. Bardos, J. C. Guillot, and J. Ralston, La relation de Poisson pour l'équation des ondes dans un ouvert non-born, *Comm. Partial Differential Equations* 7 (1982), 905–958
- [BK] M. I. Belishev and Y. V. Kurylev, To the reconstruction of a Riemannian manifold via its spectral data (BC-method). *Comm. Partial Differential Equations* 17 (1992), no. 5-6, 767–804
- [Be] P. Bérard, Transplantation et isospectralité. *Math. Ann.* 292, 547–560 (1992)
- [Be2] P. Bérard, Transplantation et isospectralit. II. *J. London Math. Soc.* (2) 48 (1993), no. 3, 565–576.
- [Be3] P. Bérard, *Spectral geometry: direct and inverse problems*. With appendixes by Grard Besson, and by Bérard and Marcel Berger. *Lecture Notes in Mathematics*, 1207. Springer-Verlag, Berlin, 1986.
- [Ber] M. Berger, P. Gauduchon, and E. Mazet, *Le spectre d'une variété riemannienne*. *Lecture Notes in Mathematics*, Vol. 194 Springer-Verlag, Berlin-New York 1971
- [BCG] G. Besson, G. Courtois, and S. Gallot, Entropies et rigidités des espaces localement symtriques de courbure strictement ngative. *Geom. Funct. Anal.* 5 (1995), no. 5, 731–799.
- [Bia] M. Bialy, Convex billiards and a theorem by E. Hopf. *Math. Z.* 214 (1993), no. 1, 147–154
- [BKL] M. Blank, G. Keller, and C. Liverani, Ruelle-Perron-Frobenius spectrum for Anosov maps Authors: (arXiv preprint nlin.CD/0104031).
- [BJP] D. Borthwick, C. Judge, and P. A. Perry, Determinants of Laplacians and isopolar metrics on surfaces of infinite area. *Duke Math. J.* 118 (2003), no. 1, 61–102.
- [BG] T.P. Branson and P. Gilkey, The asymptotics of the Laplacian on a manifold with boundary. *Comm. Partial Differential Equations* 15 (1990), no. 2, 245–272.
- [BGKV] T. Branson, P. B. Gilkey, K. Kirsten and D. Vassilevich, Heat kernel asymptotics with mixed boundary conditions, *Nuclear Phys.* B563 (1999), 603–626.
- [BP] R. Brooks and P. A. Perry, Isophasal scattering manifolds in two dimensions. *Comm. Math. Phys.* 223 (2001), no. 3, 465–474.
- [BH] J. Brüning and E. Heintze, Spektrale Starrheit gewisser Drehflächen, *Math. Ann.* 269 (1984), 95–101.
- [Ch] J. Chazarain, Construction de la paramtrix du problme mixte hyperbolique pour l'equation des ondes. *C. R. Acad. Sci. Paris Sr. A-B* 276 (1973), A1213–A1215.
- [Ch2] J. Chazarain, Formule de Poisson pour les varits riemanniennes. *Invent. Math.* 24 (1974), 65–82
- [CdV] Y. Colin de Verdière, Sur les longueurs des trajectoires priodiques d'un billard, in: Dazord, Desolneux (eds.): *Géométrie symplectique et de contact*, Sem. Sud-Rhod. Gom. (1984), 122–139.
- [CdV2] Y. Colin de Verdière, Spectre du laplacien et longueurs des godsiques priodiques. *C. R. Acad. Sci. Paris Sr. A-B* 275 (1972), A805–A808.

- [CdV3] Y. Colin de Verdière, Spectre du laplacien et longueurs des godsiques priodiques. I, II. *Compositio Math.* 27 (1973), 83–106; *ibid.* 27 (1973), 159–184.
- [CdV4] Y. Colin de Verdière, Spectre conjoint d’opérateurs pseudo-différentiels qui commutent. II. Le cas intégrable. *Math. Z.* 171 (1980), no. 1, 51–73.
- [C] C. Croke, Rigidity for surfaces of nonpositive curvature. *Comment. Math. Helv.* 65 (1990), no. 1, 150–169.
- [C2] C. Croke, Rigidity and the distance between boundary points. *J. Differential Geom.* 33 (1991), no. 2, 445–464.
- [C3] C. Croke, Rigidity theorems in Riemannian geometry, to appear in *IMA Volume 137: Geometric Methods in Inverse Problems and PDE Control*. C.B. Croke, I. Lasiecka, G. Uhlmann, and M. S.Vogelius eds.
- [CFF] C. Croke, A. Fathi and J. Feldman, The marked length-spectrum of a surface of nonpositive curvature. *Topology* 31 (1992), no. 4, 847–855.
- [CS] C. B. Croke and V. Sharafutdinov, Spectral rigidity of a compact negatively curved manifold. *Topology* 37 (1998), no. 6, 1265–1273.
- [DS] C. Deninger and W. Singhof, A note on dynamical trace formulas. *Dynamical, spectral, and arithmetic zeta functions* (San Antonio, TX, 1999), 41–55, *Contemp. Math.*, 290, Amer. Math. Soc., Providence, RI, 2001.
- [D] H. Donnelly, On the wave equation asymptotics of a compact negatively curved surface. *Invent. Math.* 45 (1978), no. 2, 115–137.
- [DG] J.J.Duistermaat and V.Guillemin, The spectrum of positive elliptic operators and periodic bicharacteristics, *Inv.Math.* 24 (1975), 39-80.
- [Gh] M. Ghomi, Shortest periodic billiard trajectories in convex bodies, to appear in *Geom. Funct. Anal.*
- [Gor] C. Gordon, Survey of isospectral manifolds. *Handbook of differential geometry*, Vol. I, 747–778, North-Holland, Amsterdam, 2000.
- [Gor2] C. Gordon, CBMS Lecture Notes (to appear).
- [GorM] C. Gordon and Y. Mao, Geodesic conjugacies of two-step nilmanifolds. *Michigan Math. J.* 45 (1998), no. 3, 451–481.
- [GMS] C. Gordon, Y. Mao, and D. Schueth, Symplectic rigidity of geodesic flows on two-step nilmanifolds. *Ann. Sci. cole Norm. Sup.* (4) 30 (1997), no. 4, 417–427.
- [GWW] C. Gordon, D. Webb, and S. Wolpert, Isospectral plane domains and surfaces via Riemannian orbifolds. *Invent. Math.* 110 (1992), no. 1, 1–22.
- [G] V. Guillemin, Wave-trace invariants. *Duke Math. J.* 83 (1996), no. 2, 287–352.
- [G2] V. Guillemin, Lectures on spectral theory of elliptic operators. *Duke Math. J.* 44 (1977), no. 3, 485–517.
- [G3] V. Guillemin, Wave-trace invariants and a theorem of Zelditch. *Internat. Math. Res. Notices* 1993, no. 12, 303–308
- [GK] V. Guillemin and D. Kazhdan, Some inverse spectral results for negatively curved 2-manifolds. *Topology* 19 (1980), no. 3, 301–312.
- [GM] V. Guillemin and R. B. Melrose, The Poisson summation formula for manifolds with boundary, *Adv. Math.* 32 (1979), 204–232.
- [GM2] V. Guillemin and R. B. Melrose, The wave equation on manifolds with boundary, *Colloque Boutet de Monvel* (2003).
- [GM3] V. Guillemin and R. B. Melrose, An inverse spectral result for elliptical regions in  $R^2$ . *Adv. in Math.* 32 (1979), no. 2, 128–148.
- [GM4] V. Guillemin and R. B. Melrose, A cohomological invariant of discrete dynamical systems. E. B. Christoffel (Aachen/Monschau, 1979), pp. 672–679, Birkhuser, Basel-Boston, Mass., 1981
- [Gutz] M. C. Gutzwiller, *J. Math. Phys.* 12, 343 (1971).
- [H] U. Hamenstädt, Cocycles, symplectic structures and intersection. *Geom. Funct. Anal.* 9 (1999), no. 1, 90–140.
- [H2] U. Hamenstädt, Time-preserving conjugacies of geodesic flows. *Ergodic Theory Dynam. Systems* 12 (1992), no. 1, 67–74.
- [HZel] A. Hassell and S. Zelditch, Quantum ergodicity of boundary values of eigenfunctions (2003), arxiv preprint math.SP/0211140.
- [HZel2] A. Hassell and S. Zelditch, Determinants of Laplacians in exterior domains. *Internat. Math. Res. Notices* 1999, no. 18, 971–1004.

- [HZ] A. Hassell and M. Zworski, Resonant rigidity of  $S^2$ , *J. Funct. Anal.* 169 (1999), no. 2, 604–609.
- [Hil] C. Hillar (private communication).
- [ISj] A. Iantchenko and J. Sjöstrand, Birkhoff normal forms for Fourier integral operators. II. *Amer. J. Math.* 124 (2002), no. 4, 817–850.
- [ISZ] A. Iantchenko, J. Sjöstrand and M. Zworski, Birkhoff normal forms in semi-classical inverse problems. *Math. Res. Lett.* 9 (2002), no. 2-3, 337–362.
- [I1] H. Ito, Convergence to Birkhoff normal forms for integrable systems, *Comment. Math. Helv.* 64 (1989), 412–461.
- [I2] H. Ito, Integrability of Hamiltonian systems and Birkhoff normal forms in the simple resonance case, *Math. Ann.* 292 (1992), 411–444.
- [K] W. Klingenberg, *Lectures on Closed Geodesics*, Grundlehren der. math. W. **230**, Springer-Verlag (1978).
- [KKL] A. Katchalov, Y. Kurylev, and Lassas, *Inverse boundary spectral problems*. Chapman and Hall/CRC Monographs and Surveys in Pure and Applied Mathematics, 123. Boca Raton, FL, 2001 Ruishi On isospectral deformations of Riemannian metrics. *Compositio Math.* 40 (1980), no. 3, 319–324.
- [Ka] A. Katok, Periodic and quasiperiodic orbits for twist maps. *Dynamical systems and chaos* (Sitges/Barcelona, 1982), 47–65, Lecture Notes in Phys., 179, Springer, Berlin, 1983.
- [KH] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*. Encyclopedia of Mathematics and its Applications, 54. Cambridge University Press, Cambridge, 1995.
- [KT] V. V. Kozlov and D. V. Treshchë v, *Billiards*, Trans. Math. Mono. 89, AMS, Providence R.I. (1991).
- [Ku2] R. Kuwabara, On isospectral deformations of Riemannian metrics. II. *Compositio Math.* 47 (1982), no. 2, 195–205.
- [Ku3] R. Kuwabara, On the characterization of flat metrics by the spectrum. *Comment. Math. Helv.* 55 (1980), no. 3, 427–444.
- [LL] L. D. Landau and E.M. Lifshitz, *Quantum Mechanics*, Course of Theoretical Physics Vol. 3, Third Edition, Pergamon Press (1977).
- [LU] M. Lassas and G. Uhlmann, On determining a Riemannian manifold from the Dirichlet-to-Neumann map. *Ann. Sci. cole Norm. Sup. (4)* 34 (2001), no. 5, 771–787
- [L] V. F. Lazutkin, Construction of an asymptotic series for eigenfunctions of the "bouncing ball" type. In *Asymptotic methods and stochastic models in wave propagation problems*. (Russian) *Trudy Mat. Inst. Steklov.* 95 (1968), 106–118, 215
- [La2] V. F. Lazutkin, Existence of caustics for the billiard problem in a convex domain. *Math. USSR-Izv.* 7 (1973), 185–214.
- [Ll] R. de la Llave, Analytic regularity of solutions of Livsic's cohomology equation and some applications to analytic conjugacy of hyperbolic dynamical systems, *Ergodic Theory Dynam. Systems* 17 (1997), 649–662.
- [L2] R. de la Llave, On necessary and sufficient conditions for uniform integrability of families of Hamiltonian systems. *International Conference on Dynamical Systems* (Montevideo, 1995), 76–109, Pitman Res. Notes Math. Ser., 362, Longman, Harlow, 1996.
- [LS] E. Lerman and N. Shirokova, Completely integrable torus actions on symplectic cones. *Math. Res. Lett.* 9 (2002), no. 1, 105–115.
- [Ma] A. Majda, High frequency asymptotics for the scattering matrix and the inverse problem of acoustical scattering. *Comm. Pure Appl. Math.* 29 (1976), no. 3, 261–291.
- [Ma2] A. Majda, A representation formula for the scattering operator and the inverse problem for arbitrary bodies. *Comm. Pure Appl. Math.* 30 (1977), no. 2, 165–194.
- [M] J. D. Meiss, Symplectic maps, variational principles, and transport. *Rev. Modern Phys.* 64 (1992), no. 3, 795–848.
- [MM] Sh. Marvizi and R. B. Melrose, Spectral invariants of convex planar regions. *J. Differ. Geom.* 17, 475–502 (1982).
- [MF] J.N. Mather and G. Forni, Action minimizing orbits in Hamiltonian systems, in: Graffi (ed.): *Transition to Chaos in Classical and Quantum Mechanics*, Springer LNM 1589 (1992), 92–186.

- [Me] R. B. Melrose, The inverse spectral problem for planar domains. *Instructional Workshop on Analysis and Geometry*, Part I (Canberra, 1995), 137–160, Proc. Centre Math. Appl. Austral. Nat. Univ., 34, Austral. Nat. Univ., Canberra, 1996.
- [Me2] R. B. Melrose, "Geometric Scattering Theory," Cambridge Univ. Press, Cambridge/New York/Melbourne, 1995
- [MS] R. B. Melrose and J. Sjöstrand, Singularities of boundary value problems. I. *Comm. Pure Appl. Math.* 31 (1978), no. 5, 593–617.
- [M1] R. Michel, Sur la rigidité imposée par la longueur des géodésiques, *Invent. Math.* 65 (1981), 71–83.
- [M2] R. Michel, Restriction de la distance géodésique à un arc et rigidité, *Bull. Soc. Math. France* 122 (1994), 435–442.
- [Mo] J. Moser, On invariant curves of area-preserving mappings of an annulus. *Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II* 1962 1962 1–20.
- [Mo1] J. Moser, Convergent series expansions for quasi-periodic motions, *Math. Annalen* 169 (1967), 136–176.
- [Mo.2] J. Moser, *Stable and Random Motions in Dynamical Systems*, Ann. Math. Studies 77, Princeton U Press, Princeton (1973).
- [Mu] R. G. Mukhometov, The reconstruction problem of a two-dimensional Riemannian metric, and integral geometry, (Russian), *Dokl. Akad. Nauk SSSR* 232 (1977), 32–35.
- [OPS] B. Osgood, R. Phillips, and P. Sarnak, Moduli space, heights and isospectral sets of plane domains. *Ann. of Math. (2)* 129 (1989), no. 2, 293–362.
- [O1] J. P. Otal, Le spectre marqué des longueurs des surfaces courbure négative. *Ann. of Math. (2)* 131 (1990), no. 1, 151–162.
- [O2] J. P. Otal, Sur les longueurs des géodésiques d'une métrique à courbure négative dans le disque, *Comment. Math. Helv.* 65 (1990), 334–347.
- [Pa] V.K. Patodi, Curvature and the fundamental solution of the heat operator. *J. Indian Math. Soc.* 34 (1970), no. 3-4, 269–285 (1971).
- [PM] R. Pérez-Marco, Convergence or generic divergence of the Birkhoff normal form. *Ann. of Math. (2)* 157 (2003), no. 2, 557–574.
- [PS] V.M. Petkov and L.N. Stoyanov, *Geometry of Reflecting Rays and Inverse Spectral Problems*, John Wiley and Sons, N.Y. (1992).
- [P] G. Popov, Invariants of the length spectrum and spectral invariants for convex planar domains. *Commun. Math. Phys.* 161, 335–364 (1994)
- [P2] G. Popov, Length spectrum invariants of Riemannian manifolds. *Math. Z.* 213 (1993), no. 2, 311–351.
- [P3] G. Popov, On the contribution of degenerate periodic trajectories to the wave-trace. *Comm. Math. Phys.* 196 (1998), no. 2, 363–383.
- [PT] G. Popov and P. Topalov, Liouville billiard tables and an inverse spectral result. *Ergodic Theory Dynam. Systems* 23 (2003), no. 1, 225–248.
- [Rau] J. Rauch, Illumination of bounded domains. *Amer. Math. Monthly* 85 (1978), no. 5, 359–361.
- [R] M. Rouleux, Semi-classical integrability, hyperbolic flows and the Birkhoff normal form, to appear in *Canadian J. Math* (see also arXiv preprint math.DS/0207026).
- [SU] V. Sharafutdinov and G. Uhlmann, On deformation boundary rigidity and spectral rigidity of Riemannian surfaces with no focal points. *J. Differential Geom.* 56 (2000), no. 1, 93–110.
- [S1] K. F. Siburg, Aubry-Mather theory and the inverse spectral problem for planar convex domains. *Israel J. Math.* 113 (1999), 285–304.
- [S2] K. F. Siburg, Symplectic invariants of elliptic fixed points. *Comment. Math. Helv.* 75 (2000), no. 4, 681–700.
- [S3] K. F. Siburg, The minimal action in geometry and dynamics (to appear).
- [S.M] C.L. Siegel and J. Moser, *Lectures on Celestial Mechanics*, Grundlehren der math. Wiss. Einz. 187, Springer-Verlag, Berlin (1971).
- [SjZ] J. Sjöstrand and M. Zworski, Quantum monodromy and semi-classical trace formulae, *J. Math. Pure Appl.* 81 (2002), 1–33.
- [Sm] L. Smith, The asymptotics of the heat equation for a boundary value problem. *Invent. Math.* 63 (1981), 467–493.

- [StU] P. Stefanov and G. Uhlmann, Rigidity for metrics with the same lengths of geodesics. *Math. Res. Lett.* 5 (1998), no. 1-2, 83–96.
- [St] L. Stoyanov, Rigidity of the scattering length spectrum. *Math. Ann.* 324 (2002), no. 4, 743–771.
- [St2] L. Stoyanov, On the scattering length spectrum for real analytic obstacles. *J. Funct. Anal.* 177 (2000), no. 2, 459–488.
- [Su] T. Sunada, Riemannian coverings and isospectral manifolds. *Ann. of Math. (2)* 121 (1985), no. 1, 169–186.
- [T] S. Tanno, A characterization of the canonical spheres by the spectrum. *Math. Z.* 175 (1980), no. 3, 267–274.
- [T2] S. Tanno, Eigenvalues of the Laplacian of Riemannian manifolds. *Tōhoku Math. J. (2)* 25 (1973), 391–403.
- [T3] S. Tanno, A characterization of a complex projective space by the spectrum. *Kodai Math. J.* 5 (1982), no. 2, 230–237
- [Wa] C.P. Walkden, C. P. Livsic theorems for hyperbolic flows. *Trans. Amer. Math. Soc.* 352 (2000), no. 3, 1299–1313
- [W] K. Watanabe, Plane domains which are spectrally determined. *Ann. Global Anal. Geom.* 18 (2000), no. 5, 447–475.
- [Z1] S. Zelditch, Spectral determination of analytic bi-axisymmetric plane domains. *Geom. Funct. Anal.* 10 (2000), no. 3, 628–677.
- [Z2] S. Zelditch, The inverse spectral problem for surfaces of revolution. *J. Differential Geom.* 49 (1998), no. 2, 207–264.
- [Z3] S. Zelditch, Wave invariants at elliptic closed geodesics. *Geom. Funct. Anal.* 7 (1997), no. 1, 145–213.
- [Z4] S. Zelditch, Wave invariants for non-degenerate closed geodesics. *Geom. Funct. Anal.* 8 (1998), no. 1, 179–217.
- [Z5] S. Zelditch, The inverse spectral problem for analytic plane domains. I: Balian-Bloch trace formula (<http://arXiv.org/abs/math.SP/0111077>, to appear in *Comm. Math. Phys.*); and II: domains with symmetry, (<http://arXiv.org/abs/math.SP/0111078>.)
- [Z6] S. Zelditch, Isospectrality in the FIO category. *J. Differential Geom.* 35 (1992), no. 3, 689–710.
- [Z7] S. Zelditch, Normal forms and inverse spectral theory, Journées Équations aux dérivées partielles, 2–5 June 1998, GDR 1151 (CNRS).
- [Z8] S. Zelditch, Maximally degenerate Laplacians. *Ann. Inst. Fourier (Grenoble)* 46 (1996), no. 2, 547–587.
- [Z9] S. Zelditch, Lectures on wave invariants. *Spectral theory and geometry* (Edinburgh, 1998), 284–328, London Math. Soc. Lecture Note Ser., 273, Cambridge Univ. Press, Cambridge, 1999.
- [Z10] S. Zelditch Inverse resonance problem for  $Z_2$  symmetric analytic obstacles in the plane, IMA Volume 137: *Geometric Methods in Inverse Problems and PDE Control*. C.B. Croke, I. Lasiecka, G. Uhlmann, and M. S. Vogelius eds.
- [Z11] S. Zelditch, Kuznecov sum formulae and Szegő limit formulas on manifolds. *Comm. P.D.E.* 17, 221–260 (1992)
- [Zw2] M. Zworski, Resonances in physics and geometry. *Notices Amer. Math. Soc.* 46 (1999), no. 3, 319–328.
- [Zw3] M. Zworski, Quantum resonances and partial differential equations. *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, 243–252, Higher Ed. Press, Beijing, 2002.

## 9. Appendix: Quantum monodromy revisited (by Johannes Sjöstrand and Maciej Zworski)

We present a few simplifications of the presentation of the *quantum monodromy* operator in [1] and [2].

We first repeat the comment made in [1, §4]: the trace formula of [2] is formulated in terms of a general Hamiltonian,  $P(z)$  (for instance an effective Hamiltonian with a non-linear dependence on  $z$ ). However, the proof can be reduced to the case



of  $P - z$ . In fact, the assumptions of [2, Theorem 2], the implicit function theorem, and the usual symbolic iteration, imply that

$$P(z) = A(z)^*(P - z)A(z),$$

with  $A(z) \in \Psi_h^{0,k/2}(X)$  elliptic near  $\gamma(0)$ , and  $P \in \Psi_h^{0,0}(X)$  self-adjoint. Replacing  $P(z)$  by  $P - z$  in [2, Theorem 2] changes the trace by  $\mathcal{O}(h^\infty)$ .

In the special case of  $P(z) = P - z$  the monodromy operator can be written quite simply (though we still believe that it is interesting to consider  $M(z)$  for the non-linear  $P(z)$  as done in [2]). Let us recall that at a point on an integral curve of  $P - z$ ,  $\gamma(z)$ ,  $m_0(z) \in \gamma$ , we can define the microlocal kernel of  $P - z$  at  $m_0(z)$ , to be the set of families  $u(h)$ , such that  $u(h)$  are microlocally defined near  $m_0$  and

$$(P - z)u(h) = \mathcal{O}(h^\infty) \text{ near } m_0.$$

We denote it by  $\ker_{m_0(z)}(P - z)$ . Since microlocally, near a given point, the operator  $P - z$  can be reduced to  $hD_{x_1}$  any solution can be continued microlocally along  $\gamma(z)$  and we denote the corresponding forward and backward continuations by  $I_\pm(z)$ . We can also define the propagator  $\exp(-it(P - z)/h)$  and we see that

$$\exp(-it(P - z)/h) : \ker_{m_0(z)}(P - z) \longrightarrow \ker_{\exp(tH_p)m_0(z)}(P - z).$$

This follows from the fact that  $(P - z)\exp(-it(P - z)/h) = \exp(-it(P - z)/h)(P - z)$ , and propagation of semi-classical wave fronts:  $WF_h(\exp(-it(P - z)/h)u)$  is contained in a neighbourhood of  $\exp(tH_p)m_0(z)$  if  $WF_h(u)$  is contained in a neighbourhood of  $m_0(z)$ . Hence we have

$$(9.1) \quad I_\pm(z) = \exp(\mp it(P - z)/h)$$

microlocally near  $(m_0(z), \exp(tH_p)(m_0(z)))$ .

To define the quantum monodromy we take  $m_1(z) \neq m_0(z)$  be another point on  $\gamma(z)$  and put

$$(9.2) \quad \begin{aligned} I_-(z)\mathcal{M}(z) &= I_+(z), \quad \text{near } m_1, \\ \mathcal{M}(z) &: \ker_{m_0(z)}(P - z) \longrightarrow \ker_{m_0(z)}(P - z). \end{aligned}$$

In view of (9.1) we now have

$$(9.3) \quad \mathcal{M}(z) = \exp(-iT(z)(P - z)/h) : \ker_{m_0(z)}(P - z) \longrightarrow \ker_{m_0(z)}(P - z),$$

where  $T(z)$  is the period of  $\gamma(z)$  but for  $z$  small we can replace it by a fixed period,  $T(0)$ .

The operator  $P(z)$  is assumed to be self-adjoint with respect to some inner product  $\langle \bullet, \bullet \rangle$ , and we define the *quantum flux* norm on  $\ker_{m_0(z)}(P - z)$  as follows: let  $\chi$  be a microlocal cut-off function supported near  $\gamma$  and equal to one near the part of  $\gamma$  between  $m_0$  and  $m_1$ , in the positive direction determined by  $H_p$ . We denote by  $[P, \chi]_+$  the part of the commutator supported near  $m_0$ , or more generally, near the left end point (using the orientation determined by  $H_p$ ) of the support of  $\chi|_\gamma$ . We then put

$$\langle u, v \rangle_{\text{QF}} \stackrel{\text{def}}{=} \langle [(i/h)P, \chi]_+ u, v \rangle, \quad u, v \in \ker_{m_0(z)}(P - z).$$

It is easy to check that this norm is independent of the choice of  $\chi$ : if  $\tilde{\chi}$  agrees with  $\chi$  near  $m_1$  we see that  $[P, \tilde{\chi} - \chi]_+ = [P, \tilde{\chi} - \chi]$  and clearly  $\langle [P, \tilde{\chi} - \chi]u, v \rangle = 0$

(see [2, Lemma 4.4] for more details). This independence leads to the unitarity of  $\mathcal{M}(z)$ :

$$\begin{aligned}
 \langle \mathcal{M}(z)u, \mathcal{M}(z)u \rangle_{\text{QF}} &= \langle (i/h)[P, \chi]_+ e^{-iT(z)(P-z)/h} u, e^{-iT(z)(P-z)/h} v \rangle \\
 (9.4) \qquad \qquad \qquad &= \langle (i/h)[P, e^{iT(z)(P-z)/h} \chi e^{-iT(z)(P-z)/h}]_+ u, v \rangle \\
 &= \langle (i/h)[P, \chi]_+ u, v \rangle = \langle u, v \rangle_{\text{QF}}
 \end{aligned}$$

As already recalled above the operator  $P - z$  can be reduced to  $hD_{x_1}$  we can identify  $\ker_{m_0(z)}(P - z)$  with  $\mathcal{D}'(\mathbb{R}^n)$ . This is done microlocally near  $(0, 0)$ , and we can choose the identification,  $K(z)$ , so that

$$K^*(z)(i/h)[P, \chi]_+ K(z) = Id.$$

This guarantees that the corresponding monodromy operator,

$$M(z) \stackrel{\text{def}}{=} K(z)^{-1} \mathcal{M}(z) K(z)^{-1} : \mathcal{D}'(\mathbb{R}^n) \longrightarrow \mathcal{D}'(\mathbb{R}^n),$$

microlocally defined near  $(0, 0)$ , is unitary (microlocally near  $(0, 0)$ ). Here  $(0, 0)$  corresponds to the closed orbit intersecting a transversal identified with  $T^*\mathbb{R}^n$ . We easily see that  $M(z)$  is a semi-classical Fourier integral operator which quantizes the Poincaré map of  $\gamma(z)$ .

Using (9.3),

$$(9.5) \qquad M(z) = K(z)^{-1} \circ \exp(-iT(0)(P - z)/h) \circ K(z).$$

This expression trivializes the proof of [2, Lemma 6.2] in the case  $P(z) = P - z$ . For  $P = hD_{x_1}$ ,  $K(z)u(x) = e^{izx_1/h}u(x')$ ,  $x = (x_1, x')$ ,  $x' \in \mathbb{R}^n$ , and hence the complexification of  $z$  in  $K(z)$  produces growth of size  $\mathcal{O}(e^{\epsilon|\text{Im } z|/h})$ . Then (9.5) shows that (for  $z$  close to 0)

$$\begin{aligned}
 (9.6) \qquad \qquad \qquad \|M(z)\| &\leq e^{-(T(0)-\epsilon)\text{Im } z/h}, \quad 0 < \text{Im } z, \\
 \|M(z)^{-1}\| &\leq e^{(T(0)-\epsilon)\text{Im } z/h}, \quad \text{Im } z < 0.
 \end{aligned}$$

The rather subtle [2, Lemma 6.1] is altogether unnecessary (unless we want the results for the general  $P(z)$ ; they are however not needed for the trace formula). The estimates (9.6) also give a slight improvement in [2, Theorem 1]: we can make the conditions on the support of  $\hat{f}$  there optimal:  $\hat{f} \in C_c^\infty(\mathbb{R})$ ,  $\text{supp } \hat{f} \subset (-NT, NT) \setminus \{0\}$ ,

For a discussion of the quantum monodromy operator in a concrete setting and a relation of the quantum flux norm to the more standard objects, see the appendix to [1].

### References

[1] A. Iantchenko, J. Sjöstrand, and M. Zworski, *Birkhoff normal forms in semi-classical inverse problems*, Math. Res. Lett. **9**(2002), 337-362.  
 [2] J. Sjöstrand and M. Zworski, *Quantum monodromy and semi-classical trace formulae*, J. Math. Pure Appl. **81**(2002), 1-33.

S.ZELDITCH, DEPARTMENT OF MATHEMATICS, JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD 21218, USA

*E-mail address:* zelditch@math.jhu.edu

J. SJÖSTRAND, ECOLE POLYTECHNIQUE, CENTRE DE MATHÉMATIQUES, UMR 7460, CNRS, F-91128 PALAISEAU CEDEX, FRANCE

*E-mail address:* johannes@math.polytechnique.fr

M. ZWORSKI, UNIVERSITY OF CALIFORNIA, MATHEMATICS DEPARTMENT, EVANS HALL, BERKELEY, CA 94720, USA

*E-mail address:* [zworski@math.berkeley.edu](mailto:zworski@math.berkeley.edu)