ЖЭТФ, 2013, том 144, вып. 3 (9), стр. 595–616

STATISTICAL MECHANICS OF COULOMB GASES AS QUANTUM THEORY ON RIEMANN SURFACES

T. Gulden [a], ***M. Janas*** [a], ***P. Koroteev*** [a,b], ***A. Kamenev*** [a,c*]

[a] *Department of Physics, University of Minnesota MN 55455, Minneapolis, USA*

[b] *Perimeter Institute for Theoretical Physics ON N2L2Y5, Canada*

[c] *William I. Fine Theoretical Physics Institute, University of Minnesota MN 55455, Minneapolis, USA*

Received March 26, 2013

Dedicated to the memory of Professor Anatoly Larkin

Statistical mechanics of a 1D multivalent Coulomb gas can be mapped onto non-Hermitian quantum mechanics. We use this example to develop the instanton calculus on Riemann surfaces. Borrowing from the formalism developed in the context of the Seiberg–Witten duality, we treat momentum and coordinate as complex variables. Constant-energy manifolds are given by Riemann surfaces of genus $g \geq 1$. The actions along principal cycles on these surfaces obey the ordinary differential equation in the moduli space of the Riemann surface known as the Picard–Fuchs equation. We derive and solve the Picard–Fuchs equations for Coulomb gases of various charge content. Analysis of monodromies of these solutions around their singular points yields semiclassical spectra as well as instanton effects such as the Bloch bandwidth. Both are shown to be in perfect agreement with numerical simulations.

DOI: 10.7868/S0044451013090125

1. INTRODUCTION

One of the very last works of Anatoliy Larkin [1] was devoted to transport through ion channels of biological membranes. An ion channel may be roughly viewed as a cylindrical water-filled tube surrounded by a lipid membrane. Its typical radius $a \approx 6\,\text{Å}$ is much smaller than its length $L \approx 120\,\text{Å}$. The important observation with far-reaching consequences, made in Ref. [1], is that the dielectric constant of water $\varepsilon_{water} \approx 80$ is significantly larger than that of the surrounding lipid membrane $\varepsilon_{lipid} \approx 2$. This defines a new length scale

$$\xi \approx \sqrt{\frac{a\varepsilon_{water}}{\varepsilon_{lipid}}} \ln \frac{\varepsilon_{water}}{\varepsilon_{lipid}} \approx 140\,\text{Å}$$

over which the electric field stays inside the channel and does not escape into the surrounding media. Since $\xi \gtrsim L$, the ions inside the channel interact essentially through the 1D Coulomb potential

$$U(x_1 - x_2) \approx eE_0|x_1 - x_2|,$$

where $E_0 = 2e/a^2\varepsilon_{water}$ is a discontinuity of the electric field created by a unit charge. This fact dictates a significant energy barrier $U(L/4) \approx 4k_BT_{room}$ for moving a single ion through the channel. If indeed present, such a barrier would essentially impede ion transport, preventing the channel from performing its biological functions.

Nature removes such Coulomb blocking by screening. A moving ion is screened either by mobile ions of dissociated salt [1], or by immobilized charged radicals attached to the walls of the channel [2–9]. Nevertheless, due to the peculiar nature of the long-range 1D Coulomb potential, the transport barrier proportional to the channel length L is always present. Its magnitude, however, is typically suppressed [1] down to about k_BT_{room}, allowing for a relatively unimpeded transport of ions. These considerations call for developing a transport theory of 1D Coulomb gases. Following the

*E-mail: kamenev@physics.umn.edu

celebrated mapping of 1D statistical mechanics onto an effective quantum mechanics, pioneered by Edwards and Lenard [10] and Vaks, Larkin, and Pikin [11], the authors of [1] mapped the problem onto quantum mechanics with a cosine potential (we briefly review this mapping in Sec. 2). The ground-state energy of such quantum mechanics is exactly the equilibrium pressure in the Coulomb plasma. Moreover, the width of the lowest Bloch band is a specific energy barrier for ion transport through the channel.

It is instructive to notice that the $2\alpha\cos\theta = \alpha(e^{i\theta}+e^{-i\theta})$ potential describes a mixture of positive, $e^{i\theta}$, and negative, $e^{-i\theta}$, monovalent ions with concentration α. We can also consider a situation where the channel is filled with a solution of dissociated multivalent salt, e. g., divalent $CaCl_2$ or trivalent $AlCl_3$. In these cases, the corresponding 1D statistical mechanics is mapped onto the quantum problem with a non-Hermitian potential such as $\alpha(e^{2i\theta}/2+e^{-i\theta})$ or $\alpha(e^{3i\theta}/3+e^{-i\theta})$ [2, 10]. This paper is devoted to efficient mathematical methods of treating non-Hermitian quantum mechanics of this sort.

Our particular focus here is on a semiclassical treatment, applicable in the regime of a sufficiently large salt concentration α. In its framework, the energy spectrum (and hence the pressure) is determined by the Bohr–Sommerfeld quantization condition for the action of classical periodic orbits. On the other hand, the bandwidth (and hence the transport barrier) is given by the exponentiated action accumulated on an instanton trajectory, running through the classically forbidden part of the phase space. The traditional techniques of Hermitian quantum mechanics call for finding classical and instanton trajectories by solving equations of motion in real and imaginary time and evaluating corresponding actions. This route cannot be straightforwardly followed in non-Hermitian quantum problems arising in the context of multivalent Coulomb gases. Even leaving aside the technical difficulties of solving complex equations of motion, there are conceptual difficulties with identifying periodic orbits as well as the meaning of classically allowed vs forbidden regions and with the imaginary time procedure.

In this paper, we borrow from the algebraic topology methods developed in the past decades in the context of the Seiberg–Witten solution [12, 13] and its applications to integrable systems [14–16] (and many follow-up contributions). The central idea is to consider both the coordinate θ and the corresponding canonical momentum p as complex variables. This leads to a four-dimensional (4D) phase space. Then (complex) energy conservation restricts the trajectories to live on 2D Riemann surfaces embedded into the 4D phase space. The dynamics of the system are essentially determined by the topology, i. e., the genus g of such Riemann surfaces. We show that mono- and divalent gases are described by tori, while trivalent and 4-valent lead to genus-2 surfaces, etc. The Cauchy theorem and the resulting freedom to deform the integration contour in the complex space allows us to avoid finding specific solutions of the equations of motion. Instead, we identify the homology cycles on the Riemann surface and find the corresponding action integrals, which depend only on the topology of the cycles and not on their specific shape. For example, the cosine potential of a monovalent gas leads to a torus, which obviously has two topologically distinct cycles (see Fig. 6 below). The two turn out to be related to classical and instanton actions correspondingly. The genus $g \geq 1$ Riemann surfaces admit $2g$ topologically distinct cycles. Below, we identify and explain the meaning of the corresponding action integrals.

The shape of the specific Riemann surface depends on the parameters of the problem, e. g., salt concentration α in our case. Such parameters are called moduli of the Riemann surface. It turns out that the action integrals, being functions of the moduli, satisfy a closed ordinary differential equation (ODE) of the order $2g$, known as the Picard–Fuchs equation. The actions can be found as solutions of this ODE in the moduli space, rather than by performing integrations over cycles on the surface. Below, we derive and solve Picard–Fuchs equations for several (positive and negative) ionic charge combinations, such as the genus $g = 1$ cases $(1,1), (2,1)$ and the genus $g = 2$ cases $(3,1), (3,2), (4,1)$. We then discuss how to connect the principal classical actions with the spectra of the corresponding quantum problem. The key observation is that in the moduli space, the actions exhibit a few isolated branching points. Going around such a branching point transforms the actions into their linear combinations, effecting an $Sp(2g,\mathbb{Z})$ monodromy transformation. The invariance of quantum observables under monodromy transformations dictates Bohr–Sommerfeld quantization for one of the principal classical actions. The remaining actions can be identified with the instanton processes, e. g., related to the Bloch bandwidth.

Statistical mechanics of 1D Coulomb gases may seem to be an isolated problem, not worthy of developing an extensive mathematical apparatus. Our goal here is to use it as a test-drive example, grounded into a well-posed physics problem, to develop a machinery applicable in other setups. Recently, the so-

called $\mathcal{PT}$ symmetric non-Hermitian quantum mechanics attracted much attention for its application in active optics [17] and open quantum systems [18], as well as in the description of antiferromagnetic lattices [19] and calculating energy states in larger molecules [20]. Our examples also belong to the class of $\mathcal{PT}$ symmetric problems. It seems likely that the methods developed here may be applied to advance analytic understanding of a broader class of $\mathcal{PT}$ symmetric quantum mechanics. Another context where complexified quantum mechanics was proven to be extremely useful, is dynamics of large molecular spins [21, 22]. Indeed, functional integral representation of the spin dynamics leads naturally to the Hamiltonian formulation, where the projective coordinates $(z, \bar{z})$ on the sphere play the role of the canonical pair [23]. It was realized in [21, 22] that to find instanton trajectories, one has to consider z and $\bar{z}$ as independent complex variables, thus expanding the dynamics into 4D phase space. The Riemannian geometry methods seem to be well-suited to advance this subject as well.

This paper is organized as follows. In Sec. 2, we outline the relation between 1D multivalent Coulomb gases and non-Hermitian quantum mechanics and discuss general symmetries of the latter. In Sec. 3, we summarize major numerical observations regarding complex spectra and the band structure for the family of Hamiltonians considered here. In Sec. 4, we illustrate the machinery of algebraic geometry on Riemann surfaces for the familiar Hermitian cosine potential quantum mechanics, which corresponds to the monovalent $(1, 1)$ gas. There, we introduce the complexified phase space and Riemann tori of constant energy; we then derive, solve, and analyze solutions of the Picard–Fuchs equations. In Sec. 5, we apply the developed methods for the divalent $(2, 1)$ Coulomb gas, which is also described by a genus-1 torus. In Sec. 6, we extend the method for genus-2 example of a trivalent $(3, 1)$ gas, which exhibits some qualitatively new features. The $(3, 2)$ and $(4, 1)$ gases are briefly discussed in Sec. 7. In Sec. 8, we outline connections to the Seiberg–Witten theory. We conclude with a brief discussion of the results in Sec. 9.

2. MAPPING OF COULOMB GASES ONTO QUANTUM MECHANICS

We consider a 1D gas of cations with charge $n_1 e$ and anions with charge $-n_2 e$, where (n_1, n_2) are positive integers. By Gauss's theorem, the electric field at a distance x larger than the radius of the channel a from a unit charge is $E_0 = 2e/a^2\varepsilon_{water}$. At the location of a charge $n_{1,2}$, the electric field exhibits a discontinuity $\pm 2E_0 n_{1,2}$. Since all charges are integers, the field is conserved modulo $2E_0$ along the channel. This allows defining the order parameter [1, 3] $q = E(x) (\mathrm{mod}\, 2E_0)$, which acts like an effective boundary charge $\pm q$ at the two ends of the channel. The Poisson equation in 1D is $\nabla^2\phi = -2E_0\delta(x)$, leading to the 1D Coulomb potential $\phi(x) = -E_0|x|$. The potential energy of the gas is thus

$$U = -\frac{eE_0}{2}\sum_{i,j}\sigma_i\sigma_j|x_i - x_j|, \qquad (1)$$

where σ_j is the charge n_1 or $-n_2$ of an ion at the position x_j and we omit the $\pm q$ boundary charges for brevity. Our goal is to evaluate the grand canonical partition function of the gas in the channel of length L,

$$\mathcal{Z}_L = \sum_{N_1,N_2=0}^{\infty}\frac{f_1^{N_1}f_2^{N_2}}{N_1!N_2!}\prod_{i=1}^{N_1}\int_0^L dx_i\prod_{j=1}^{N_2}\times \\ \times\int_0^L dx_j\exp\left(-\frac{U}{k_BT}\right), \qquad (2)$$

where $f_{1,2}$ are fugacities of the two charge species. We can now introduce the charge density using a delta-function $\delta[\rho(x) - \sum_j \sigma_j\delta(x - x_j)]$. The delta-function is elevated to the exponent with the help of the auxiliary field $\theta(x)$. This procedure decouples all x_j integrals [1], bringing them to the form

$$\sum_N\frac{[f\int dx\, e^{i\sigma\theta(x)}]^N}{N!} = \exp\left\{f\int dx\, e^{i\sigma\theta(x)}\right\}.$$

Interaction potential (1), being inverse of the 1D Laplace operator, leads to $\exp\{(T/eE_0)\int dx\,\theta\partial_x^2\theta\}$. As a result, partition function (2) is identically written as a Feynman path integral, in an "imaginary time" x, for the quantum mechanics with the Hamiltonian

$$\hat{H} = (i\partial_\theta - q)^2 - [\alpha_1\exp(in_1\theta) + \alpha_2\exp(-in_2\theta)], \qquad (3)$$

where $\alpha_{1,2} = f_{1,2}k_BT/eE_0$ are dimensionless ion concentrations. Such a Feynman integral is the expectation value of the evolution operator during "time" L, leading to

$$\mathcal{Z}_L = \left\langle q\middle|\mathcal{X}\exp\left(-\frac{eE_0}{k_BT}\int_0^L dx\,\hat{H}\right)\middle|q\right\rangle = \\ = \sum_m|\langle q|m\rangle|^2\exp\left(-\frac{eE_0L}{k_BT}\epsilon_m(q)\right), \qquad (4)$$

where $\mathcal{X}$ stands for the x-ordered exponential. Here, $\epsilon_m(q)$ are eigenvalues of the effective Hamiltonian $\hat{H}$ and $|m\rangle = \psi_m(\theta)$ are its eigenvectors in the Hilbert space of periodic functions $\psi_m(\theta) = \psi_m(\theta + 2\pi)$, and finally the matrix elements are $\langle q|m\rangle = \int_0^{2\pi} d\theta e^{-iq\theta}\psi_m(\theta)$. The boundary charge q plays the role of the Bloch quasimomentum and the spectrum is obviously periodic in q with the unit period (reflecting the fact that the integer part of the boundary charge may be screened by mobile ions and thus becomes inconsequential).

The pressure of the Coulomb gas is its free energy per unit length,

$$P = k_B T \frac{\partial \ln \mathcal{Z}_L}{\partial L} \xrightarrow{L\to\infty} -eE_0\epsilon_0(q), \tag{5}$$

where $\epsilon_0(q)$ is the eigenvalue with the smallest real part. In equilibrium, the system minimizes its free energy by choosing an appropriate boundary charge q. In all cases considered below, the minimum turns out to be a nonpolarized state of the channel, i. e., $q = 0$ (see Refs. [2] for exceptions to this rule, however). Adiabatic charge transfer through the channel is associated with the boundary charge q sweeping through its full period. As a result, the (free) energy barrier for ion transport is

$$U_0 = eE_0 L\Delta_0, \tag{6}$$

where Δ_0 is the width of the lowest Bloch band. Therefore, the ground-state energy and the width of the lowest Bloch band of Hamiltonian (3) determine thermodynamic and transport properties of the (n_1, n_2) Coulomb gas. The rest of this paper is devoted to a semiclassical theory of the spectral properties of such Hamiltonians. We start by discussing some general symmetries of non-Hermitian Hamiltonian (3).

2.1. $\mathcal{PT}$ Symmetry

Although the Hamiltonian in (3) is non-Hermitian for $n_1 \neq n_2$, it obeys $\mathcal{PT}$ symmetry [24, 25]. Here, the parity operator $\mathcal{P}$ acts as $\theta \to -\theta$ and the time-reversal operator $\mathcal{T}$ works as complex conjugation $i \to -i$. Clearly, the two operations combined leave Hamiltonian (3) unchanged. It can be proved [25, 26] that all eigenvalues of $\mathcal{PT}$-symmetric Hamiltonians are either real or occur in complex-conjugate pairs. As shown below for positive values of concentrations $\alpha_{1,2} > 0$, the lowest-energy band $\epsilon_0(q)$ is entirely real, ensuring the positivity of the partition function. The higher bands $\epsilon_m(q)$ are in general complex. It is interesting to note that for unphysical negative concentrations $\alpha_{1,2} < 0$, already the lowest band $\epsilon_0(q)$ is complex, making the free energy ill-defined.

2.2. Isospectrality

The spectrum of Hamiltonian (3) is invariant under shifts of the coordinate $\theta \to \theta + \theta_0$, where θ_0 is an arbitrary complex number. Under this transformation (preserving the periodic boundary conditions), the dimensionless concentrations $\alpha_{1,2}$ renormalize as $\alpha_1 \to \alpha_1 \exp(in_1\theta_0)$ and $\alpha_2 \to \alpha_2 \exp(-in_2\theta_0)$. We note that the combination $\alpha_1^{n_2}\alpha_2^{n_1}$ remains invariant. We hence conclude that the family of Hamiltonians (3) with

$$\alpha_1^{n_2}\alpha_2^{n_1} = \text{const} \tag{7}$$

is isospectral [10]. Therefore, without loss of generality, we can pick one representative from each isospectral family. It is convenient to choose such a representative to manifestly enforce charge neutrality in the bulk reservoirs. For this, we take $\alpha_1 n_1 = \alpha_2 n_2 = \alpha$, which brings Hamiltonian (3) to the form

$$\hat{H} = \alpha\left[\hat{p}^2 - \left(\frac{1}{n_1}\exp(in_1\theta) + \frac{1}{n_2}\exp(-in_2\theta)\right)\right], \tag{8}$$

where we have defined the momentum operator as

$$\hat{p} = \alpha^{-1/2}(-i\partial_\theta + q), \quad [\theta, \hat{p}] = i\alpha^{-1/2}. \tag{9}$$

The commutation relation shows that $\alpha^{-1/2}$ plays the role of the effective Planck constant. With the help of isospectrality condition (7), a proper α can always be chosen such that the spectrum of Hamiltonian (8) is identical with that of a Hamiltonian with arbitrary $\alpha_{1,2}$. The physical reason for this symmetry is that the interior region of the long channel always preserves charge neutrality, allowing the edge regions to screen charge imbalance of the reservoirs. Therefore, irrespective of the relative fugacities of cations and anions in the reservoirs, the thermodynamics of the long channel is equivalent to the one in contact with neutral reservoirs with an appropriate salt concentration α. Hereafter, we restrict ourselves to the neutral Hamiltonian (8) with a single parameter α.

3. NUMERICAL ANALYSIS

In this section, we discuss numerical simulation of the spectrum of Hamiltonian (8). We focus on unequal charges $n_1 \neq n_2$, because the case of $n_1 = n_2$ reduces to the well-known Hermitian cosine potential [27, 28].

For unequal charges, the Hamiltonian is non-Hermitian but $\mathcal{PT}$ symmetric, allowing for complex eigenvalues, which appear in conjugate pairs [25, 26].

Since the Hamiltonian $\hat{H}$ acts in the Hilbert space of periodic functions, we can choose the complete basis in the form $\{e^{im\theta}\}_{m\in\mathbb{Z}}$. In this basis, the Hamiltonian is represented by an infinite-size real matrix [2]

$$\hat{H}_{m,m'} = (m-q)^2\delta_{m,m'} - \\ - \alpha\left(\frac{1}{n_1}\delta_{m+n_1,m'} + \frac{1}{n_2}\delta_{m-n_2,m'}\right). \quad (10)$$

The boundary charge q plays the role of quasimomentum residing in the Brillouin zone $q \in [-1/2, 1/2]$. To numerically calculate the energy spectrum $\epsilon_m(q)$, we truncate the matrix at a large cutoff, after verifying that a further increase in the matrix size does not change the low-energy spectrum. We left the boundary conditions "open", i. e., did not change the matrix elements near the cutoff, after verifying that different boundary conditions do not affect the result. It is easy to see that the matrix size should be much more than $\sqrt{\alpha}$ to accurately represent the low-energy spectrum. As an illustration, we show the Hamiltonian cut to a 5×5 matrix for the divalent $(2,1)$ gas:

$$\begin{pmatrix} (-2-q)^2 & 0 & -\alpha/2 & 0 & 0 \\ -\alpha & (-1-q)^2 & 0 & -\alpha/2 & 0 \\ 0 & -\alpha & (0-q)^2 & 0 & -\alpha/2 \\ 0 & 0 & -\alpha & (1-q)^2 & 0 \\ 0 & 0 & 0 & -\alpha & (2-q)^2 \end{pmatrix}.$$

For reasons that become apparent below, it is convenient to represent the spectrum ϵ on the complex plane of the normalized energy u defined as

$$u = \frac{n_1 n_2}{n_1+n_2}\,\frac{\epsilon}{\alpha}. \quad (11)$$

For the divalent $(2,1)$ gas, $u = 2\epsilon/3\alpha$ and the corresponding spectra are shown in Fig. 1. The spectrum consists of a sequence of complex Bloch bands. The number of narrow bands within the unit circle $|u| = 1$ scales as $\sqrt{\alpha}$. They form three branches, which terminate at $u = -1$ and $u = e^{\pm i\pi/3}$ and approximately line up along the lines connecting the termination points with the point $u = 1$. We discuss the corresponding bandwidths below. Outside the unit circle, the bands are wide and centered near the positive real axis of energy.

Figure 2 shows the band structure in the first Brillouin zone $|q| < 1/2$ for $\alpha = 1$. We note that the lowest Bloch band is purely real (this is always the case for

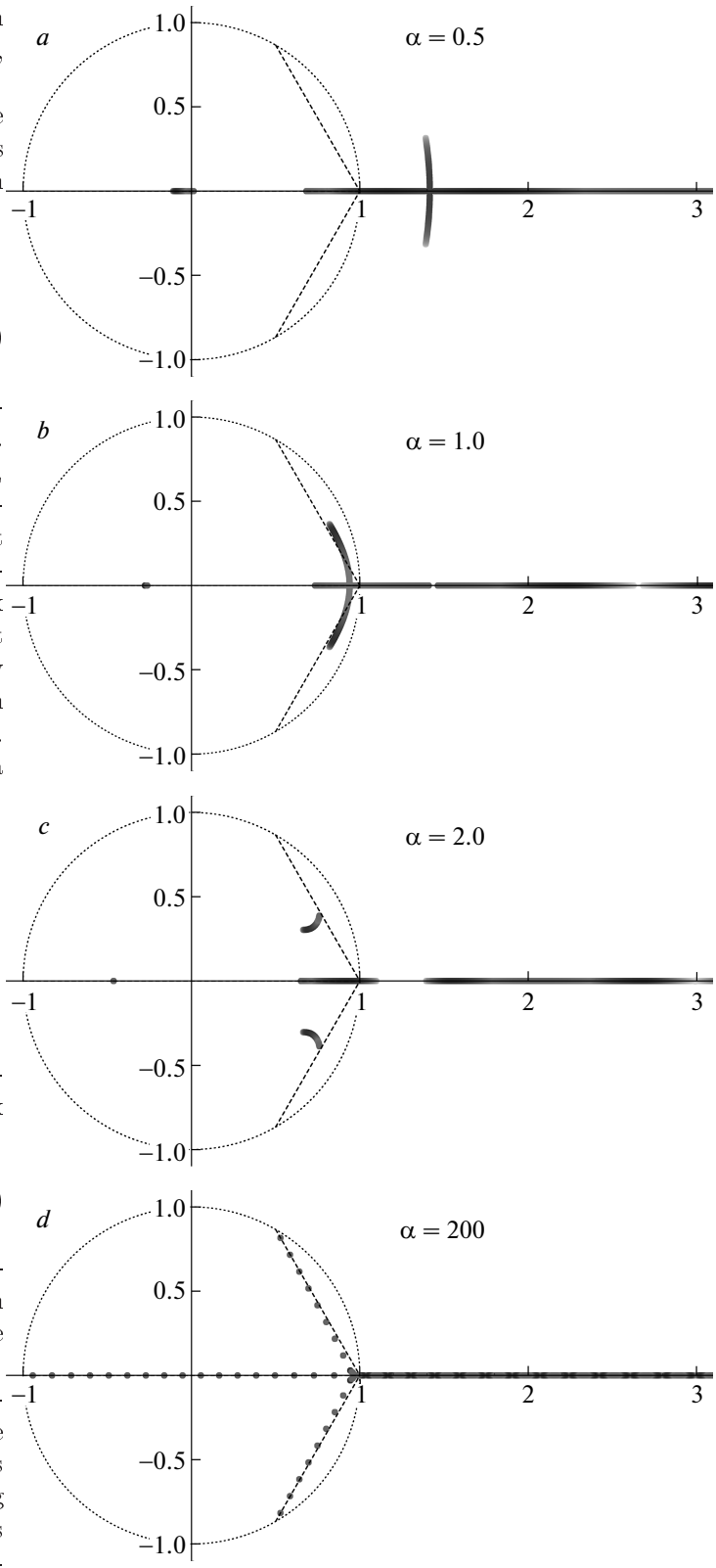

Fig. 1. Complex plane of the normalized energy $u = 2\epsilon_m(q)/3\alpha$ for the $(2,1)$ gas. The color corresponds to different values of the quasimomentum q; dark gray stands for $q = 0$ and light gray for $q = \pm 1/2$. The dotted circle is $|u| = 1$, the dashed lines connect between $u = 1$ and $u = e^{\pm i\pi/3}$, indicating positions of the narrow complex bands in the limit of large α

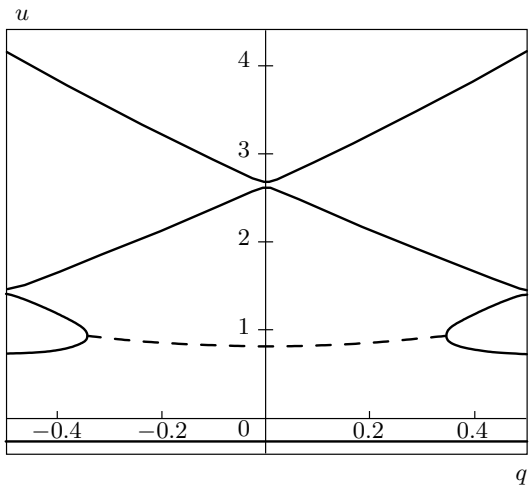

Fig. 2. Band structure for the $(2,1)$ gas with $\alpha = 1$ (cf. Fig. 1*b*) vs the boundary charge (quasimomentum) q. For complex bands, the real part of $\epsilon_m(q)$ is shown in dashed line

$\alpha > 0$), ensuring positive partition function (4) and real pressure (5). The next two bands are complex. For $|q| < q_c \approx 0.36$, they exhibit opposite imaginary parts (not shown), but turn real at $|q| > q_c$. The next two bands are real (cf. Fig. 1*b*). The higher bands form an alternating sequence of two real and two complex bands. For larger values of α, there is a sequence of entirely complex narrow bands (cf. Fig. 1*d*).

Figure 3 shows normalized spectra for several different combinations of charges on the complex energy plane of u, Eq. (11), at the large concentration $\alpha = 200$. One may notice an odd number $n_1 + n_2$ or $n_1 + n_2 - 1$ of spectral sequences, consisting of order-$\sqrt{\alpha}$ exponentially narrow bands, seen as points. The central sequence goes along the real axis terminating at the bottom of the spectrum near $u = -1$. The others appear in conjugate pairs terminating near the roots of unity $u = -(1)^{1/(n_1+n_2)}$. Close to the termination points, the band sequences align along the lines pointing towards $u = 1$. Farther away from the termination points, they deviate from these lines and may coalesce.

Although thermodynamics and transport properties of the Coulomb gases are merely determined by the lowest band $\epsilon_0(q)$, below we address the wider spectral properties of Hamiltonians (8), presented in Figs. 1–3. For this, we develop a semiclassical theory that is best suited for the description of exponentially narrow bands present at a large concentration $\alpha \gtrsim 1$.

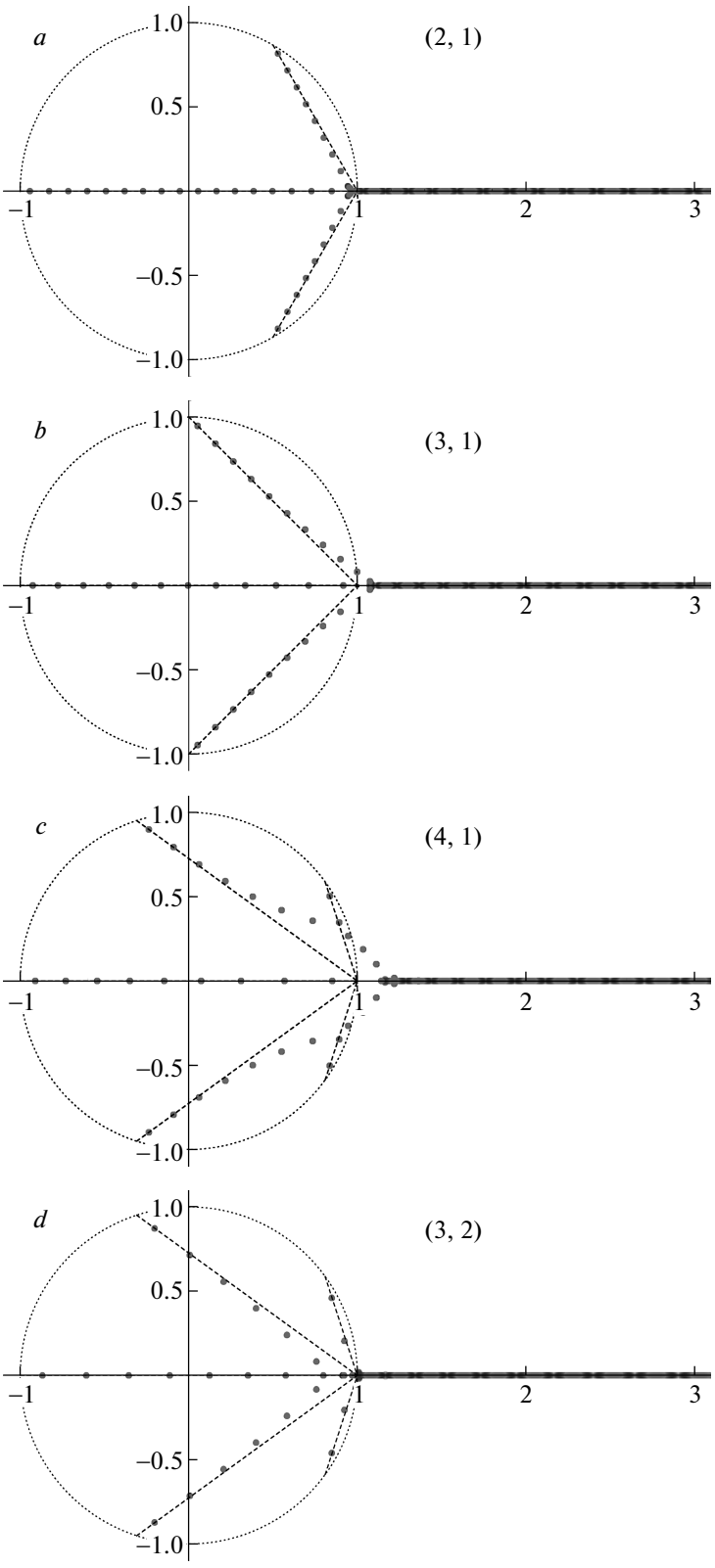

Fig. 3. Complex plane of the normalized energy u, Eq. (11), for $\alpha = 200$ and various valences (n_1, n_2). The dotted circle is $|u| = 1$, the dashed lines connect spectrum termination points $u = -(1)^{1/(n_1+n_2)}$ and $u = 1$, indicating positions of narrow complex bands

4. MONOVALENT (1,1) GAS

To introduce the methods, we first develop a semiclassical spectral theory for the Hermitian Hamiltonian in (8), (9) with $n_1 = n_2 = 1$. For this, we look for wavefunctions in the form $\psi = \exp(i\alpha^{1/2}S)$, where S is an action for the classical problem with the normalized Hamiltonian

$$2u = p^2 - 2\cos\theta, \tag{12}$$

where $u = \epsilon/2\alpha$, such that $u = \mp 1$ correspond to the bottom (top) of the cosine potential. The semiclassical calculations require knowledge of the action integrals. Our approach to such integrals is based on complex algebraic geometry. First, we let $z = e^{i\theta}$ and consider (z, p) as complex variables. Since $p(z)$ resides on the constant-energy hypersurface

$$2u = p^2 - \left(z + \frac{1}{z}\right), \tag{13}$$

we have a family of complex algebraic curves

$$\mathcal{E}_u : \quad \mathcal{F}(p,z) = p^2 z - (z^2 + 2uz + 1) = 0 \tag{14}$$

parameterized by u. For $u \neq \mp 1$, it can be verified that $(\partial\mathcal{F}/\partial z, \partial\mathcal{F}/\partial p)$ does not vanish on $\mathcal{E}_u$, and hence each $\mathcal{E}_u$ is nonsingular. Then $\mathcal{F}(p,z)$ implicitly defines a locally holomorphic map $p = p(z)$. The exceptions to this occur at $z = 0, \infty, z_\pm$, where

$$z_\pm = -u \pm i\sqrt{1-u^2} \tag{15}$$

are the roots of $p^2 = 0$ (i. e., classical turning points). In the vicinity of these four branching points, $p(z)$ behaves as

$$p \sim z^{-1/2} \quad (z \sim 0,) \tag{16}$$

$$p \sim z^{1/2} \quad (z \sim \infty), \tag{17}$$

$$p \sim (z - z_\pm)^{1/2} \quad (z \sim z_\pm), \tag{18}$$

i. e., $p(z)$ is locally double-valued. (Note that we have added a point at $z = \infty$ to the complex plane, thereby rendering it compact and topologically equivalent to a Riemann sphere, Fig. 4). To make sense of this double-valuedness, we first introduce two cuts between the four branching points. For convenience, we have chosen to do so between $0, \infty$ and the turning points $z_\pm$. On this cut domain, $p(z)$ is locally holomorphic.

We then introduce a second sheet of the z-plane and the corresponding Riemann sphere, cut in the same way as the first. We then analytically continue $p(z)$ on the first sheet across the cuts to the second sheet. If $p(z)$ is analytically continued across the branch cut again, we return to the first sphere where we started. In this way, we obtain $p(z)$ as a locally holomorphic function, whose domain is a double-branched cover of the Riemann sphere. Furthermore, suppose we open up the branch cuts, keeping track of where on the other branch $p(z)$ would be if we cross one side of a cut. Identifying these edges, we obtain a torus as in Fig. 5 (where the arrows are used to signify the glued together edges). Thus the complex algebraic curve $\mathcal{E}_u$ can be understood

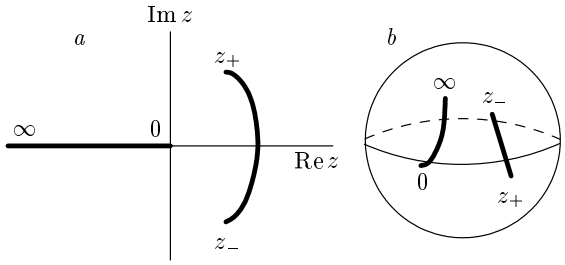

Fig. 4. (a) Complex z-plane with two cuts. (b) It compactifies to the Riemann sphere with two cuts

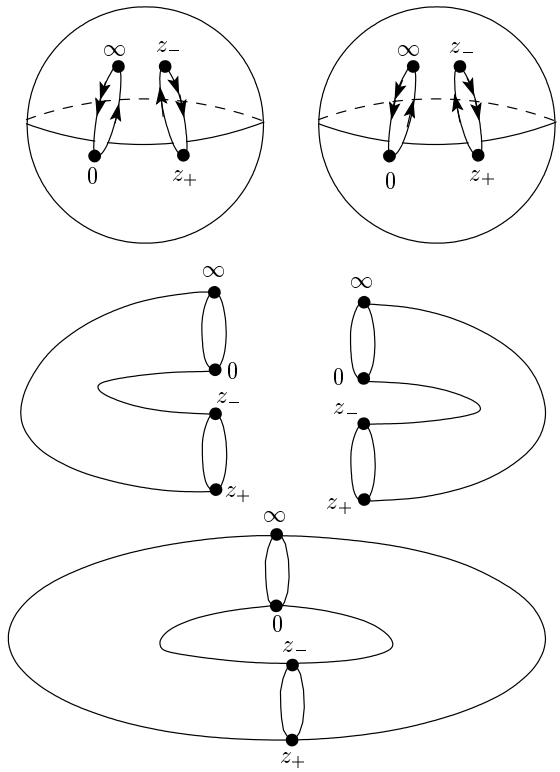

Fig. 5. Construction of a Riemann surface of genus 1. Two Riemann spheres with two cuts each are deformed into tubes to make the gluing in the final step more clear

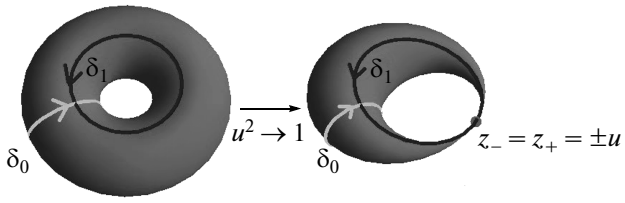

Fig. 6. Riemann surface of genus 1 with two basic cycles δ_0 and δ_1 on it. In the limit $u \to \mp 1$, the torus degenerates into a singular surface. This coincides with the loop δ_0 (but not δ_1) becoming contractible to a point

as a compact Riemann surface of genus $g = 1$ (generally, every compact Riemann surface is topologically a sphere with some number of handles g, called the genus of the surface).

At the exceptional points $u = \mp 1$, the two turning points collide ($z_+ = z_- = \pm 1$) and the branch cut between them collapses. The Riemann surface degenerates into a sphere with two points identified, a singular surface of genus 0. This coincides with one of the loops of the torus becoming contractible to a point (Fig. 6).

4.1. Integration and topology on the torus

The action integrals can be understood as $S = \oint_\gamma \lambda$ over classical trajectories, where

$$\lambda(u) = p(\theta)\, d\theta = p(z)\frac{dz}{iz} = \frac{(z^2 + 2uz + 1)^{1/2}}{iz^{3/2}}\, dz \quad (19)$$

is the action 1-form that is meromorphic on the torus. To visualize the relevant trajectories, we momentarily return to θ and consider it complex. In this representation, there are square-root branch cuts along the real axis, connecting the classical turning points. The action integrals run just above or below the real axis between the turning points. Combining them into closed cycles, we can push these cycles off the real axis and away from the turning points without altering the action integrals (by the Cauchy theorem). The two deformed cycles, shown in Fig. 7, are hereafter called γ_0 and γ_1.

Translating these two cycles to the complex z-plane yields the contours in Fig. 8. We note that these are indeed cycles (i. e., closed contours) owing to the crossing of branch cuts. On the Riemann surface, both wind around the torus. For this reason, the integrals $S_j(u) = \oint_{\gamma_j} \lambda$ are known as periods of $\mathcal{E}_u$ with respect to $\lambda(u)$. It can be verified that the residue of the action form (19) at infinity is zero. Indeed, we have $\lambda \sim dp$

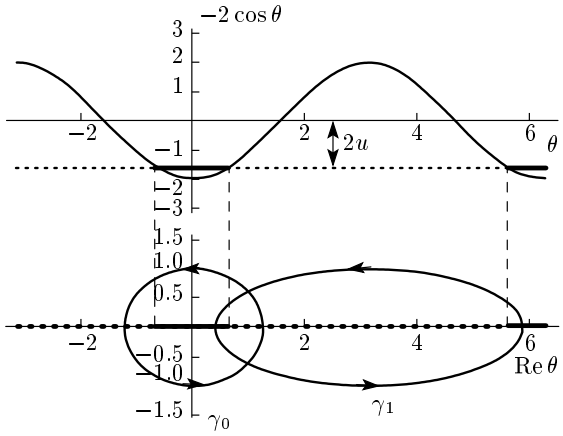

Fig. 7. The classically allowed (forbidden) region at energy $2u$ are shown by the bold solid (dashed) line. A classical (instanton) periodic orbit, in the complex θ-plane, leads to the γ_0 (γ_1) cycle

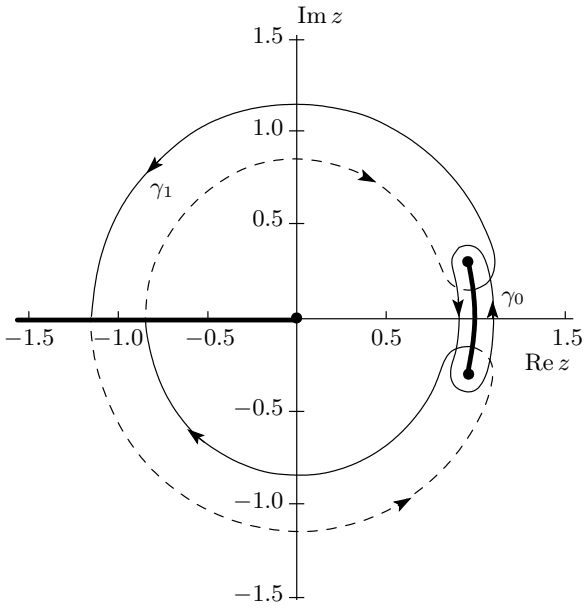

Fig. 8. Cycles γ_0 and γ_1 on the complex z-plane for $u = -0.9$. The cycle γ_1 crosses twice the two cuts from the first branch (solid line) to the second branch (dashed line) and back

at large z. Therefore, we can safely deform the contour around infinity in the z-plane. We consider cycles δ_0 and δ_1 as defined in Fig. 6. Any closed cycle on the torus (after an appropriate deformation) can be decomposed into a superposition of an integer number of these two basic cycles. For example, the cycles γ_0 and γ_1 are

$$\gamma_0 = \delta_0, \quad \gamma_1 = 2\delta_1 - \delta_0. \quad (20)$$

This is evident if we examine the manner in which these cycles encircle around the torus. Formally, the basic cycles generate the first homology group of the torus (since cycles that are alike in this manner are homologous).

We can also consider the first cohomology group of the torus, generated by two independent 1-forms on the Riemann surface modulo exact 1-forms (the latter integrate to zero for all cycles on the torus by Stokes' theorem). In this work, we consider meromorphic 1-forms with zero residues. Modulo exact forms, they are dual to 1-cycles on the torus by the de Rham theorem [29]. The duality implies that there are exactly as many independent 1-forms to integrate upon the surface as independent 1-cycles to integrate along the surface. For the torus, the cohomology, like the homology, is two-dimensional, i. e., any three (or more) 1-forms on the torus are linearly dependent up to an exact form.

4.2. Picard–Fuchs equation

As a result, there must exist a linear combination of 1-forms $\{\lambda''(u), \lambda'(u), \lambda(u)\}$ that is an exact form (here, primes denote derivatives w.r.t. u). This combination can be found by allowing for (u-dependent) coefficients in front of the three 1-forms and seeking an exact form $d_z[P_2(z)/\sqrt{z(z^2+2uz+1)}]$, where $P_2(z)$ is a second-degree polynomial with u-dependent coefficients. Matching the coefficients for powers of z leads to five equations for six unknown parameters, determining the sought combination up to an overall multiplicative factor. This way, we find that the operator $\mathcal{L} = (u^2-1)\partial_u^2 + 1/4$ acts on $\lambda(u)$ as

$$\mathcal{L}\lambda(u) = \frac{d}{dz}\left[\frac{i}{2}\,\frac{1-z^2}{\sqrt{z(z^2+2uz+1)}}\right]. \qquad (21)$$

It follows from Stokes' theorem and the exactness of $\mathcal{L}\lambda(u)$ that $\mathcal{L}S_j(u) = 0$ since γ_j is a cycle on the torus. Therefore, $S_j(u)$ satisfies the linear second-order ODE [16]

$$(u^2-1)S_j''(u) + \frac{1}{4}\,S_j(u) = 0. \qquad (22)$$

This is an example of the Picard–Fuchs equation [30, 31] (see Ref. [32] for a review). Exactly this equation appears extensively in the context of the Seiberg–Witten theory.

Inspecting the coefficient in front of the highest derivative shows that Eq. (22) has regular singular points at $u = \infty$ and $u = \mp 1$, where the torus degenerates into a sphere (see Fig. 6). Changing the variable to u^2, this equation can be brought to the standard hypergeometric form [33]. In the domain $|\arg(1-u^2)| < \pi$, it admits two linearly independent solutions of the form $F_0(u^2)$ and $uF_1(u^2)$, where

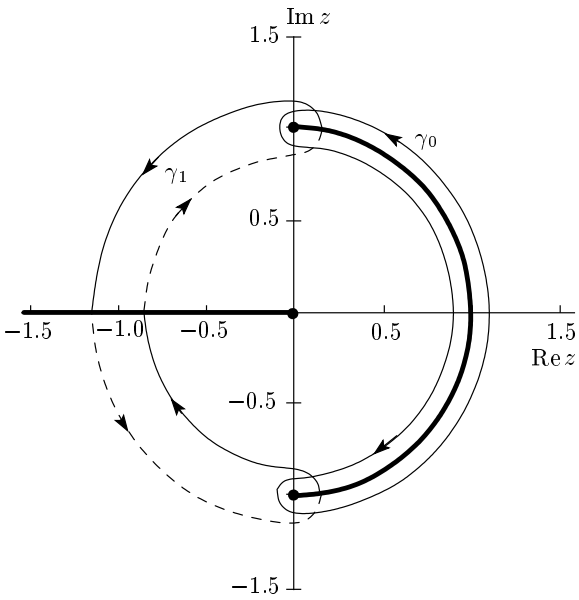

Fig. 9. The two cycles $\gamma_{0,1}$ for $u = 0$. Here, γ_1 can be mapped to γ_0 by rotating through 180°

$$F_0(u^2) = {}_2F_1\left(-\frac{1}{4}, -\frac{1}{4}; \frac{1}{2};\, u^2\right), \qquad (23)$$

$$F_1(u^2) = {}_2F_1\left(+\frac{1}{4}, +\frac{1}{4}; \frac{3}{2};\, u^2\right). \qquad (24)$$

These solutions form a basis out of which $S_j(u)$ (and indeed any period of (14)) must be composed:

$$S_0(u) = C_{00}F_0(u^2) + C_{01}\,uF_1(u^2), \qquad (25)$$

$$S_1(u) = C_{10}F_0(u^2) + C_{11}\,uF_1(u^2). \qquad (26)$$

To find the coefficients C_{jk}, $j,k = 0,1$ appropriate for the action cycles γ_j, we need to evaluate the periods at one specific value of u. Employing the fact that the hypergeometric functions (23)–(24) are normalized and analytic at $u = 0$, i. e., $F_k = 1 + \mathcal{O}(u^2)$, we see that $S_j(u) = C_{j0} + uC_{j1} + \mathcal{O}(u^2)$. Hence, to identify C_{jk}, we expand $S_j(u)$ to the first order in u and evaluate the integrals at $u = 0$. The corresponding cycles in the z-plane are shown in Fig. 9 and explicit calculation yields

$$C_{00} = e^{-i\pi/2}C_{10} = 8\pi^{-1/2}\Gamma(3/4)^2, \qquad (27)$$

$$C_{01} = e^{i\pi/2}C_{11} = \pi^{-1/2}\Gamma(1/4)^2. \qquad (28)$$

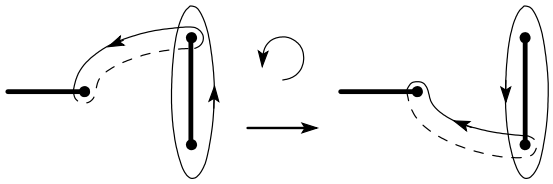

Fig. 10. Monodromy transformation $(u+1) \to (u+1)e^{2\pi i}$ rotates the branch cut between $[z_-, z_+]$ by 180° counter-clockwise. This changes the cycle $\delta_1 \to \delta_1' = \delta_1 - \delta_0$ along with it

The relations between C_{0k} and C_{1k} are not accidental. They originate from the fact that for $u = 0$, the turning points are $\pm i$ and hence the cycle γ_1 transforms into γ_0 under the substitution $z' = e^{-i\pi} z$, Fig. 9. Together with Eqs. (25) and (26), these relations imply global symmetry between the two periods,

$$S_0(u) = e^{-i\pi/2} S_1(e^{i\pi} u). \tag{29}$$

4.3. Structure of $S_j(u)$ near $u = -1$

Equations (23)–(28) fully determine the two actions $S_{0,1}(u)$ in terms of the hypergeometric functions[1)]. We should now relate them to physical observables. For this, we consider the structure of $S_j(u)$ in the neighborhood of $u = -1$. As noted above, the cycle $\gamma_0 = \delta_0$ contracts to a point as $u \to -1$, and therefore $S_0(-1) = 0$ by Cauchy's theorem. By contrast, $S_1(-1)$ remains finite. Moreover, while S_0 is analytic near $u = -1$, it turns out that S_1 is not. To see this, we choose some $u \gtrsim -1$ and allow u to wind around -1 (i. e., $(u+1) \to (u+1)e^{2\pi i}$). Since $u \approx -1$, the roots $z_\pm$ in (15) are of the form $z_\pm = -1 \pm i\sqrt{2(u+1)}$, and we see that this transformation exchanges these branch points via a counter-clockwise half-turn; the branch cut in effect rotates by 180°. For the cycle δ_0, which encloses the turning cut, this has no effect: the cut turns within it. Not so for δ_1: as the cut rotates, we must allow δ_1 to continuously deform if δ_1 is never to intersect the branch points. The overall effect is shown in Fig. 10. The effect of this monodromy transformation is to produce a new cycle δ_1'. Thus, while we have returned to the initial value of u, the period $S_1(u)$ (unlike $S_0(u)$) does not return to its original value and therefore $S_1(u)$ cannot be analytic near $u = -1$.

[1)] Since the integrals considered here are in fact elliptic integrals over a closed cycle, the hypergeometric functions presented here could have been given directly in terms of the complete elliptic integrals of the first and second kind [33].

These facts are consistent, of course, with the origin of the integrals as the classical and instanton actions. As $u \to -1$, the classically allowed region collapses and $p(\theta) \to 0$, and hence the classical action at the bottom of the cosine potential approaches that of the harmonic oscillator $S_0(u) \propto (1+u)$ (indeed, the classical period $T \propto \partial_u S_0$ is a constant). For the instanton trajectory γ_1, the action S_1 does not vanish. Moreover, as $u \to -1$, the period on the instanton trajectory is logarithmically divergent because the trajectory goes to the extrema of the cosine potential (see Fig. 7). This implies that $S_1(u) \propto \text{const} + (1+u)\ln(1+u)$.

In fact, more can be said. Under the monodromy transformation, the basis cycle δ_1' relates to the original basis as $\delta_1' = \delta_1 - \delta_0$ (as can be seen by counting intersections of cycles or by moving onto the torus). Thus, $(\delta_0, \delta_1) \to (\delta_0, \delta_1 - \delta_0)$. From the decomposition of γ_0 and γ_1 noted in (20), it follows that the $S_j(u)$ must transform as

$$\begin{pmatrix} S_0(u) \\ S_1(u) \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} S_0(u) \\ S_1(u) \end{pmatrix} = \\ = M_{-1} \begin{pmatrix} S_0(u) \\ S_1(u) \end{pmatrix}, \tag{30}$$

where we have introduced the monodromy matrix M_{-1} of the actions near $u = -1$. Since this variation of S_1 occurs for every such monodromy near $u = -1$, S_1 must have a component that depends logarithmically on $1 + u$. Indeed, $\ln(1+u)$ increases by $2\pi i$ under the monodromy and since S_1 changes by $-2S_0$ it must have the functional form,

$$S_1(u) = Q_1(u) + \frac{i}{\pi} S_0(u) \ln(1+u), \tag{31}$$

where $Q_1(u)$ and $S_0(u)$ are analytic functions of $1 + u$.

As an immediate corollary, we can use relation (29) between S_0 and S_1 to find the structure of the solution near $u = 1$. Then the functional form of $S_0(u)$ near $u = 1$ is

$$S_0(u) = Q_0(u) - \frac{i}{\pi} S_1(u) \ln(1-u),$$

where $Q_0(u) = -iQ_1(-u)$ and $S_1(u) = iS_0(-u)$ are analytic functions of $1 - u$. The corresponding monodromy matrix is

$$M_1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}. \tag{32}$$

While the structure of the periods near $u = \pm 1$ has been shown through geometric reasoning, it can also

be found directly by seeking solutions of Picard–Fuchs equation (22) as power series in $1 \pm u$. Such a procedure along with the demand of a constant Wronskian leads to a realization that one of the two solutions must include $(1 \pm u)\ln(1 \pm u)$ terms along with the iterative sequence for finding the coefficients of the polynomials. This allows directly verifying Eq. (31).

4.4. Semiclassical results

We now seek semiclassical results for the sequence of low-energy bands terminated at $u = -1$. We interpret the period $S_0(u)$ that is analytic around $u = -1$ as a classical action. It should be quantized according to the Bohr–Sommerfeld rule to determine the normalized energies u_m of the bands,

$$S_0(u_m) = 2\pi\alpha^{-1/2}(m + 1/2), \qquad m = 0, 1, \ldots \quad (33)$$

(we do not discuss the origin of the Maslov index 1/2 here). The second nonanalytic period $S_1(u)$ is identified as the instanton action, which determines the bandwidth $(\Delta u)_m$ according to Gamow's formula

$$(\Delta u)_m = \frac{\omega}{\pi\sqrt{\alpha}} \exp\left(i\alpha^{1/2}\frac{S_1(u_m)}{2}\right), \quad (34)$$

where $\omega = 2$ is the classical frequency for Hamiltonian (12). The monodromy of u around -1, Eq. (30), carries over to the bandwidth as the factor $\exp[(i/2)\alpha^{1/2}(-2S_0(u_m))]$. Then the Bohr–Sommerfeld quantization in (33) is also a condition for the bandwidth to be invariant with respect to monodromies.

To illustrate these results, we expand the periods in Eqs. (25)–(28) near $u = -1$ to find the physical energy levels $\epsilon_m = 2\alpha u_m$. To the first order, we find

$$S_0(u) = 2\pi(u + 1), \quad (35)$$

$$Q_1(u) = 16i - \frac{i}{\pi}(u + 1)\ln(32e), \quad (36)$$

implying $\epsilon_m = -2\alpha + 2\alpha^{1/2}(m + 1/2)$. As a result, the pressure of a monovalent gas, Eq. (5), is

$$P = -eE_0\epsilon_0 = 2k_BTf - \sqrt{k_BTeE_0 f}. \quad (37)$$

The two terms here are respectively the pressure of the ideal gas with the fugacity f and the mean-field Debye–Hückel interaction correction [2].

The instanton action, Eq. (31), at the quantized u_m is

$$S_1(u_m) = 16i + 2i\left(m + \frac{1}{2}\right)\ln\left(\frac{m + 1/2}{32e\alpha^{1/2}}\right), \quad (38)$$

where the linear term in $Q_1(u)$ is absorbed into the logarithm. Gamow formula (34) leads to

$$(\Delta\epsilon)_m = 2\alpha(\Delta u)_m = 2\alpha\,\frac{\omega}{\pi\sqrt{\alpha}} \times \exp\left(i\alpha^{1/2}\frac{S_1(u_m)}{2}\right) = \frac{4}{\pi}\left(\frac{32e}{m + 1/2}\right)^{m+1/2} \times \exp\left[-8\alpha^{1/2} + \left(\frac{m}{2} + \frac{3}{4}\right)\ln\alpha\right], \quad (39)$$

This coincides with the known asymptotic results for the Mathieu equation [27, 28, 34].

4.5. Neighborhood of $u = \infty$

For completeness, we also consider the behavior of the actions at high energy. In the limit $u \to \infty$, Picard–Fuchs equation (22) is of the form $u^2S''(u) + S(u)/4 = 0$. Seeking a solution in the form $S = u^r$, we find $r(r-1) + 1/4 = (r - 1/2)^2 = 0$ and thus there must be two independent solutions with the leading behavior $u^{1/2}$ and $u^{1/2}\ln(u)$. Therefore, the two periods should be of the form

$$S_i(u) = u^{1/2}\left[V_i(u) + W_i(u)\ln u\right], \quad (40)$$

where W_i and V_i are analytic functions of $1/u$. To find these functions, we note that while the continuation to infinity for S_1 is unambiguous, the result obtained for S_0 depends on whether the path to infinity passes above or below $u = 1$. This is because S_0 exhibits a nontrivial monodromy around $u = 1$, Eq. (32). In other words, whether u goes to infinity below or above the real axis determines which of the two turning points $z_\pm$ goes to zero or infinity. Since these are also branching points for the torus, the path of analytic continuation determines how the cycles on the torus are carried along in the process.

Thus, looking for the asymptotic behavior of periods (25)–(28) at $u \to \infty \pm i0$, we finds [34]

$$V_0(u) = i\pi W_1(u) \mp V_1(u), \quad (41)$$

$$W_0(u) = \mp W_1(u), \quad (42)$$

$$V_1(u) = 4i\sqrt{2}\left[\ln\left(e^2/8\right) + 2/u\right], \quad (43)$$

$$W_1(u) = -4i\sqrt{2}\left[1 - (4u)^{-2}\right] \quad (44)$$

to leading corrections in $1/u$. Since $S_0(u) \pm S_1(u) = i\pi W_1(u)u^{1/2}$, it readily follows that under the monodromy $u \to ue^{2\pi i}$, the two actions transform with the monodromy matrices

$$M_{\infty - i0} = \begin{pmatrix} -3 & 2 \\ -2 & 1 \end{pmatrix}, \quad M_{\infty + i0} = \begin{pmatrix} 1 & 2 \\ -2 & -3 \end{pmatrix}. \quad (45)$$

It can be verified that the three monodromy matrices satisfy

$$M_{\infty-i0} = M_1 M_{-1}, \quad M_{\infty+i0} = M_{-1} M_1, \tag{46}$$

as expected [33]: winding around 0 in a large counterclockwise circle is the same as winding -1 and 1 sequentially counterclockwise.

From Eqs. (40)–(44), we find the unique nonsingular period at $u \to \infty \pm i0$ to be given by $S_0(u) \pm \pm S_1(u) = -i\pi W_1(u) u^{1/2}$. As discussed above, it must be identified with the classical action and subject to the Bohr–Sommerfeld quantization

$$[S_0(u_m) \pm S_1(u_m)]/2 = 2\pi\alpha^{-1/2} m.$$

This leads to $u_m \approx m^2/2\alpha$ and hence $\epsilon_m = 2\alpha u_m = m^2$, as expected for the high-energy spectrum.

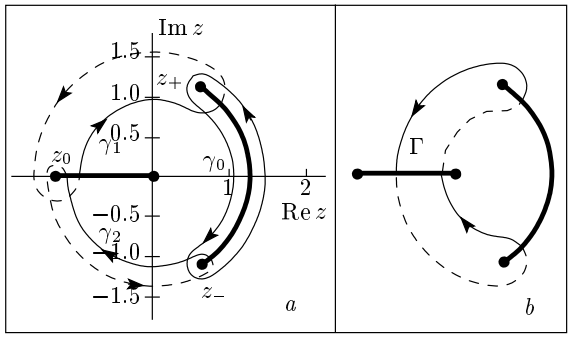

Fig. 11. Complex z-plane with two branch cuts, shown in bold solid lines. (a) Three integration cycles $\gamma_0, \gamma_1, \gamma_2$ are displayed for $u = 0$. (b) The instanton cycle $\Gamma = -\gamma_1 + \gamma_2$. The solid (dashed) lines denote parts of the cycles going over the first (second) branch

5. DIVALENT (2,1) GAS

The divalent (2,1) gas is the simplest case where Hamiltonian (8) is non-Hermitian. In terms of the complex variable $z = e^{i\theta}$ and normalized energy $u = 2\epsilon/3\alpha$, it takes the form

$$\frac{3}{2}\, u = p^2 - \left(\frac{z^2}{2} + \frac{1}{z}\right). \tag{47}$$

Similarly to Eq. (13), this defines a family of complex algebraic curves

$$\mathcal{E}_u: \quad \mathcal{F}(p,z) = 2p^2 z - \left(z^3 + 3uz + 2\right) = 0. \tag{48}$$

The map $p = p(z)$ is locally holomorphic away from the zeros $z_0, z_\pm$ (Fig. 11). At these three branching points as well as at the singularity at $z = 0$, the function $p(z)$ is locally double-valued and behaves as $p \sim (z - z_j)^{1/2}$, $j = 0, \pm$ and $p \sim z^{-1/2}$, respectively. In contrast to the monovalent $(1,1)$ case in Sec. 4, the function $p(z)$ is single-valued at $z \sim \infty$, where it behaves as $p \sim z$, and hence no branch cut extends to $z = \infty$. Nevertheless, there are again four branching points. To construct the Riemann sphere, we draw two branch cuts: one between $[0, z_0]$ and the other between $[z_+, z_-]$. The resulting Riemann surface is again a $g = 1$ torus, analogous to that in Fig. 5.

Its moduli space u contains four singular points $u = -1, e^{\pm i\pi/3}$, and $u = \infty$, where the torus degenerates into the sphere. (There were only three such points in the (1,1) case.) For $u = -1$, the branching points $z_\pm$ coalesce, while for $u = e^{\pm i\pi/3}$, the branching point z_0 collides with $z_\pm$, correspondingly. As $u \to \infty$, the branching point z_0 approaches $z = 0$, while $z_\pm \to \pm i\infty$.

The action integrals are again defined as $S_j = \oint_{\gamma_j} \lambda$, where the 1-form $\lambda(u) = p(z)dz/iz$ is meromorphic on the torus. In general, the counterparts of the turning points in the complex θ-plane are not real. This makes it more convenient to discuss the action cycles γ_j in the z-plane. With three turning points $z_0, z_\pm$, it is convenient to take three paths of integration $\gamma_0, \gamma_1, \gamma_2$, depicted in Fig. 11. In terms of the two basic cycles δ_0 and δ_1 on the torus (see Fig. 6), the three paths are given by

$$\gamma_0 = \delta_0, \quad \gamma_1 = -\delta_1 + \delta_0, \quad \gamma_2 = \delta_1. \tag{49}$$

We note that $\gamma_0 - \gamma_1 - \gamma_2 = 0$, and hence $S_0 = S_1 + S_2$. This equality holds because there are only two independent closed cycles on Riemann surface of genus 1. It follows from de Rham's theorem [29] that there are exactly two independent 1-forms. Therefore, the three forms $\{\lambda''(u), \lambda'(u), \lambda(u)\}$ are linearly dependent up to an exact form. Following the root outlined in Sec. 4.2 (where $P_2(z)$ is replaced with $P_3(z)$, a polynomial of degree 3), we obtain the Picard–Fuchs equation

$$(u^3 + 1) S_j''(u) + \frac{u}{4}\, S_j(u) = 0. \tag{50}$$

In agreement with the above discussion, there are regular singular points at the third roots of negative unity, i. e., $u = -1, e^{\pm i\pi/3}$, where the coefficient in front of the highest derivative goes to zero, and at $u = \infty$. Two linearly independent solutions $F_0(u^3)$ and $uF_1(u^3)$ of this second-order ODE are given in terms of the hypergeometric functions

$$F_0(u^3) = {}_2F_1\left(-\frac{1}{6}, -\frac{1}{6}; \frac{2}{3}; -u^3\right), \quad (51)$$

$$F_1(u^3) = {}_2F_1\left(+\frac{1}{6}, +\frac{1}{6}; \frac{4}{3}; -u^3\right). \quad (52)$$

In this basis, the three periods $S_j(u)$, where $j = 0, 1, 2$, are given by

$$S_j(u) = C_{j0}F_0(u^3) + C_{j1}uF_1(u^3). \quad (53)$$

Since the hypergeometric functions $F_j(u^3 \to 0) = 1 + \mathcal{O}(u^3)$, it follows that $S_j(u) = C_{j0} + uC_{j1} + \mathcal{O}(u^3)$ as $u \to 0$. We can thus find constants C_{jk} by explicitly evaluating the actions at $u = 0$, i. e., $C_{j0} = S_j(0)$ and $C_{j1} = S_j'(0)$. The corresponding integration paths are shown in Fig. 11 and straightforward integration yields

$$C_{00} = C_{10}e^{\pi i/3} = C_{20}e^{-\pi i/3} = \frac{2^{11/6}3\pi^{3/2}}{\Gamma(1/6)\Gamma(1/3)}, \quad (54)$$

$$C_{01} = C_{11}e^{-\pi i/3} = C_{21}e^{\pi i/3} = \frac{3^{1/2}\Gamma(1/6)\Gamma(1/3)}{2^{11/6}\pi^{1/2}}. \quad (55)$$

These relations along with Eq. (53) imply the three-fold symmetry between the actions (cf. Eq. (29))

$$S_0(u) = e^{i\pi/3}S_1\left(e^{-2i\pi/3}u\right) = e^{-i\pi/3}S_2\left(e^{2i\pi/3}u\right). \quad (56)$$

We now need to connect periods (53) with the quantum spectrum. We start by discussing the real branch of the spectrum terminating at the singular point $u = -1$ (see Fig. 1). As $u \to -1$, the two branching points $z_\pm$ coalesce. As a result, the γ_0 cycle degenerates to a point, leading to $S_0(u \to -1) \to 0$, while $S_{1,2}$ remain finite and actually turn out to be nonanalytic. This can be seen by considering the monodromy for a winding of u around -1, i. e., $(u+1) \to (u+1)e^{2\pi i}$ (cf. Sec. 4.3). Such a transformation exchanges branching points $z_\pm$ by a counter-clockwise 180° rotation. This leaves the cycle $\delta_0 = \gamma_0$, which encloses these two points, unchanged. On the other hand, the cycle δ_1 picks up a contribution of $-\delta_0$: $\delta_1' = \delta_1 - \delta_0$. Thus $\gamma_{1,2}$, Eq. (49), pick up a contribution of $\pm\delta_0$. As a result, for every monodromy cycle, $S_{1,2}$ pick up a contribution of $\pm S_0$, and therefore locally they are of the form

$$S_{1,2}(u) = Q_{1,2}(u) \mp \frac{i}{2\pi}S_0(u)\ln(1+u), \quad (57)$$

where $Q_{1,2}(u)$ and $S_0(u)$ are analytic functions of $1+u$ (moreover $Q_1 + Q_2 = S_0$, cf. Eq. (49)). This allows identifying the period $S_0(u) = (\sqrt{6}\pi/2)(1+u) + \mathcal{O}((1+u)^2)$ as the classical action, while the instanton action is a combination of the two nonanalytic periods $S_{1,2}(u)$.

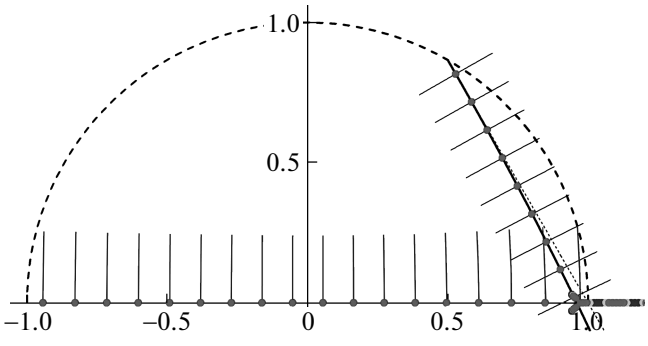

Fig. 12. Narrow energy bands (∘) in the upper half-plane of the complex energy u for $\alpha = 200$ (cf. Fig. 3*a*). $\mathrm{Im}\, S_0(u) = 0$ along the real axis, where the small lines mark $\mathrm{Re}\, S_0(u) = 2\pi\alpha^{-1/2}(m + 1/2)$. The line $\mathrm{Im}\, S_1(u) = 0$ emerges from $u = e^{i\pi/3}$ and intersects the real axis at $u \approx 0.96$. To the right of this point, we observe bands with narrow gaps and use the same coloring convention as in Figs. 1, 3. The small perpendicular lines mark $\mathrm{Re}\, S_1(u) = 2\pi\alpha^{-1/2}(m + 1/2)$

The corresponding monodromy matrix M_{-1}, e. g., in the basis (S_0, S_1) (since $S_2 = S_0 - S_1$ is linearly dependent) is

$$\begin{pmatrix} S_0(u) \\ S_1(u) \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} S_0(u) \\ S_1(u) \end{pmatrix} = M_{-1}\begin{pmatrix} S_0(u) \\ S_1(u) \end{pmatrix}. \quad (58)$$

Employing Eqs. (49) and (56), we find that at the singular point $e^{i\pi/3}$ ($e^{-i\pi/3}$), the period $S_1(u)$ ($S_2(u)$) is nonsingular and goes to zero. It should be thus identified with the classical actions for the branch of the spectrum terminating at the respective singular point (see Fig. 1). A combination of the remaining two actions S_0 and S_2 (S_1) form the corresponding instanton. The respective monodromy matrices (again in the basis (S_0, S_1)) are found as

$$M_{e^{i\pi/3}} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \quad M_{e^{-i\pi/3}} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}. \quad (59)$$

To find positions of the bands along the three branches of the spectrum, terminating at the three singular points $u = -1, e^{\pm i\pi/3}$, we use the Bohr–Sommerfeld quantization for the proper classical action $S_j(u)$ with $j = 0, 1, 2$, correspondingly:

$$S_j(u_m^{(j)}) = 2\pi\alpha^{-1/2}(m + 1/2), \quad m = 0, 1, \ldots \quad (60)$$

Figure 12 shows the lines $\operatorname{Im} S_0(u) = 0$ and $\operatorname{Im} S_1(u) = 0$ intersected with the set of lines $\operatorname{Re} S_j(u) = 2\pi\alpha^{-1/2}(m+1/2)$. The numerically computed spectrum sits right at the semiclassical complex energies $u_m^{(j)}$. The excellent agreement holds all the way up to the point $u \approx 0.96$, where all three periods S_j happen to be purely real. Beyond this point, the semiclassical approximation seems to break down, which manifests, e. g., in the appearance of wide Bloch bands. Expanding $S_0(u)$ near $u = -1$, we finds for the energy levels $\epsilon_m = 3u_m^{(0)}\alpha/2$ in the semiclassical approximation, $\epsilon_m \approx -3\alpha/2 + \sqrt{6\alpha}(m+1/2)$. The corresponding pressure $P = -eE_0\epsilon_0$ in (5) consists of two contributions: that of the ideal $(2,1)$ gas and of the mean-field Debye–Hückel interaction correction.

Taking into account that there is no physical difference between S_1 and S_2 and that the monodromy around $u = -1$ in Eq. (57) should leave the bandwidth in Gamow's formula (34) invariant (i. e., it adds the factor $\exp\{(i/2)\alpha^{1/2}(-2S_0(u_m^{(0)}))\}$), we identify the instanton cycle with $\Gamma = -\gamma_1 + \gamma_2$ (see Fig. 11): $S_{inst}(u) = -S_1(u) + S_2(u)$. This can also be found by inspecting the cycles in Fig. 11: we see that the combined $\Gamma = -\gamma_1 + \gamma_2$ cycle connects $z_\pm$ turning points through the "classically forbidden region", similarly to the γ_1 instanton cycle in the $(1,1)$ case (cf. Fig. 8). However, we do not have a rigorous proof of this fact. Rather, our choice of the integration cycle should be considered as an educated guess, which is verified by the numerics.

Expanding the $S_{1,2}(u)$ actions near $u = -1$ and substituting $u_m^{(0)}$ from the Bohr–Sommerfeld quantization in (60) with $j = 0$, we find the Bloch bandwidths of the central spectral branch (cf. Eq. (34) with $\omega = \sqrt{6}$)

$$(\Delta\epsilon)_m = \frac{3}{2}\alpha(\Delta u)_m = \frac{2\sqrt{6}}{\pi}\left(\frac{36\sqrt{6}e}{m+1/2}\right)^{m+1/2} \times \times \exp\left[-3\sqrt{6\alpha} + \left(\frac{m}{2} + \frac{3}{4}\right)\ln\alpha\right]. \quad (61)$$

Of special interest is the bandwidth of the lowest energy band, due to its direct relation to the transport barrier of the ion channel, Sec. 2. Setting $m = 0$ yields

$$(\Delta\epsilon)_0 \approx 34.14\,\alpha^{3/4}\,e^{-7.35\sqrt{\alpha}}. \quad (62)$$

This is in very good agreement with the numerical simulations (Fig. 13).

Finally, we focus on the behavior at $u = \infty$. The Picard–Fuchs equation is of the form $u^3S'' + uS/4 = 0$. Searching for a solution of the form $S(u) = u^r$ leads to $(r - 1/2)^2 = 0$, signifying two independent solutions with the leading asymptotic behavior $u^{1/2}$ and $u^{1/2}\ln u$. Upon the monodromy transformation $u \to ue^{2\pi i}$, the first of these solutions changes sign, while the second, along with the sign change, picks up a contribution from the first one. Considering the asymptotic forms of $S_{1,2}(u)$, Eq. (53), at $u \to +\infty$, we find the $SL(2,Z)$ monodromy matrix

$$M_\infty = \begin{pmatrix} -1 & 0 \\ 3 & -1 \end{pmatrix}. \quad (63)$$

It can be verified that

$$M_\infty = M_{e^{i\pi/3}} M_{-1} M_{e^{-i\pi/3}}, \quad (64)$$

as it should be: winding once around 0 in a large counterclockwise rotation is identical to winding counterclockwise in sequence around the other three singular points.

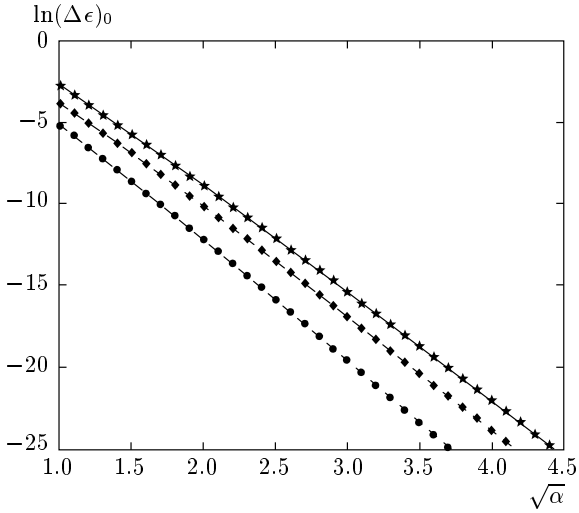

Fig. 13. Analytic (numerical) results for the logarithm of the bandwidth of the lowest band versus square root of the charge concentration with $(1,1)$ as dotted line (circles), $(2,1)$ dashed line (diamonds), and $(3,1)$ as solid line (stars)

6. TRIVALENT (3,1) GAS

The trivalent (3,1) Hamiltonian with the normalized energy u is

$$\frac{4}{3}u = p^2 - \left(\frac{z^3}{3} + \frac{1}{z}\right). \quad (65)$$

It gives a family of algebraic curves

$$\mathcal{E}_u: \quad \mathcal{F}(p,z) = 3p^2z - (z^4 + 4uz + 3) = 0 \quad (66)$$

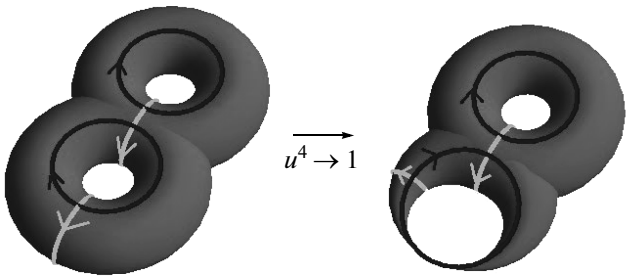

Fig. 14. (a) Double torus curve $\mathcal{E}_u$ for $u^4 \neq 1$, having four basic cycles. (b) When $u^4 = 1$, the $g = 2$ torus degenerates into a singular $g = 1$ surface. This makes one of the basic cycles to pass through the singularity, and renders another cycle contractible to a point

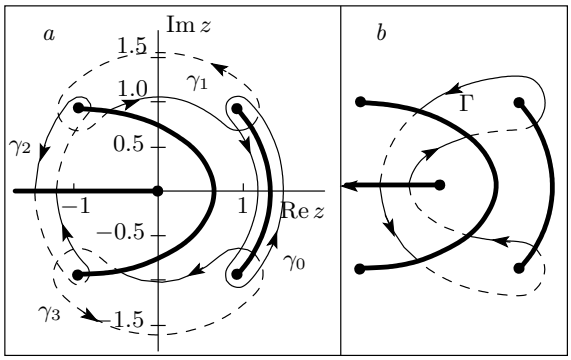

Fig. 15. The Riemann surface is doubly branched with a total of three cuts, shown in bold lines. The four cycles γ_j with $j = 0, 1, 2, 3$, along with the instanton cycle Γ (defined for later reference) are displayed for $u = 0$. The solid (dashed) lines denote parts of the cycles going over the first (second) branch

over complex (z, p). They are nonsingular if $u^4 \neq 1$, and therefore $\mathcal{F}(p, z)$ implicitly defines a locally holomorphic map $p = p(z)$ almost everywhere on (p, z). In this case, there are six square-root branching points at $z = 0, \infty$ and at the four turning points (i. e., four roots of $p^2(z) = 0$).

Hence, while $\mathcal{E}_u$ is a double-branched cover of the Riemann sphere, three cuts (instead of two as in the genus-1 case) are required per branch. After opening up cuts and identifying edges under analytic continuation, this leads to a double torus, i. e., a sphere with two handles, Fig. 14a. Unlike the mono- or divalent cases, the trivalent channel gives a family of genus-2 Riemann surfaces. The exceptional $u^4 = 1$ cases make $\mathcal{E}_u$ singular at $(p, z) = (0, -u)$, due to collision of two turning points, Fig. 14b. The double torus then degenerates into a simple torus with two points identified (a singular surface of genus 1).

As in the genus-1 cases, the actions can be understood as integrals $S_j = \oint_{\gamma_j} \lambda$ of the meromorphic action 1-form $\lambda(u) = p(z)dz/iz$ upon these Riemann surfaces. Owing to the four turning points, there are four such cycles γ_j with $j = 0, 1, 2, 3$. These are chosen as in the divalent case, with the inner arcs of each being taken to start on the principal branch. They are shown for $u = 0$ in Fig. 15a. The u-dependence of these periods is governed by the Picard–Fuchs equation.

Because the double torus is a genus-2 surface, there are four independent cycles (as opposed to two for genus 1). Therefore, the homology — and so too, as argued before, the cohomology — is not two- but four-dimensional: any five meromorphic 1-forms on the double torus are linearly dependent up to an exact form. Thus $\lambda(u)$ and its first four derivatives can be used to produce an exact form; this is done by finding coefficients in a polynomial entering the exact form, as discussed in Sec. 4.2. Stokes' theorem implies that $S(u) = \oint_\gamma \lambda(u)$ must satisfy a 4th-order linear ODE in u, i. e., a Picard–Fuchs equation, which in the present case takes the form

$$(u^4 - 1)S^{(4)} + 8u^3 S^{(3)} + \frac{217}{18}\, u^2 S'' + \\ + uS' + \frac{65}{144}\, S = 0. \quad (67)$$

It has regular singular points at fourth roots of unity, i. e., $u \in \{\pm 1, \pm i\}$ and at $u = \infty$. By changing the variable to u^4, we can cast the Picard–Fuchs equation as a generalized hypergeometric equation. In the cut domain $|\arg(1 - u^4)| < \pi$, it has four linearly independent solutions of the form $u^k F_k(u^4)$, where $k = 0, 1, 2, 3$ and

$$F_0(u^4) = {}_4F_3\left(-\frac{1}{8}, -\frac{1}{8}, \frac{5}{24}, \frac{13}{24}; \frac{1}{4}, \frac{1}{2}, \frac{3}{4}; u^4\right), \quad (68)$$

$$F_1(u^4) = {}_4F_3\left(+\frac{1}{8}, +\frac{1}{8}, \frac{11}{24}, \frac{19}{24}; \frac{1}{2}, \frac{3}{4}, \frac{5}{4}; u^4\right), \quad (69)$$

$$F_2(u^4) = \\ = {}_4F_3\left(+\frac{3}{8}, +\frac{3}{8}, \frac{17}{24}, \frac{25}{24}; \frac{3}{4}, \frac{5}{4}, \frac{3}{2}; u^4\right), \quad (70)$$

$$F_3(u^4) = \\ = {}_4F_3\left(+\frac{5}{8}, +\frac{5}{8}, \frac{23}{24}, \frac{31}{24}; \frac{5}{4}, \frac{3}{2}, \frac{7}{4}; u^4\right) \quad (71)$$

are generalized hypergeometric series. We note that the parameters of each ${}_4F_3(\{a_i\}; \{b_j\}; u^4)$ satisfy $\sum b_i - \sum a_i = 1$; such hypergeometric series are known as one-balanced or Saalschützian [35].

Writing the actions in this basis as

$$S_j(u) = \sum_{k=0}^{3} C_{jk}\, u^k F_k(u^4), \tag{72}$$

we note that $S_j(u) = \sum_{k=0}^{3} C_{jk} u^k + \mathcal{O}(u^4)$ (because generalized hypergeometric functions are unity at zero and analytic nearby). We expand each $S_j(u)$ up to u^3 around $u = 0$ and evaluate the resulting integrals (Fig. 15*a*) to obtain the $\{C_{jk}\}$[2)]. For S_0, for example, this yields

$$C_{00} = +2^{7/2} \cdot 3^{-9/8} \pi^{-1/2} \Gamma(5/8)\Gamma(7/8), \tag{73}$$

$$C_{01} = +2^{-1/2} \cdot 3^{-7/8} \pi^{-1/2} \Gamma(1/8)\Gamma(3/8), \tag{74}$$

$$C_{02} = -2^{-5/2} \cdot 3^{-13/8} \pi^{-1/2} \Gamma(1/8)\Gamma(3/8), \tag{75}$$

$$C_{03} = -7 \cdot 2^{-1/2} \cdot 3^{-27/8} \pi^{-1/2} \Gamma(5/8)\Gamma(7/8). \tag{76}$$

When $u = 0$, the turning points satisfy $z^4 + 3 = 0$ and therefore lie on a certain circle in the complex plane. Hence, γ_j and γ_{j+1} are only different by a $\pi/2$ rotation (Fig. 15*a*). As a result, we find the four-fold symmetry relations

$$S_0(u) = e^{\pi i/4} S_1(e^{-\pi i/2} u) = e^{\pi i/2} S_2(e^{-\pi i} u) = \\ = e^{-\pi i/4} S_3(e^{\pi i/2} u) \tag{77}$$

for u in the cut domain $|\arg(1 - u^4)| < \pi$.

We now consider the periods in the neighborhood of $u = -1$. As before, the cycle γ_0 becomes contractible to a point as $u \to -1$, and therefore $S_0(-1) = 0$ by Cauchy's theorem. The other three actions remain finite, but S_1 and S_3 are nonanalytic. This can be seen by considering the monodromy around $u = -1$. As in the genus-1 cases, the shrinking branch cut near $z = 1$ makes a half-turn. Examining the action cycles, it is only γ_1 and γ_3 that intersect the cut rotating under the monodromy within the γ_0 cycle. Hence, it is these two cycles that change under monodromy and thus have logarithmic nonanalyticity near $u = -1$. More precisely, $(S_1, S_3) \to (S_1 + S_0, S_3 - S_0)$ under the monodromy, and therefore these actions are of the form

$$S_{1,3}(u) = Q_{1,3}(u) \mp \frac{i}{2\pi} S_0(u) \ln(1 + u), \tag{78}$$

where $Q_{1,3}(u)$ as well as $S_0(u)$ and $S_2(u)$ are analytic near $u = -1$. Since $S_1(u) + S_3(u)$ is seen to be invariant under the monodromy, there are a total of three independent periods that have trivial monodromy around $u = -1$. This is again supported by considering series solutions of Picard–Fuchs equation (67) near $u = -1$. This way, we find three regular solutions with the leading behavior $(1 + u)^0$, $(1 + u)^1$, $(1 + u)^2$ along with an irregular solution with the leading behavior $(1 + u)\ln(1 + u)$. For reasons of space, we omit the corresponding 4×4 monodromy matrix.

Although analytic facts about the ${}_4F_3$ series are sparse (see [33, 35] for the relevant discussion), there are simple consistency checks that our solutions (72) must pass. First, the vanishing of the classical action $S_0(u)$ at $u = -1$ implies the identity

$$\sum_{k=0}^{3} C_{0k} (-1)^k F_k(1) = 0 \tag{79}$$

for the hypergeometric functions given above. In addition, inspection of Hamiltonian (65), shows that the classical frequency near $u = -1$ is $\omega = \sqrt{8}$. This implies $S_0'(-1) = (4/3)2\pi/\omega$ and thus

$$\sum_{k=0}^{3} C_{0k} \frac{d}{du} \Big(u^k F_k(u^4) \Big)_{u=-1} = \frac{\sqrt{8}\pi}{3}. \tag{80}$$

Being checked numerically, both relations hold up to 10^{-16}.

Now we turn to the analysis of the spectrum of Hamiltonian (65) at large α. There are three spectral branches terminating at the singular points $u = -1, \pm i$ (see Fig. 3*b*) (notice that the fourth point $u = 1$ lies in the middle of the spectrum and does not have an obvious semiclassical interpretation). To determine positions of the bands, we quantize the corresponding actions $j = 0, 1, 3$ (but not $j = 2$, which is responsible for the period vanishing at $u = 1$) according to the Bohr–Sommerfeld rule:

$$S_j(u_m^{(j)}) = 2\pi\alpha^{-1/2}(m + 1/2), \\ m = 0, 1, \ldots, \quad j = 0, 1, 3. \tag{81}$$

Figure 16 shows the semiclassical energies $u_m^{(j)}$ along with numerically found energy bands. We notes the perfect agreement between these two for $\mathrm{Re}\, u \lesssim 1.09$. At the point $u \approx 1.09$, all three actions $S_{0,1,3}$ are purely real and the corresponding instanton action (see below) goes through zero. Beyond this point, energy bands are not exponentially narrow and the semiclassical approximation may not be applicable. This point is unmistakably different from the singular point $u = 1$. Focusing on the real energies at the bottom of the spectrum and expanding near $u = -1$, we use identities (79) and (80) to find $S_0(u) = (\sqrt{8}\pi/3)(1 + u) + \mathcal{O}(1 + u)^2$.

[2)] Note that the integrals that arise at the u^2-order and higher are divergent near the turning points; however, they are convergent near 0 and ∞ and can be calculated by deforming the contours to run between these points.

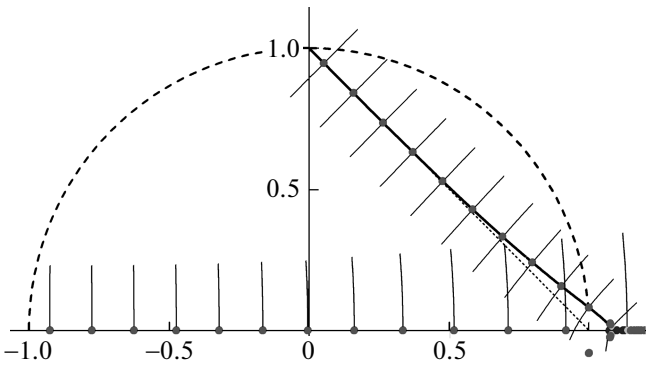

Fig. 16. Narrow energy bands in the upper half-plane of the complex energy u for $\alpha = 200$ (cf. Fig. 3*b*). $\mathrm{Im}\, S_0(u) = 0$ along the real axis, where the short lines mark $\mathrm{Re}\, S_0(u) = 2\pi\alpha^{-1/2}(m + 1/2)$. The line $\mathrm{Im}\, S_1(u) = 0$ emerges from $u = i$ and intersects the real axis at $u \approx 1.09$. To the right of this point, we observe bands with narrow gaps and use the same coloring convention as in Figs. 1, 3. The small perpendicular lines mark $\mathrm{Re}\, S_1(u) = 2\pi\alpha^{-1/2}(m+1/2)$; ∘ dots are the numerically computed narrow bands

Bohr–Sommerfeld rule (81) leads to $\epsilon_m = 4u_m^{(0)}\alpha/3 = -4\alpha/3 + 2\sqrt{2}\alpha^{1/2}(m + 1/2)$. Employing Eq. (5), this yields the pressure of the trivalent Coulomb gas as $P = 4\alpha/3 - \sqrt{2\alpha}$. The two terms here are respectively the ideal gas pressure and the mean-field Debye–Hückel correction.

We now focus on the width of the Bloch bands near $u = -1$. This requires identifying a cycle corresponding to the instanton action. Guided by the cosine potential example (cf. Fig. 7), we take the corresponding cycle as connecting the turning points of the classical action S_0 through the "classically forbidden region". This suggests the cycle Γ shown in Fig. 15*b*, which is essentially of the same form as the γ_1 instanton cycle in $(1, 1)$ case. Considering intersections of these cycles shows that $\Gamma = \gamma_3 - \gamma_2 - \gamma_1$. Upon the monodromy transformation around $u = -1$, the instanton action thus acquires a contribution $-2S_0(u)$, Eq. (78), which leaves the bandwidth invariant thanks to Bohr–Sommerfeld quantization (81). The resulting instanton action is

$$S_{inst}(u) = Q_{inst}(u) + \frac{i}{\pi} S_0(u) \ln(1 + u), \qquad (82)$$

where $Q_{inst} = Q_3 - S_2 - Q_1$ is the regular part of $S_{inst}(u)$ (cf. Eq. (78)). To the first order in $1 + u$, this is $Q_{inst}(u_m) \approx 14.12i - 6.71i(1 + u)$, where, e. g., the leading term originates from

$$Q_{inst}(-1) = S_{inst}(-1) = \sum_{j=0} (C_{3j} - C_{2j} - C_{1j})\,(-1)^j F_j(1) \approx 14.12i.$$

Then, for $u_m^{(0)}$ along the real u-axis satisfying the Bohr–Sommerfeld quantization, Gamow's formula yields the bandwidth

$$(\Delta\epsilon)_m = \frac{4\alpha}{3}(\Delta u)_m = \frac{4\alpha}{3}\,\frac{3\omega}{2\pi\sqrt{\alpha}} \times \exp\left(i\alpha^{1/2}\frac{S_{inst}(u_m)}{2}\right) \approx \frac{4\sqrt{2}}{\pi}\left(\frac{581.14}{m+1/2}\right)^{m+1/2} \times \exp\left[-7.06\sqrt{\alpha} + \left(\frac{m}{2} + \frac{3}{4}\right)\ln\alpha\right]. \qquad (83)$$

The width of the lowest band $(\Delta\epsilon)_0$ is compared with the numerical results in Fig. 13. As in the previous cases, the two results are in strong accord[3)].

For completeness, we address the $u = \infty$ behavior. For large u, the Picard–Fuchs equation is of the form

$$u^4 S^{(4)} + 8u^3 S^{(3)} + 217u^2 S''/18 + uS' + 65S/144 = 0.$$

The trial $S(u) = u^r$ brings four independent solutions with the leading asymptotic forms $\{u^{1/2},\ u^{1/2}\ln(u),\ u^{-5/6},\ u^{-13/6}\}$. The first two are familiar from the genus-1 cases, but the last two are novel to the genus-2 case. The fractional powers proportional to $1/6$ may seem unexpected, given the four-fold symmetries of the periods. However, this symmetry is manifest at the level of cycles at $u = 0$, where four turning points are equally spaced on a circle in the complex z-plane. By contrast, as $u \to \infty$, the turning points must satisfy either $z^3 \sim -u$ or $1/z \sim -u$, thus only three of the four turning points tend towards infinity and one towards zero. This leads to the three-fold exchange of actions upon the monodromy transformation around $u = \infty$. Thus the u^r behavior of the periods with $r = -\text{integer}/(2\cdot 3)$ is exactly what is needed to construct a proper $SL(4, Z)$ monodromy matrix.

7. HIGHER-VALENCE GASES

Here, we briefly summarize our current state of understanding of the higher valence $(4, 1)$ and $(3, 2)$ gases.

[3)] In writing the Gamow formula above, we conjectured the overall preexponential factor of 3 rather than 2 as in the (1,1) and (2,1) cases, possibly due to the different structure of the uctuation determinant. A detailed evaluation of the preexponential factor is beyond the scope of the present work.

The corresponding Hamiltonians are

$$(4,1): \quad \frac{5}{4}\,u = p^2 - \left(\frac{z^4}{4} + \frac{1}{z}\right), \tag{84}$$

$$(3,2): \quad \frac{5}{6}\,u = p^2 - \left(\frac{z^3}{3} + \frac{1}{2z^2}\right). \tag{85}$$

In both cases, there are five turning points in the z-plane given by the equation $p^2(z) = 0$. The behavior at $z = 0$ and $z = \infty$ is somewhat different: for $(4,1)$, there is a branching point at $z = 0$, but not at $z = \infty$ (cf. the (2,1) problem); while for $(3,2)$, the opposite is true: there is no branching point at $z = 0$, but there is one at $z = \infty$. In either case, there are six branching points, which dictate three branch cuts. The resulting Riemann surface is the double torus, as in the $(3,1)$ case (see Fig. 14). In these cases, it is not degenerate as long as $u^5 \neq -1$; otherwise, two of the five turning points collide, leading to a contraction of one of the cycles. Therefore, we expect five branches of the spectrum terminating at $u = (-1)^{1/5}$, in agreement with Figs. 3*c,d*.

Since the Riemann surfaces are of genus 2, there is a linear combination of the 1-form $\lambda(u) = p(z)dz/iz$ and its four u-derivatives that sum up to an exact form. Therefore, any period $S = \oint\lambda$ must satisfy a 4th-order ODE in u. This is found by matching coefficients in a polynomial entering the exact form (see Sec. 4.2), yielding the Picard–Fuchs equations

$$(4,1): \quad (u^5+1)S^{(4)}(u) + \frac{9u^5-1}{u}\,S^{(3)}(u) + \\ + \frac{235}{16}\,u^3S''(u) + \frac{5}{4}\,u^2S'(u) + \frac{39}{64}\,uS(u) = 0, \tag{86}$$

$$(3,2): \quad (u^5+1)S^{(4)}(u) + \frac{9u^5-1}{u}\,S^{(3)}(u) + \\ + \frac{140}{9}\,u^3S''(u) + \frac{5}{4}\,u^2S'(u) + \frac{119}{144}\,uS(u) = 0. \tag{87}$$

While the coefficients seem arbitrary, some features are notable. First, changing the variable to u^5, the equations can be brought to the generalized hypergeometric form; we then find four independent solutions of the form $u^kF_k(u^5)$, where $k = 0,1,2,4$ and F_k are certain ${}_4F_3$ hypergeometric series[4]. We note the absence of a $k = 3$ solution. This can be verified directly from the Picard–Fuchs equations, whose leading behavior near $u = 0$ is given by $S^{(4)}(u) - u^{-1}S^{(3)}(u) = 0$. Substituting $S \propto u^k$ gives $k(k-1)(k-2)(k-4) = 0$.

[4] While we omit the parameters of these series for reasons of space, we note that they satisfy the one-balanced condition [35] stated in the (3,1) case.

Second, we focus on the vicinities of fifth roots of -1, e. g., on $u = -1$. Notably, both Eqs. (86) and (87) have the same leading behavior $5(u+1)S^{(4)}(u) + +10S^{(3)}(u) = 0$, with all other terms being subleading. Looking for a solution in the form $S(u) \sim (1+u)^s$, we find $5s(s-1)^2(s-2) = 0$ for the s-exponent. Therefore, in both cases there are three analytic solutions with the leading behavior $(1+u)^0, (1+u)^1, (1+u)^2$, while the double root at $s = 1$ signifies that the fourth independent solution is of the form $(1+u)\ln(1+u)$[5].

This observation indicates a nontrivial monodromy matrix M_{-1}, allowing one to identify the polynomial in front of the $\ln(1+u)$ with the classical action $S_0(u)$. Being quantized according to Bohr–Sommerfeld, the latter determines the spectrum along the branch terminating at $u = -1$ (see Figs. 3*c,d*).

Finally, we consider the behavior at $u \to \infty$. By taking trial solutions in the form $S(u) \sim u^r$, we obtain 4-th order algebraic equations for the exponent r. The four roots of these equations are $\{1/2,\ 1/2,\ -3/4,\ -13/4\}$ in the $(4,1)$ case and $\{1/2,\ 1/2,\ -7/6,\ -17/6\}$ in the $(3,2)$ case. Remarkably, there is a doubly degenerate root at $r = 1/2$ in both cases, leading to two solutions with the leading asymptotic behavior $u^{1/2}$ and $u^{1/2}\ln(u)$. This was also the case in all the examples considered above. The first of these solutions, being quantized, leads to $\epsilon_m = m^2$, expected at large energies. The other two roots bring two additional solutions with the respective leading behavior $u^{-3/4}, u^{-13/4}$ or $u^{-7/6}, u^{-17/6}$ for the $(4,1)$ and $(3,2)$ cases. The denominators of these fractional powers may be related to the fact that four and three turning points go to infinity as $u \to \infty$ in the two respective cases. The monodromy transformation M_∞ interchanges the corresponding periods (possibly with a sign change). This is achieved by having $-\text{integer}/4$ and $-\text{integer}/(2\cdot 3)$ powers in the corresponding solutions.

8. CONNECTIONS TO THE SEIBERG–WITTEN SOLUTION

Here, we briefly review the main features of the Seiberg–Witten (SW) solution [12, 13], which were adopted in our calculations [36]. The original SW construction gives the spectrum of a four-dimensional supersymmetric $SU(2)$ Yang Mills theory (SYM). The

[5] An existence of 3 = 4–1 analytic solutions near $u = -1$ follows from a theorem on generalized hypergeometric equations going back to Pochhammer [33]; the analogous behavior of the mono-, di-, and trivalent gases near $u = -1$ also provides instances of this theorem.

spectrum of the infrared theory appears to be given by the set of electrically and magnetically charged particles (BPS dyons), which are different from the fundamental particles of the initial UV theory. The latter consists of a vector multiplet transforming in the adjoint representation of $SU(2)$, whose components are one complex scalar field ϕ, a pair of Weyl fermions (gluini), and an $SU(2)$ gauge field (gluon). In the classical UV vacuum, ϕ aligns along the Cartan generator of $SU(2)$ as $\langle\phi\rangle = a\sigma_3/2$, where the complex expectation value a parameterizes the manifold of classical vacua. In the quantum theory, a more convenient coordinate is

$$u = \langle \operatorname{tr} \phi^2 \rangle \tag{88}$$

(such that in the classical limit $u \to \infty$, one has $u \sim a^2$), defining the moduli space $\mathcal{M}_u$ of quantum vacua of the theory.

Given the expectation value a, one defines the generating function (prepotential) $\mathcal{F}(a)$ as a logarithm of the partition function of the theory, restricted by $\langle\phi\rangle = a\sigma_3/2$. It allows introducing a canonically conjugate complex variable

$$a_D = \frac{\partial \mathcal{F}(a)}{\partial a}, \tag{89}$$

where one may regard (a, a_D) as the coordinate and momentum on $\mathcal{M}_u$. The underlying supersymmetry allows arguing that $a(u)$ and $a_D(u)$ are holomorphic functions on the moduli space, except possibly for few isolated singular points. In the UV limit $u \to \infty$, one finds a one-loop correction of the form

$$a_D \sim \frac{ia}{\pi}\left(1 + \ln\frac{a^2}{\Lambda^2}\right), \tag{90}$$

where Λ is a dynamical scale. We recall that $a \sim \sqrt{u}$ in this region. Therefore, when the argument of u changes by $2\pi i$, a changes its sign and a_D transforms as $a_D \to -a_D + 2a$. This rule can be parameterized using the following monodromy matrix in the (a_D, a) basis:

$$M_\infty = \begin{pmatrix} -1 & 2 \\ 0 & -1 \end{pmatrix}. \tag{91}$$

Finding the spectrum of the IR theory means computing masses of particles that are protected by supersymmetry (so-called BPS dyons). The BPS mass formula is

$$M_{n_e,n_m}(u) = |n_e a(u) + n_m a_D(u)|, \tag{92}$$

where (n_e, n_m) are electric and magnetic charges of a dyon; for example, a monopole has $(n_e, n_m) = (0, \pm 1)$. The above relation can be understood semiclassically (at large u) by evaluating the energy functional for the UV theory on the electrically and magnetically charged configurations. The $\mathcal{N} = 2$ supersymmetry guarantees that the very same formula works at a strong coupling as well. There are special loci in the u plane where the masses in (92) vanish. These points can be identified as singularities for a and a_D.

We consider the point u_0 where the monopole becomes massless, $a_D(u_0) = 0$. By a conformal transformation, we can always scale $u_0 = 1$. In the vicinity of this point, a_D behaves as $a_D \propto u - 1$, and hence near this point $a_D(u)$ is holomorphic, while $a(u)$ is expected to be singular. A one-loop calculation similar to the one near $u = \infty$, in the framework of the dual theory, gives a relation similar to (90),

$$a \sim \frac{ia_D}{\pi} \ln \frac{a_D}{\Lambda}. \tag{93}$$

Recalling that $a_D \propto u - 1$, we find the monodromy matrix near $u = 1$, again in (a_D, a) basis:

$$M_1 = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}. \tag{94}$$

From the symmetry considerations one may argue that there should be at least one more singularity in addition to $u = \infty$ and $u = 1$. This follows from the fact that if a singularity exists at some value of u_0, then there ought to be another one at $-u_0$. The $\mathbb{Z}_2$ symmetry, which flips the sign of u, is a result of breaking the global $U(1)$ symmetry (so-called R-symmetry) of the IR action. That symmetry is a remnant of the analogous symmetry in the UV theory, which is common for gauge theories with an extended supersymmetry. It exists on the classical level, but is broken by quantum corrections (both perturbative and instanton) down to $\mathbb{Z}_2$ for $u = \langle \operatorname{tr}\phi^2 \rangle$. Therefore, there are at least three singularities in $\mathcal{M}_u$, e. g., at $u = \infty$ and $u = \pm 1$. The third singular point $u = -1$ corresponds to a massless dyon of unit electric and magnetic charges $a(-1) + a_D(-1) = 0$. The monodromy matrix around it can be computed using the completeness relation $M_1 M_{-1} = M_\infty$ in the complex u-plane.

The nontrivial realization of the SW construction is that complex variables $(a_D(u), a(u))$ with the analytic properties deduced above can be viewed as periods of algebraic curves (tori) $\mathcal{E}_u$, defined over the moduli space $\mathcal{M}_u$, with respect to some meromorphic differential λ_{SW}. The simplest way to parameterize such a curve is

$$\mathcal{E}_u: \quad \mathcal{F}(y,x) = y^2 - (x-u)(x-1)(x+1) = 0, \qquad (95)$$

where x and y are complex. The above equation describes a double cover of the x-plane branched over the four points $x = \pm 1, u$ and $x = \infty$. Moreover, the cover is singular any time two of these points coalesce, i. e., at $u = \pm 1, \infty$, as required. A basis in the first (co)homology of $\mathcal{E}_u$ (2D in this case) is given by integrals of a one-form over one-cycles. We pick the homology basis δ_0, δ_1 in Fig. 6 and the one-form $\lambda_{SW}(u)$ (SW differential) such that

$$a_D(u) = \int\limits_{\delta_0} \lambda_{SW}, \quad a(u) = \int\limits_{\delta_1} \lambda_{SW}. \qquad (96)$$

To pick a proper SW differential $\lambda_{SW}(u)$, we recall that there are only two linearly independent meromorphic 1-forms on the torus up to an exact form. These two forms can be chosen as $\lambda_1 = dx/y$ and $\lambda_2 = x\,dx/y$, whence

$$\lambda_{SW} = \beta_1(u)\frac{dx}{y} + \beta_2(u)\frac{x\,dx}{y},$$

where $\beta_{1,2}(u)$ are functions of u only. The requirement that the period integrals in (96) reproduce the correct asymptotic behavior of $a(u)$ and $a_D(u)$ at $u = 1$ and $u = \infty$ (Eqs. (90) and (93)) allows determining $\beta_{1,2}(u)$. Finally, we obtain

$$\lambda_{SW} = \frac{\sqrt{2}}{2\pi}\,\frac{\sqrt{x-u}}{\sqrt{x^2-1}}\,dx. \qquad (97)$$

This allows evaluating the periods in (96) in terms of elliptic integrals. They in turn yield the entire information about BPS mass spectrum (92) and prepotential (89).

Close parallels to our calculations are apparent. In fact, the SW construction outlined above essentially mirrors the $(1,1)$ gas calculations. Elliptic curve (95) is isogenic to torus (14) and the two SW periods in (96) are directly related to the two action integrals as $S_0 \sim a_D$ and $S_1 \sim a + a_D$. In fact, they can be shown [15, 16] to satisfy exactly the same Picard–Fuchs equation (22) as our actions. Therefore, the two basis solutions (25) and (26), expressible through the complete elliptic integrals of the first and second kind[6)], are also a basis for the SW periods $a_D(u), a(u)$.

An interesting open question is whether our multivalent examples have analogs in SYM theories. For example, the $(2,1)$ case corresponding to a torus with the residual $\mathbb{Z}_3$ symmetry in the u-plane may be related to the $SU(2)$ theory with several fundamental hypermultiplets added. Other examples, leading to $g = 2$ surfaces with $\mathbb{Z}_4$ and $\mathbb{Z}_5$ symmetries, may be related to certain $SU(3)$ SYM theories with matter.

Another captivating observation is related to the peculiar structure of the spectra near $u \approx 0.96$ in the $(2,1)$ gas, $u \approx 1.09$ in the $(3,1)$ gas, etc. These points are marked by the condition $\operatorname{Im} S_1(u)/S_0(u) = 0$, which is reminiscent of wall crossing phenomena in $\mathcal{N} = 2$ theories (for a comprehensive review and references see, e. g., [37]). It is observed that the moduli space $\mathcal{M}_u$ has domains separated by walls such that when "crossing" a wall, the spectrum of the IR theory changes dramatically. For instance, for the $SU(2)$ theory at small $|u|$, there are only two states in the spectrum: a monopole $(0, \pm 1)$ and a dyon $(\pm 1, \mp 1)$. But at large $|u|$, these particles can form bound states with higher electric charges $(n, \pm 1)$ for any integer n. The wall is given by $\operatorname{Im} a_D(u)/a(u) = 0$.

9. DISCUSSION OF THE RESULTS

In this paper we developed a semiclassical treatment for a family of non-Hermitian $\mathcal{PT}$-symmetric Hamiltonians. These Hamiltonians appear upon transfer-matrix mapping of 1D classical statistical mechanics of multivalent Coulomb gases onto quantum mechanics. The low-energy spectra of the Hamiltonians directly translate into thermodynamic and adiabatic transport coefficients of the corresponding Coulomb gases.

We use methods of algebraic topology, traditionally employed in the context of the Seiberg–Witten theory. The main advantage of this strategy is that it allows us to avoid solving equations of motion and finding classical trajectories explicitly. The latter task is rather nontrivial (if at all attainable) in the 4D phase space. Instead, we argue that any constant-energy surface is a 2D Riemann surface of a genus $g \geq 1$. The action along any closed trajectory (not necessarily satisfying equations of motion) can be written as an integer-valued linear combination of $2g$ basic periods of the surface. The periods can be found as solutions of the Picard–Fuchs ODE in the space of parameters. Finally, relations between basic periods and the quantum spectra are established by considering special points in the parameter space, where the surface degenerates into a genus-$(g-1)$ singular surface. Consideration of monodromy transformations in the vicinity of these points allows identifying classical actions, quantized according to Bohr–Sommerfeld, as well as the instanton action, which determines the bandwidth.

The results obtained this way are in excellent agreement with numerical simulations in a broad

[6)] See the footnote [1)] on p.

range of parameters. One of the reasons for this success is that the method provides with preexponetial factors on the same footing with the exponential itself. Another appealing feature of the approach is that none of our semiclassical calculations required the concept of imaginary time. In fact, the "time" (i. e., the 1D coordinate of the Coulomb gas) does not appear at all. In a sense, it is substituted by evolution in the space of parameters of the Hamiltonian (the moduli space). We expect the method to be useful in a broad class of problems that require instanton calculations in complex spaces.

We are indebted to A. Gorsky for introducing us to the algebraic geometry methods and sharing his unpublished notes. The work was partially supported by U.S.–Israel Binational Science Foundation Grant 2008075. Research of P. K. at the Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation.

REFERENCES

1. A. Kamenev, J. Zhang, A. I. Larkin, and B. Shklovskii, Physica A **359**, 129 (2006).
2. J. Zhang, A. Kamenev, and B. Shklovskii, Phys. Rev. E **73**, 051205 (2006).
3. J. Zhang, A. Kamenev, and B. Shklovskii, Phys. Rev. Lett. **95**, 148101 (2005).
4. R. MacKinnon, Angew. Chem., Int. Ed. **43**, 4265 (2004).
5. D. A. Doyle, J. M. Cabral, R. A. Pfuetzner et al., Science **280**, 69 (1998).
6. B. Roux, T. Allen, S. Berneche, and W. Im, Q. Rev. Biophys. **37**, 15 (2004).
7. T. W. Allen, P. S. Andersen, and B. Roux, Proc. Natl. Acad. Sci. USA **101**, 117 (2004).
8. S.-H. Chung and B. Corry, Eur. Phys. J. E **1**, 417 (2005).
9. A. M. Berezhkovskii, M. A. Pustovoit, and S. M. Bezrukov, J. Chem. Phys. **116**, 6216 (2002); **116**, 9952 (2002); **119**, 3943 (2003).
10. S. F. Edwards, and A. Lenard, J. Math. Phys. **3**, 778 (1962).
11. V. G. Vaks, A. I. Larkin, and S. A. Pikin, Zh. Eksp. Teor. Fiz. **53**, 281 (1967) [Sov. Phys. JETP **24**, 240 (1967)].
12. N. Seiberg, and E. Witten, Nucl. Phys. B **426**, 19 (1994) [Erratum-ibid. B **430**, 485 (1994)].
13. N. Seiberg and E. Witten, Nucl. Phys. B **431**, 484 (1994).
14. R. Donagi and E. Witten, Nucl. Phys. B **460**, 299 (1996).
15. A. Gorsky, I. Krichever, A. Marshakov, A. Mironov, and A. Morozov, Phys. Lett. B **355**, 466 (1995).
16. A. Gorsky, *On the Properties of Complex Dynamical Systems* (unpublished).
17. C. E. Rüter, K. G. Makris, R. El-Ganainy, D. N. Christodoulides, M. Segev, and D. Kip, Nat. Phys. **6**, 192 (2010).
18. I. Rotter, J. Phys. A **42**, 153001 (2009).
19. D. L. Huber, and W. Y. Ching, Phys. Rev. B **47**, 3220 (1993).
20. J. C. Tully and C. M. Truesdale, J. Chem. Phys. **65**, 1002 (1976).
21. A. Garg, E. Kochetov, K.-S. Park, and M. Stone, J. Math. Phys. **44**, 48 (2003).
22. E. Kececioglu and A. Garg, Phys. Rev. B **67**, 054406 (2003).
23. A. Altland and B. Simons, *Condensed Matter Field Theory*, Cambridge Univ. Press, Cambridge UK (2007).
24. C. M. Bender, D. C. Brody, and H. F. Jones, Phys. Rev. Lett. **89**, 270401 (2002).
25. C. M. Bender, D. C. Brody, and H. F. Jones, Amer. J. Phys. **71**, 1095 (2003).
26. Y. N. Joglekar and J. L. Barnett, Phys. Rev. A **84**, 024103 (2011).
27. J. N. L. Connor, T. Uzer, R. A. Marcus, and A. D. Smith, J. Chem. Phys. **80**, 10 (1984).
28. J. Meixner and F.W. Schäfke, *Mathieusche Funktionen und Sphaeroidfunktionen mit Anwendungen auf Physikalische und Technische Probleme*, Springer, Berlin (1954).
29. R. Miranda, *Algebraic Curves and Riemann Surfaces*, Amer. Math. Soc., Providence RI (1995).
30. P. A. Griffiths, Ann. Math., Second Series **90**, 460 (1969).
31. P. Deligne, Lecture Notes in Math. **163**, 133 (1970).
32. D. R. Morrison, arXiv:185-199hep-th/9111025 (1991).

33. G. Heckman, *Tshinghua Lectures on Hypergeometric Functions* [http://www.math.ru.nl/heckman/ /tsinghua.pdf (2013)].

34. M. Abramowitz, and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publ., New York (1972).

35. W. Bühring and H. M. Srivastava, in *Approximation Theory and Applications*, Hadronic Press, Florida (1998), p. 17.

36. A. Bilal, arXiv:hep-th/9601007.

37. D. Gaiotto, G. W. Moore, and A. Neitzke, arXiv: 0907.3987[hep-th].