ELSEVIER

Journal of Geometry and Physics 38 (2001) 81–139

JOURNAL OF GEOMETRY AND PHYSICS

Statistical mechanics of $2+1$ gravity from Riemann zeta function and Alexander polynomial: exact results

A.L. Kholodenko

375 H.L. Hunter Laboratories, Clemson University, Clemson, SC 29634-0973, USA

Received 14 June 2000

Abstract

In the recent publication [J. Geom. Phys. 33 (2000) 23], we have demonstrated that dynamics of $2+1$ gravity can be described in terms of train tracks. Train tracks were introduced by Thurston in connection with description of dynamics of surface automorphisms. In this work, we provide an example of utilization of general formalism developed earlier. The complete exact solution of the model problem describing equilibrium dynamics of train tracks on the punctured torus is obtained. Being guided by similarities between the dynamics of two-dimensional liquid crystals and $2+1$ gravity the partition function for gravity is mapped into that for the Farey spin chain. The Farey spin chain partition function, fortunately, is known exactly and has been thoroughly investigated recently. Accordingly, the transition between the pseudo-Anosov and the periodic dynamic regime (in Thurston's terminology) in the case of gravity is being reinterpreted in terms of phase transitions in the Farey spin chain whose partition function is just the ratio of two Riemann zeta functions. The mapping into the spin chain is facilitated by recognition of a special role of the Alexander polynomial for knots/links in study of dynamics of self-homeomorphisms of surfaces. At the end of paper, using some facts from the theory of arithmetic hyperbolic 3-manifolds (initiated by Bianchi in 1892), we develop systematic extension of the obtained results to noncompact Riemann surfaces of higher genus. Some of the obtained results are also useful for $3+1$ gravity. In particular, using the theorem of Margulis, we provide new reasons for the black hole existence in the Universe: black holes make our Universe arithmetic, i.e. the discrete Lie groups of motion are arithmetic.

MSC: 83C45

Subj. Class.: Quantum gravity

Keywords: Surface automorphisms; Gravity; Dynamical systems

E-mail address: string@clemson.edu (A.L. Kholodenko).

PII: S0393-0440(00)00049-8

1. Introduction and summary

The Riemann zeta function $\zeta(\beta)$ has been an object of intensive study in both mathematics [1–3] and physics [4,5] for quite some time. The reason for physicists interest in this function can be easily understood if one writes it in the form of a partition function $Z(\beta)$ given by

$$\zeta(\beta) \equiv Z(\beta) = \sum_{n=1}^{\infty} \exp\{-\beta \ln n\}. \tag{1.1}$$

If β is interpreted as the inverse temperature then, naturally, the following questions arise:

1. What is the explicit form of the quantum mechanical Hamiltonian whose eigenvalues E_n are given by $E_n = \ln n$?
2. Can such system undergo phase transition(s) if one varies the temperature? The goal of providing answers to both questions is at the forefront of current research activities both in physics [4,5] and mathematics. Answers to these questions are being sought in connection with theories of random matrices and quantum chaos [4,5], non-commutative geometry and Yang–Lee zeros [6]. According to the theory of Yang and Lee, the problem of existence of phase transitions can be reduced to the problem of existence of zeros of the partition function in the complex z-plane (where z may be related to either fugacity or the magnetic field, etc.). In the case of $Z(\beta)$, Eq. (1.1), one is also looking at analytic behavior of the partition function in the complex β-plane. Riemann had conjectured that

$$Z(\tfrac{1}{2} + \mathrm{i}t) = 0, \tag{1.2}$$

and $t \in \mathbb{R}$, i.e. all "nontrivial" zeros of the partition function $Z(\beta)$ are located at the critical line $\operatorname{Re}\beta = \frac{1}{2}$. The "trivial" zeros are known [1–3] to be located at $\beta = -2, -4, -6, \ldots$, i.e.,

$$Z(\beta = -2m) = 0, \qquad Z(\beta = -2m + 1) = -\frac{\mathrm{B}_{2m}}{2m} \tag{1.3}$$

for $m = 1, 2, \ldots$ and B_{2m} being the Bernoulli numbers.

Combinations of the Riemann zeta functions are also of physical interest. In particular, in this paper we shall be concerned with the following combination:

$$\hat{Z}(\beta) = \frac{\zeta(\beta - 1)}{\zeta(\beta)} = \sum_{n=1}^{\infty} \phi(n) n^{-\beta}, \tag{1.4}$$

where $\phi(n)$ is the Euler totient function,

$$\phi(n) = n\left(1 - \frac{1}{p_1}\right) \cdots \left(1 - \frac{1}{p_r}\right), \tag{1.5}$$

which is just the number of numbers less than n and prime to n, provided that $n = p_1^{m_1} \cdots p_r^{m_r}$, and p_1, etc. are primes with respect to n. This partition function had appeared in mathematical physics literature in connection with the partition function for the number-theoretic spin chain [7–9] and in connection with calculations of the scattering S matrix for the "leaky torus" quantum mechanical problem [10–12]. Remarkably enough, the results for the number-theoretic spin chain can be obtained as well from earlier works on mode locking and circle maps [13,14], e.g. see in particular [14, Eq. (30)], as the authors of Ref. [9] acknowledge. This fact is not totally coincidental as we shall explain below (in Sections 2 and 3) and has been already anticipated based on our earlier works [15,16] on dynamics of $2+1$ gravity.

In this paper, we would like to demonstrate that the partition function, Eq. (1.4), can adequately describe statistical mechanics of Einsteinian $2+1$ gravity if the underlying surface is the punctured torus. The restriction to the punctured torus case is not too severe and is motivated mainly by illustrative purposes: recall, that both the Seifert surfaces of the figure eight and the trefoil knots are just punctured toruses [17]. This observation allows us to make an easy connection between the dynamics of surface self-homeomorphisms and the associated with its time evolution 3-manifolds which fiber over the circle [16]. These manifolds are just complements of the figure eight and trefoil knots in S^3, respectively. The Seifert surfaces of more complicated knots may naturally be of higher genus but, since both the trefoil and the figure knots belong to the category of fibered knots, only those knots and links which are fibered and the associated with them Seifert surfaces are relevant to the dynamics of $2+1$ gravity [16]. The 3-manifolds associated with the figure eight and the trefoil knots are fundamentally different: the first one is known to belong to the simplest representative of the hyperbolic manifolds while the second corresponds to the so-called Seifert-fibered manifolds [18]. The surface dynamics associated with the first is associated with pseudo-Anosov type of surface self-homeomorphisms while the second one is associated with periodic self-homeomorphisms. Both types of 3-manifolds are topologically very interesting and potentially contain wealth of useful physical information. In this work, we only initiate their study with hope of returning to this subject in future publications.

In order for the reader to keep focus primarily on physical aspects of the problem, we feel, that some simple explanation of what follows is appropriate at this point. To avoid repetitions, we expect that our readers have some background knowledge of the results presented in our earlier published papers [15,16]. In particular, to help our intuition, we would like to exploit the fact that statics and dynamics of $2+1$ gravity is isomorphic with statics and dynamics of textures in two-dimensional liquid crystals. According to the existing literature on liquid crystals, e.g. [19], the liquid crystalline state can be found in several phases which physicists classify as liquid, solid, hexatic and gas. In mathematical literature the textures, e.g. like those in liquid crystals, are known as foliations [18,19]. Dynamics of textures is known accordingly as dynamics of foliations [20]. Some of these foliations may contain singularities. These singularities are mistakenly being treated as Coulombic charges (while in $2+1$ gravity it is well documented [16,21] that these singularities do not interact) since the nonorientable line fields [20] are being confused with the orientable vector fields. The phase transitions in two-dimensional liquid crystals are described in terms of the phase transitions in two-dimensional Coulomb gas. These are known as the Kosterlitz–Thouless-type transitions [22]. Although mathematically such explanation of phase transition is not satisfactory, nevertheless, one has to respect the experimental data associated with the liquid crystalline phases. The extent to which the Kosterlitz–Thouless interpretation of phase transition is appropriate is discussed in Appendix A where it is being argued that, by analogy with transitions in the liquid helium (where the Kosterlitz–Thouless interpretation is normally used) the dynamical phase transition in $2+1$ gravity resemble to some extent the Bose gas condensation type of transition. This analogy is incomplete, however, and it is only being used for the sake of comparison with the existing literature. As Remark 4.1 indicates, the partition function of $2+1$ gravity without any approximations can be recast into the Lee–Yang form [6]. The calculations associated with such form require knowledge of distribution of zeros of the Riemann zeta function and, hence, effectively, the proof of the Riemann hypothesis. Since this proof is not yet available [5], we employ alternative methods associated with recently developed thermodynamic/statistical mechanic formalism for description of phase transitions in the number-theoretic Farey spin chains [7–9]. To prove that dynamical transitions in gravity can be described in terms of transitions in the Farey spin chains several steps are required. In our earlier works [15,16], we have demonstrated that dynamic of $2+1$ gravity is best described in terms of dynamic of train tracks. In Section 2, we demonstrate how dynamic of train tracks can be mapped into dynamic of geodesic laminations. In turn, the dynamic of geodesic laminations is reformulated in terms of the sequence of Dehn twists. This sequence is actually responsible for the dynamic in the Teichmüller space of the punctured torus. Such dynamic is subject to the number-theoretic constraints associated with the Markov triples. The Markov triples had been known in physics literature for a while in connection with the trace maps [23] used for description of quasicrystals, one-dimensional tight band Scrödinger equations, etc. In this work the Markov triples play a somewhat different role: they make the set of closed nonperipheral geodesics on the punctured torus discrete. Mathematically, the sequence of Dehn twists is written in terms of the product of the "right" and the "left" 2×2 Dehn matrices. The modulus of eigenvalues (the stretch factors) associated with such matrix product can be either greater than one or equal to one. In the first case one is dealing with the pseudo-Anosov and in the second, with the periodic (Seifert-fibered) dynamical regime. The results of Section 2 (for the figure eight and the trefoil knots) acquire new meaning in Section 3 where they are reobtained with the help of the associated Alexander polynomials. The stretch factors of Section 2 are reobtained as zeros of the related Alexander polynomials. In the same Section 3, we discuss the fiber bundle construction of 3-manifolds complementary to the figure eight and the trefoil knots in S^3. The sequence of Dehn twists, discussed in Section 2, in this section is being associated with the operation of Dehn surgery (Dehn filing) performed on the 3-manifold related to the figure eight knot. The stretch factors produced as a result of such surgery are reobtained with help of the Mahler measures which allow us to reinterpret these factors in terms of the topological entropies. Introduction of the Mahler measures, in addition, allows us to make direct connection between the dynamical phase transitions and the thermodynamic transitions in the sense of Yang and Lee [6]: zeros of the Alexander polynomial play similar role in dynamics as Yang–Lee zeros in thermodynamics. In Section 4, we provide direct connection between the results of Section 3 just described and the statistical mechanics formalism developed for the number-theoretic Farey spin chains [7–9]. With such connection established, dynamical phase transitions in $2+1$ gravity can be treated with formalism which is more familiar to physicists. Unfortunately, this more familiar formalism is applicable (at least at the present level of our understanding) only to the case of punctured torus. Fortunately, the final results obtained with help of such type of formalism can be reobtained in several different and independent ways. We discuss these alternative ways at the end of Section 4 and in Section 5 which is entirely devoted to development and refinements of these alternatives. The major reason of doing this lies in the opportunity of extension of the punctured torus results to the noncompact Riemann surfaces of any genus. This is accomplished using some results from the scattering theory for Poincaré (Eisenstein) series acting on 3-manifolds. These series had been recently discussed in our earlier work, [24], to which we refer for more details. The key result of Section 4, the partition function for the case of punctured torus, happens to coincide with the scattering S matrix (up to unimportant constant) obtained for 3-manifolds with one $\mathbf{Z} \oplus \mathbf{Z}$ cusp [25]. Three-manifolds containing multiple $\mathbf{Z} \oplus \mathbf{Z}$ cusps are associated with fibered links (as explained in Section 3 and in Appendix C). The S matrix for this case has been also obtained recently [26]. The determinant of this matrix produces the desired exact partition function for $2+1$ gravity. In addition, the formalism allows us to obtain the volumes of the associated 3-manifolds exactly. Extension of the punctured torus results to surfaces of higher genus requires some careful analysis of the nontrivial mathematical problem of circle packing. This fact has some profound impact at development of the whole formalism. It happens, that such type of problems had been comprehensively studied by Bianchi already in 1892 [27] who championed study of the arithmetic hyperbolic manifolds. The notion of arithmeticity is rather involved. Since in physics literature (to our knowledge) it did not find its place yet, in Appendices B and C we supply the essentials needed for the uninterrupted reading of the main text. Surely, selection of the material in these appendices is subjective. But, it is hoped, that interested reader will be able to restore the missing details if it is required. The arithmeticity of 3-manifolds associated with $2 + 1$ gravity stems from some very deep results of Riley [28] and Margulis [29] which we discuss to some extent in Appendix C. The arithmeticity leads to some restrictions on groups of motions in symmetric spaces (e.g. hyperbolic space is symmetric space). Thanks to the Margulis Theorem C.11 and some results of Helgason [30] and Besse [31], the notion of arithmeticity is extendable to $3 + 1$ gravity as well. In the case of $2+1$ gravity we demonstrate, that the very existence of black holes makes such $2+1$ Universe arithmetic. Since most of the Einstein spaces happen to be symmetric, we expect that they are arithmetic in addition in view of the Margulis theorem [29]. This possibility is realized in nature only if the black holes exist in our Universe. The black holes make our Universe arithmetic.

2. From train tracks to geodesic laminations

2.1. Dynamics of train tracks on punctured torus

Dynamics of pseudo-Anosov homeomorphisms on the four punctured sphere was studied in some detail in [32]. Closely related but more physically interesting case is associated with

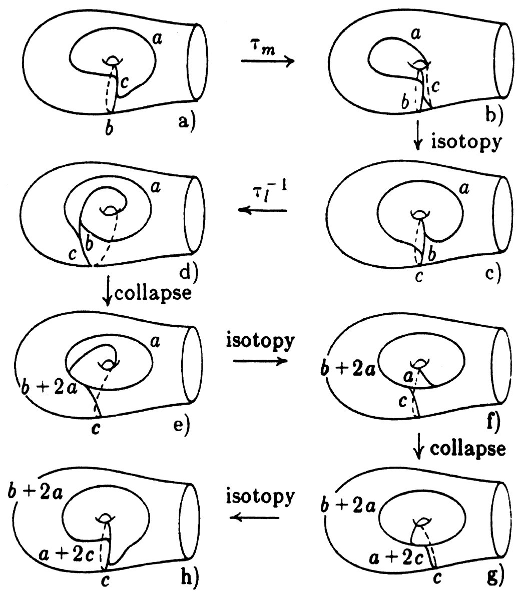

Fig. 1. Fragment of the train track dynamics on once punctured torus.

study of self-homeomorphisms of the punctured torus. The Poincaré–Hopf index theorem requires existence of two Y-type singularities, each having index $-\frac{1}{2}$, as it is explained in Ref. [15]. These singularities can move on the surface of the punctured torus thus giving rise to the train tracks dynamics as depicted in Fig. 1.

From this picture it follows, that the nontrivial dynamics is effectively caused by sequence of meridional τ_{m} and longitudinal τ_{l} Dehn twists. Using the rules set up for dynamics of train tracks [15], one can easily calculate the transition matrix by noticing that topologically the state h) is the same as a) while the weights on the branches are different. This allows us to write the following system of equations:

$$a' = b + 2a, \qquad b' = c, \qquad c' = a + 2c. \tag{2.1}$$

These results can be neatly presented in the matrix form

$$\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}. \tag{2.2}$$

The incidence matrix just obtained can be found in Penner's [33] article, where it was presented without derivation. The largest eigenvalue λ is found to be

$$\lambda = \tfrac{1}{2}(3 + \sqrt{5}). \tag{2.3}$$

Since $\lambda > 1$, this indicates that the dynamics depicted in Fig. 1 is of pseudo-Anosov type. We shall reobtain this result for λ below, e.g. see Eq. (2.16), and in Section 3 using totally different methods. In fact, the presentation above is only given for the sake of comparison with these new methods to be discussed in Section 3. These new methods provide the most natural connections between the dynamics of $2 + 1$ gravity and the theory of knots and 3-manifolds. The remainder of this section is devoted to exposition of some mathematical results which will be used in the rest of this paper.

2.2. The Markov triples

To begin, let us recall [15,16,18], that a *geodesic lamination* on a hyperbolic surface S is a closed subset of S made of union of disjoint simple geodesics. When lifted to the universal cover, i.e. to the Poincaré disc $\mathcal{D}$ whose boundary is a circle S^1_∞ at infinity, the endpoints of the geodesic lamination determine a closed subset (actually, a Mobius strip)

$$\mathcal{E} = \frac{S^1_\infty \times S^1_\infty - \Delta}{Z_2}, \tag{2.4}$$

where Δ is diagonal (x, x), $x \in S^1_\infty$ and the factor Z_2 reflects the fact that the circle segments representing these geodesics are unoriented, i.e. the picture remains unchanged if the ends of each geodesic which lie on S^1_∞ are interchanged. This fact has been discussed and used already in our earlier work, [15]. Since, by definition, the lamination is made of disjoint set of geodesics, when lifted to $\mathcal{D}$, there are no circle segments (representing geodesics) which intersect with each other. It is intuitively clear, that the dynamics of train tracks should affect the dynamics of geodesics. We would like to make this intuitive statement more precise. To this purpose, we notice that, when lifted to the universal cover, this dynamics causes some homeo(diffeo)morphisms of the circle S^1_∞ thus making clear the connections with circle maps and mode locking [13,14].

If $G = \pi_1(S)$ is the fundamental group of surface S, the endpoint subset $\mathcal{E}$ remains invariant under the action of $\pi_1(S)$. Suppose, that the surface S has boundaries, e.g. a hole in the case of a torus. Let $P \sqsubseteq G$ be the set of *peripheral elements*, e.g. those elements of G which correspond to loops freely homotopic to the boundary components. In the case of a punctured torus, G is just a free group of two generators: $G = \langle a, b\rangle$, and the subset P is determined by the commutator $aba^{-1}b^{-1} \equiv [a, b]$ whose trace equals to -2 [34]. Such restriction on the trace indicates that the group element $[a, b]$ is parabolic. Parabolic elements are always associated with cusps (punctures) on the Riemann surface (for more details, e.g. consult Refs. [24,28]). This restriction affects the presentation of the group G. Using generators a and b one can construct a word W_r

$$W_r = a^{\alpha_1} b^{\beta_1} \cdots a^{\alpha_r} b^{\beta_r} \tag{2.5}$$

where α_1 and β_r can be any integers while α_i and β_i, $i \neq 1, r$, can be any integers, except zero. Not all words thus constructed are different. Those words which are conjugate, i.e. $W_r W_l = W_l W_r$, are considered to be equivalent. If we define the equivalence relation as $\sim$, then the quotient $G/\sim$ can be identified with the set Ω of homotopy classes of closed curves on the torus T [34]. Let $\hat{\Omega}$ be a subset of Ω which corresponds to the nontrivial nonperipheral simple closed curves on T. To account for the presence of a puncture, we require that both a and b belong to $\hat{\Omega}$ while $[a, b]$ should correspond to the peripheral simple closed curve(s). The question now arises: how to find an explicit form of the generators a and b? If we require these generators to act by isometries in the Poincaré disc $\mathcal{D}$ (or the upper-half plane), we need to find a mapping of the group G into the group $\mathrm{PSL}(2, \mathrm{R})$. This can be achieved with help of the following identities (in the case of $\mathrm{SL}(2, \mathrm{R})$):

$$2 + \mathrm{tr}[a, b] = (\mathrm{tr}\, a)^2 + (\mathrm{tr}\, b)^2 + (\mathrm{tr}\, ab)^2 - \mathrm{tr}\, a\, \mathrm{tr}\, b\, \mathrm{tr}\, ab, \tag{2.6a}$$

$$\mathrm{tr}\, a\, \mathrm{tr}\, b = \mathrm{tr}\, ab + \mathrm{tr}\, ab^{-1}, \tag{2.6b}$$

which had been discovered by Fricke [35]. The above identities can be analytically extended to $\mathrm{SL}(2, \mathrm{C})$ [34]. The projectivized versions of these groups, that is $\mathrm{PSL}(2, \mathrm{R})$ and $\mathrm{PSL}(2, \mathrm{C})$, are groups of isometries of hyperbolic H^2 and H^3 spaces, respectively. Such an extension is useful for dealing with problems discussed in our earlier work, [24], and will be also discussed later in this work in Section 5.

Since, as we know already, $\mathrm{tr}[a, b] = -2$, the above identities can be conveniently rewritten as

$$x^2 + y^2 + z^2 = xyz, \tag{2.7}$$

$$xy = z + w, \tag{2.8}$$

where, $x = \mathrm{tr}\, a$, $y = \mathrm{tr}\, b$, $z = \mathrm{tr}\, ab$ and $w = \mathrm{tr}\, ab^{-1}$. As it is argued by Bowdich in [36] (who, in turn, attributes it to Jorgensen [37]) the first identity, Eq. (2.7), is sufficient for restoration of the explicit form of matrices a and b. These are given by

$$a = \frac{1}{z}\begin{pmatrix} xz - y & x \\ x & y \end{pmatrix}, \qquad b = \frac{1}{z}\begin{pmatrix} yz - x & -y \\ -y & x \end{pmatrix}. \tag{2.9a}$$

The above matrices differ slightly from those given in [36]. This difference is essential, however, and originates from the fact that we require $\det a = \det b = 1$. Such requirement leads automatically to Eq. (2.7) as required. The above choice of matrices is not unique since, e.g. Eq. (2.7) is also to be satisfied by the choice of

$$a = \frac{1}{z}\begin{pmatrix} y & x \\ x & xz - y \end{pmatrix}, \tag{2.9b}$$

etc. For the integer values of x, y and z the identity, Eq. (2.7), is known as equation for the Markov triples discovered in the number theory by Markov [35]. Given this equation, one is interested to obtain all integer solutions (triples x, y and z). In mathematics literature, sometimes, related equation is known as an equation for the Markov triples [38]. Specifically, one introduces the following redefinitions:

$$x = 3m_1, \qquad y = 3m_2, \qquad z = 3m_3,$$

so that Eq. (2.7) acquires the standard form

$$m_1^2 + m_2^2 + m_3^2 = 3m_1m_2m_3. \tag{2.10}$$

The simplest solution of this equation is just $m_1 = m_2 = m_3 = 1$. To generate other solutions it is sufficient to have a "seed" (m_1, m_2, m_3) which, by definition, obeys the Markov equation (2.10). Then, the first generation of Markov triples is given by

$$(m_1', m_2, m_3), \quad (m_1, m_2', m_3) \quad \text{and} \quad (m_1, m_2, m_3'),$$

where

$$m_1' = 3m_2m_3 - m_1, \qquad m_2' = 3m_1m_3 - m_2, \qquad m_3' = 3m_1m_2 - m_3. \tag{2.11}$$

Other solutions can be generated via mapping: $(m_1, m_2, m_3) \to (-m_1, -m_2, m_3)$ along with cyclical permutation of m_1, m_2 and m_3. If one starts with $(1, 1, 1)$ then, the Markov tree (its part, of course) is depicted in Fig. 2.

Emergence of numbers m_1', etc., can be easily understood if, instead of Eq. (2.10), we would consider the following quadratic form:

$$f(x) = x^2 - 3m_2m_3x + m_2^2 + m_3^2. \tag{2.12}$$

Then, for $f(x) = 0$ we can identify $x = m_1$ and $x' = m_1' = 3m_2m_3 - m_1$. In physics literature the dynamics of Markov triples has actually been studied already in connection with quasicrystals [23], one-dimensional tight binding Schrödinger equations (with quasiperiodic potential), etc. [39], and is known as dynamics of trace maps. This dynamics can be easily understood based on Eqs. (2.10)–(2.12). In short, one studies maps $\mathcal{F}$ of the type

$$\mathcal{F} : \begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} 3yz - x \\ y \\ z \end{pmatrix}, \tag{2.13}$$

etc., which possess an integral of motion

$$I(x, y, z) = x^2 + y^2 + z^2 - 3xyz \tag{2.14}$$

invariant under action of $\mathcal{F}$.

From the theory of Teichmüller spaces [40] it is known, that the length $l(\gamma)$ of closed geodesics associated with $\gamma \in G$ is given by

$$\mathrm{tr}^2(\gamma) = 4\cosh^2(\tfrac{1}{2}l(\gamma)). \tag{2.15}$$

Let, $\mathrm{tr}\,\gamma = x$ (or y, or z), then the Markov triple $(1, 1, 1)$ corresponds to the geodesic whose hyperbolic length $l(\gamma)$ is given by

$$l(\gamma) = 2\cosh^{-1}(\tfrac{3}{2}) = 2\ln(\tfrac{1}{2}(3 + \sqrt{5})) = 2\ln\lambda, \tag{2.16}$$

where in the last equality use had been made of Eq. (2.3). Obtained result provides us with the first piece of evidence that the train tracks are directly associated with closed geodesic laminations. Since in the case of the punctured torus the Teichmüller space coincides with the Poincaré upper-half plane model of hyperbolic space, the Teichmüller distance d_{T} is the same as hyperbolic distance l_{H} [40]. Therefore, we obtain

$$d_{\mathrm{T}} = \tfrac{1}{2}\ln K = l_{\mathrm{H}} = 2\ln\lambda, \tag{2.17}$$

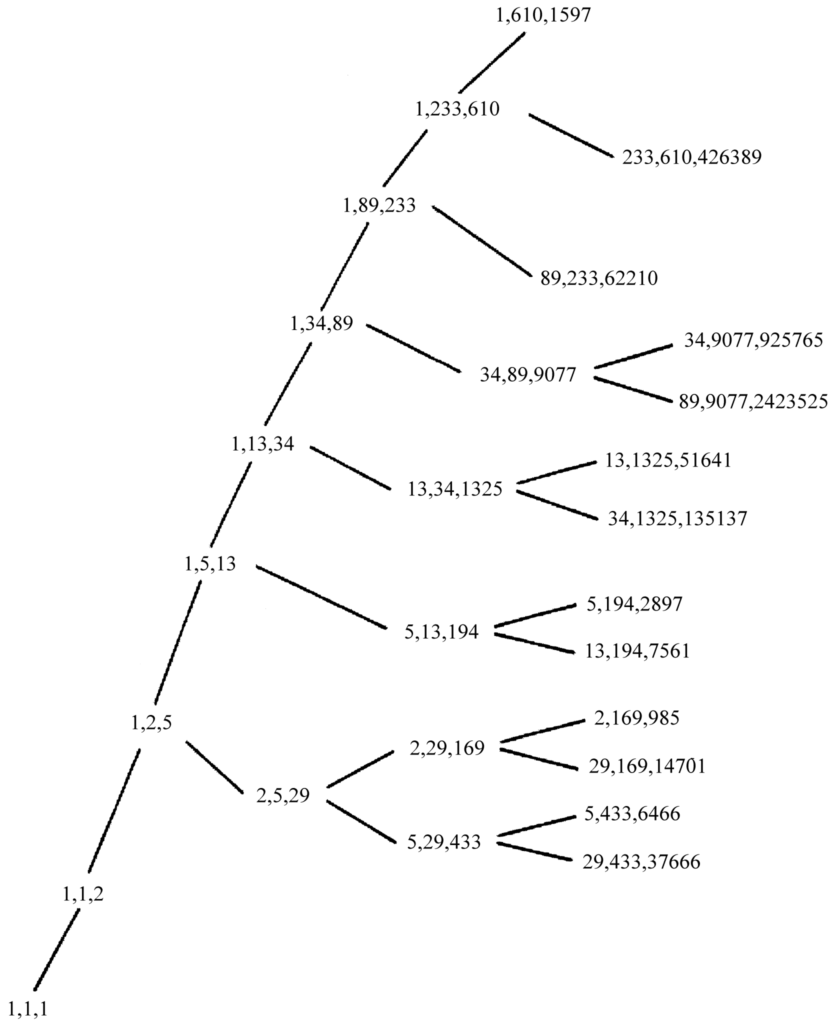

Fig. 2. Fragment of Markov tree close to the original "seed".

which leads to the equation

$$\sqrt{K} = \lambda^2. \tag{2.18}$$

That is the stretching factor λ is associated directly with the Teichmüller dilatation factor K (for more details on K, please, consult our earlier work, [16]).

In the light of previous discussion, we notice that the spectrum of stretching factors is discrete. This is reminiscent already to the energy spectrum of some quantum mechanical system. The question arises at this point: is the stretching factor λ, defined by Eq. (2.3), represents the maximum or the minimum among possible stretching factors? A simple calculation based on Eq. (2.15) and Fig. 2 indicates that λ, given by Eq. (2.15), corresponds to the minimum. In the case of higher genus Riemann surfaces (perhaps with punctures) Penner [41] had demonstrated that the stretching factors are bounded from above and from below.

Let us now obtain the explicit form of matrices a and b, given by Eqs. (2.9a) and (2.9b), for the case when $x = y = z = 3$. An easy calculation produces

$$a = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}, \tag{2.19}$$

yielding the trace of commutator $[a, b]$ being equal to -2 as required. Introduce now two basic matrices

$$L = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad R = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{2.20}$$

The letters stand for the "left" (L) and the "right" (R) Dehn twist matrices (e.g. see Section 3 for more details). Evidently,

$$a = RL \tag{2.21}$$

as can be seen by direct calculation. It is also not difficult to check that

$$b = L^{-1}aR^{-1} = L^{-1}RLR^{-1}. \tag{2.22}$$

Let us now take a note of the fact that

$$L^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \qquad R^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}. \tag{2.23}$$

If we introduce the matrix

$$\hat{I} = \pm \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{2.24}$$

such that

$$\hat{I}^2 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.25}$$

then, Eqs. (2.23) and (2.24) can be conveniently rewritten as

$$L^{-1} = \hat{I}R\hat{I}, \tag{2.26}$$

$$R^{-1} = \hat{I}L\hat{I}, \tag{2.27}$$

which also can be checked by direct calculation. Thus, obtained relations can be equivalently presented as follows:

$$L\hat{I}L = R, \qquad R\hat{I}R = L, \qquad R\hat{I}L = \hat{I}, \qquad L\hat{I}R = \hat{I}. \tag{2.28}$$

Using the results just obtained, it is clear, that every word W, e.g. see Eq. (2.5), can be written in terms of positive powers of L, R and $\hat{I}$. Moreover, since originally we had only matrices a and b, things can be simplified further. In particular, let us consider the word

$$W_1 = L^{\alpha_1} \cdots RL \cdots L^{\alpha_r} R^{\beta_r}, \tag{2.29}$$

and the related word

$$W_2 = L^{\alpha_1} \cdots R\hat{I}L \cdots L^{\alpha_r} R^{\beta_r}. \tag{2.30}$$

We can eliminate the contribution $R\hat{I}L$ in Eq. (2.30) by replacing it with $\hat{I}$ using Eq. (2.28). If we continue to use Eq. (2.28) we can evidently get rid of all $\hat{I}$ factors in the middle of the word W_2 so that in the end we have to consider only the totality of words of the type W_1 with nonnegative integer coefficients. To the totality of these words one has to add words like $\hat{I}W_1$, $W_1\hat{I}$ and $\hat{I}W_1\hat{I}$. Since in Section 4 we shall be interested in the traces of these words, evidently, only words of the type W_1 and $\hat{I}W_1$ need to be considered. Incidentally, use of Eq. (2.28) allows us to reduce b, Eq. (2.22), to $b = \hat{I}LR\hat{I}$ and, hence, from now on we shall use $a = RL$ and $b = LR$. Such choice is in accord with that known in the literature [42,43]. Moreover, to make connections with the results from knot theory and 3-manifolds (to be discussed in Sections 3 and 5) only words of the type W_1 should be considered. This peculiarity will be explained further below.

2.3. *The Farey numbers and the Farey tesselation of H^2*

Words of the type W_1 are represented by the set of matrices M,

$$M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \qquad \alpha\delta - \gamma\beta = \pm 1 \tag{2.31}$$

with integer coefficients. Such matrices belong to the group $\mathrm{GL}(2, \mathrm{Z})$ and play an important role in the number theory [44]. In the number theory they are associated with the Farey numbers. Recall that the Farey series of numbers $\mathfrak{F}_n$ is the ascending series of irreducible fractions between 0 and 1 whose denominators do not exceed n. If p/q and r/s are two consecutive terms (neighbors) of $\mathfrak{F}_n$, then

$$ps - qr = \pm 1. \tag{2.32}$$

In addition, if p/q, r/s and h/k are three consecutive terms, then the mediant r/s is obtained via

$$\frac{r}{s} = \frac{p \pm h}{q \pm k}. \tag{2.33}$$

Hence, α/γ and β/δ in the matrix M correspond to the neighbors in $\mathfrak{F}_n$. Given this information, one can make two additional steps. First, one can let n go to $\pm\infty$ [45]. Thus,

$$\mathfrak{F}_1 : -\infty, -1, 0, 1, \infty, \qquad \mathfrak{F}_2 : -\infty, -2, -1, -\tfrac{1}{2}, 0, \tfrac{1}{2}, 1, 2, \infty, \text{etc.}$$

Second, by noticing that projectivisation $\mathrm{PSL}(2, \mathrm{R})$ of $\mathrm{SL}(2, \mathrm{R})$ is just an isometry of H^2 it is advantageous to map conformally $(-\infty, \infty)$ into S^1_∞. Once this is done, locations of the Farey numbers on S^1_∞ acquire new geometrical meaning (see also Section 5.1). To understand its significance, let us consider the Möbius-like transformations associated with matrices L and R, i.e.

$$L : z' = \frac{z}{z+1}, \qquad R : z' = z + 1. \tag{2.34}$$

Let us notice first that for L-type transformation

$$z = 0 \to z' = 0, z = \pm\infty \to z' = 1, \qquad z = -1 \to z' = -\infty. \tag{2.35a}$$

At the same time, for R-type transformation we have

$$z = 0 \to z' = 1, z = \pm\infty \to z' = \pm\infty. \tag{2.35b}$$

Using these simple results, we obtain at once

$$\begin{aligned} &b(\infty) = LR(\infty) = 1, \qquad a(0) = RL(0) = 1, \\ &a(-1) = RL(-1) = -\infty, \qquad b(-1) = LR(-1) = 0. \end{aligned} \tag{2.36}$$

It is instructive to depict these results graphically, e.g. see Fig. 3.

Points $-\infty, -1, 0, 1, +\infty$ on the circle S^1_∞ are just the members of $\mathfrak{F}_1$. These are joined by the circular arcs (not to be mistaken for the hyperbolic geodesics which are going to be discussed below). The arrows are in accord with the results given by Eq. (2.36). Identifying sides of the polygon using the arrows depicted in Fig. 3 we obtain an orbifold known in physics literature as "leaky torus" [12]. To construct the next level of

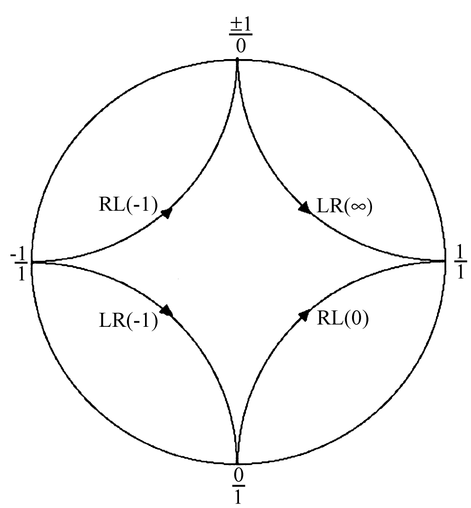

Fig. 3. Leaky torus in the Poincaré disc $\mathcal{D}$ model.

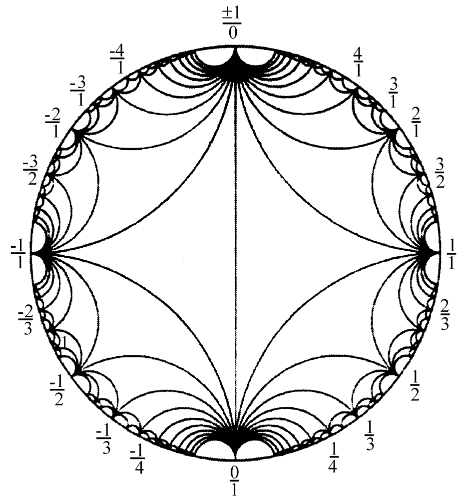

Fig. 4. Farey tesselation of $\mathcal{D}$.

Farey numbers, $\mathfrak{F}_2$, and, for this matter, $\mathfrak{F}_n$, etc., let us consider the most general modular transformation

$$z' = \frac{az+b}{cz+d} \equiv M(z), \tag{2.37}$$

where the integer coefficients a, b, c, and d are subject to the constraint $ad - bc = 1$ (to be compared with Eqs. (2.31) and (2.32)). Using this transformation, we obtain

$$z = \pm\infty \to z' = \frac{a}{c}, \qquad z = 0 \to z' = \frac{b}{d}, \qquad z = \pm 1 \to z' = \frac{\pm a + b}{\pm c + d}. \tag{2.38}$$

But the numbers a/c and b/d are just the Farey neighbors! Using Eq. (2.33), we obtain the following sequence of numbers:

$$-2 = \frac{-1-1}{0+1}, \qquad -\frac{1}{2} = \frac{-1+0}{1+1}, \qquad \frac{1}{2} = \frac{0+1}{1+1}, \qquad 2 = \frac{1+1}{0+1}. \tag{2.39}$$

If we place these numbers on the circle S^1_∞, we obtain all the members of $\mathfrak{F}_2$, etc., as it is depicted in Fig. 4. The question arises now: what combinations of L and R will lead us to these numbers? Let us consider the generic case of $\frac{1}{2}$. With help of Eqs. (2.38) and (2.39), we obtain two transformations

$$z' = \frac{z}{z+1}, \qquad z' = \frac{az+1}{cz+2}, \qquad 2a - c = 1, \tag{2.40}$$

connecting, respectively, the numbers 1 and 0 with the Farey number $\frac{1}{2}$. Since both a and b are integers, the simplest choice is to take $a = 1 = c$. This then produces immediately

$$z' = L(z), \qquad z' = LR(z). \tag{2.41}$$

Other examples can now be constructed without problems. Evidently, each successive level of $\mathfrak{F}_n$ can be obtained by some application of combinations of L's and R's. To make this procedure systematic, let us consider some Farey number, say p/q, which has a continued fraction expansion of the type [44]

$$\frac{p}{q} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \frac{1}{a_3 + \cdots}}} \equiv [a_0, a_1, \dots, a_n], \tag{2.42}$$

where $a_1, \dots, a_n$ are some integers. Suppose, we would like to connect the point 0 (or ∞) with the point p/q. To this purpose, using results given by Eq. (2.23), we need to take into account that

$$x' = L^{-1}x = \frac{x}{-x+1} = \frac{1}{(1/x)-1}, \quad x' = R^{-1}x = x - 1. \tag{2.43}$$

Using these results, it is clear that

$$R^{-a_0}\left(\frac{p}{q}\right) = \frac{1}{a_1 + 1/b_1}, \tag{2.44}$$

where

$$\frac{1}{b_1} = \frac{1}{a_2 + (1/c_1)}, \text{ etc.} \tag{2.45}$$

Using Eqs. (2.43) and (2.44), we obtain

$$L^{-1}R^{-a_0}\left(\frac{p}{q}\right) = \frac{1}{a_1 + (1/b_1) - 1}, \tag{2.46}$$

and, since

$$L^{-a}x = \frac{1}{(1/x) - a}, \tag{2.47}$$

we may write as well, instead of Eq. (2.46),

$$L^{-a_1}R^{-a_0}\left(\frac{p}{q}\right) = \frac{1}{b_1}. \tag{2.48}$$

Taking into account Eq. (2.45), we also get

$$L^{-a_2}\left(L^{-a_1}R^{-a_0}\left(\frac{p}{q}\right)\right) = a_3 + \frac{1}{d_1}. \tag{2.49}$$

Now one has to apply to this R^{-a_3} in order to obtain expression similar to Eq. (2.44) so that this process may continue. The end result will depend upon whether $a_n > 1$ or $a_n = 1$. When $a_n > 1$ we can write $a_n = (a_n - 1) + 1$ and then use the operator $L^{-(a_n-1)}$ to $1/a_n$. Based on the results just obtained, it should be clear by now that it is possible to construct any Farey number starting from $0, 1, -1$ or $\pm\infty$ with the result being some word W of the type

$$W = L^{\alpha_1}R^{\beta_1} \cdots L^{\alpha_r}R^{\beta_r}, \tag{2.50}$$

where the exponents $\alpha_1, \beta_1, \dots, \alpha_r, \beta_r$ are related directly to numbers $a_1, \dots, a_n$.

Consistent triangulation of H^2 (or $\mathcal{D}$) invariant with respect to the action of $\mathrm{PSL}(2, \mathrm{Z})$ can be achieved now. To this purpose first, we connect the point $\pm\infty$ with the point $\frac{0}{1}$ of Fig. 4 with help of the equation

$$LRL^{-1}R^{-1}(-\infty) = 0, \tag{2.51}$$

which follows easily from Eq. (2.36). Since the points $\frac{1}{1}$ and $\frac{-1}{1}$ are mediants already, it should be clear how to produce the rest of the triangles from these two which are basic as it is depicted in Fig. 4. By direct observation, it is easy to notice that this figure has an axial symmetry. It can be shown [43], that the r.h.s corresponds to words of type W_1 while the l.h.s. corresponds to words of the type $\hat{I}W_1$ and that these two sets are nonoverlapping. Hence, we shall use here and in Sections 3 and 4 only the r.h.s. of this figure. Although lines on this figure look like geodesics, actually, they are not geodesics in a usual sense (so that the statement made in [46], e.g. see page 566, that these lines are geodesics should be treated with caution) as it will be explained below. Hence, the triangulation in Fig. 4 should be understood only in a topological sense (to keep track of how the Farey numbers are related to each other). An alternative and very effective geometrical description of Farey numbers (via the circle packing in H^2) had been proposed by Rademacher [47]. We shall discuss it briefly in Section 5.

Although the Farey tesselation of $\mathcal{D}$ depicted in Fig. 4 is helpful, it cannot be used directly for our calculations since we are interested in geodesics related to the Markov triples. Therefore, now we would like to relate these Markov geodesics to the Farey tesselation. The simplest Markov geodesics are associated with matrices given by Eq. (2.19) (in view if Eq. (2.15)). Since we require these matrices to correspond to the closed geodesics on the Riemann surface of leaky torus, when lifted to $\mathcal{D}$ (or H^2), their ends must lie on S^1_∞ (or on the line $y = 0$ in H^2 model). The location of the ends of geodesics is determined by the fixed points equation which in the case of matrix a is given by

$$x = \frac{2x+1}{x+1} = 1 + \frac{1}{1+1/x}. \tag{2.52}$$

Iteration of this equation leads to the periodic continued fraction expansion with period 1 [44]

$$x = [\dot{1}] = \frac{\sqrt{5}+1}{2}. \tag{2.53}$$

We had used the notations of [44, p. 46], to reflect the periodicity. Surely, this result can be obtained as well directly from the quadratic equation

$$x^2 - x - 1 = 0, \tag{2.54}$$

which is equivalent to Eq. (2.52). Eq. (2.54) has two roots

$$x_{1,2} = \tfrac{1}{2}(1 \pm \sqrt{5}). \tag{2.55}$$

To recover the second root from Eq. (2.52) we have to introduce the following change of variable: $x = -(1/y)$, in Eq. (2.52). This then produces after a few trivial manipulations

$$y = -\frac{1}{1 + 1/(1-y)}. \tag{2.56}$$

Iterating this expression, we obtain

$$y = -\frac{1}{1 + \frac{1}{1 + \frac{1}{1+\cdots}}}. \tag{2.57}$$

This is equivalent to the continued fraction expansion of $\frac{1}{2}(1-\sqrt{5})$ [44] as required. Obtained results are special cases of general theorem proven by Series [45] which states that

$$x_1 = x_\infty = [n_1, n_2, \ldots], \qquad x_2 = x_{-\infty} = \frac{-1}{[n_0, n_{-1}, n_{-2}, \ldots]}, \tag{2.58}$$

so that in both cases we have an infinite continued fractions corresponding to irrational numbers. In the case of fixed points for the Markov matrices the continued fractions are always periodic [48]. This fact has profound topological significance as it will be explained in Section 3. In the meantime, we still need to clarify the connections between the Farey tesselation of H^2 (or $\mathcal{D}$) and the results just obtained. The comprehensive treatment of this problem by mathematicians, surprisingly, had been performed only quite recently [42,46]. Moreover, to our knowledge, there are no similar comprehensive treatments for surfaces of higher genus (perhaps, with exception of Ref. [49]). The major reason for utilization of the Farey triangulation of H^2 is exactly the same as used in approximations of irrational numbers by the rationals which belong to the Farey series [44]. According to the theory of numbers [44] all rational numbers are equivalent since they are connected with each other via modular transformation, Eq. (2.37). In our case this means that the vertices of the basic quadrangle depicted in Fig. 3 upon identification (needed for formation of the leaky torus) all correspond to a puncture. Subsequent higher levels of the Farey tesselations do not change this result: all Farey numbers correspond to a puncture. Hence, from here it also follows that the ends of geodesics which correspond to the nonperipheral elements of G are represented by the irrational numbers connected by bi infinite sequence of Möbius transformations. In particular case of Eq. (2.52) this becomes obvious if we rewrite it in the form

$$x = RL(x). \tag{2.59}$$

Iterating, we obtain,

$$x = RLRLRL\cdots(x). \tag{2.60}$$

To clarify the meaning of this result, please, recall that, according to Eq. (2.21), $RL = a$. The eigenvalues $\lambda_{1,2}$ of a can be easily found from the equation

$$\lambda^2 - 3\lambda + 1 = 0, \tag{2.61}$$

thus producing

$$\lambda_{1,2} = \tfrac{1}{2}(3 \pm \sqrt{5}). \tag{2.62}$$

Not surprisingly, λ_1 coincides with earlier obtained result, Eq. (2.3). These results will acquire completely new topological meaning in Section 3.

Consider now the Möbius transformation $U_\lambda(z)$ given by

$$z' = U_\lambda(z) = \left(\frac{\lambda_1}{\lambda_2}\right) z \equiv \hat{\lambda} z. \tag{2.63}$$

This transformation has two fixed points: $z^* = 0$ and $z^* = \infty$. Consider yet another Möbius transformation $W(z)$ such that

$$W(x_1) = 0, \qquad W(x_2) = \infty, \quad \text{e.g. } W(z) = \frac{z - x_1}{z - x_2}, \tag{2.64}$$

where x_1 and x_2 are given by Eq. (2.55). Hence, using such transformation, we obtain

$$WaW^{-1}(z) = U_\lambda(z), \tag{2.65}$$

so that Eq. (2.60) can be rewritten as

$$x = WaW^{-1}WaW^{-1}\cdots(x) = U_\lambda U_\lambda \cdots U_\lambda(x). \tag{2.66}$$

Therefore, indeed, in view of Eq. (2.63), we obtain the expected sequence of iterates connecting two fixed points. These points are the initial and the final limiting points representing “motion” along the geodesics in H^2 which are just semicircles passing through points x_1 and x_2. It is very important to realize that in the case of a torus (and also punctured torus) the hyperbolic H^2 plane coincides with the Teichmüller space. Hence, the “motion” along the geodesic in H^2 coincides with the real motion in the Teichmüller space as was discussed qualitatively in our earlier work on $2+1$ gravity. More details will be provided in Sections 3 and 4.

Consider now the possibility of joining of two Farey numbers by the geodesics. That is, let us consider the fixed point of equation

$$x = \frac{ax+b}{cx+d}, \qquad ad - cb = 1. \tag{2.67}$$

We need to look for solution of this equation only for the integer values of a, b, c and d. It is instructive to discuss some special cases first. Let us begin with the case $c = b = 0$. In this case, we obtain

$$x = \frac{a}{d}x. \tag{2.68}$$

Although this equation looks exactly like Eq. (2.63), these equations are not the same since according to Eq. (2.67) we should require $ad = 1$ which leaves us with the only choice: $a = d = \mp 1$. The obtained identity result indicates that the hyperbolic-like transformations are not possible if we use just $\mathrm{PSL}(2, \mathrm{Z})$. Let us now consider the parabolic transformations, e.g. let $c = 0$ then, given that $ad = 1$, we obtain

$$x = x + b. \tag{2.69}$$

This equation has only one fixed point, $x^* = \infty$, typical for all parabolic transformations. Let us now put $b = 0$ in Eq. (2.67). Then, again, $ad = 1$, and we obtain

$$x = \frac{x}{cx + 1}, \tag{2.70}$$

thus producing another acceptable solution: $x^* = 0$. Finally, let $d = 0$ or $a = 0$. In the first case, we obtain $cb = -1$ so that

$$x = \frac{ax - 1}{x}, \tag{2.71}$$

thus leading to the equation

$$x_{1,2} = \frac{a}{2} \pm \frac{1}{2}\sqrt{a^2 - 4}, \tag{2.72}$$

while in the second case we get $bc = -1$ so that

$$x = \frac{-1}{x + d}, \tag{2.73}$$

thus producing

$$x_{1,2} = -\frac{d}{2} \pm \frac{1}{2}\sqrt{d^2 - 4}. \tag{2.74}$$

In both cases the problem is reduced to finding the Pythagorean numbers, i.e. to finding all integer solutions of equation

$$a^2 + b^2 = c^2. \tag{2.75}$$

As known results indicate [50], the only solution for d in Eq. (2.74) is $d = 2$ (accordingly, $a = 2$ in Eq. (2.72)) thus leaving us with just one fixed point. Clearly, we can identify thus obtained fixed points with $0, 1, -1$ and $\pm\infty$ as depicted in Fig. 3 since all other points can be obtained by successive applications of modular transformations. Hence, the semicircles depicted in Fig. 4 are not true geodesics contrary to the statements made in [46]. Consideration of the general case (Section 3) produces the same negative result: only one trivial solution of Eq. (2.75), thus leading to just one (parabolic) fixed point for $\mathrm{PSL}(2, \mathrm{Z})$.

Although we had provided enough evidence which connects the geodesic laminations and the train tracks, more systematic treatment of this subject would lead us somewhat away from the topics which we had discussed so far. Fortunately, Sections 1.5–1.7 of Chapter 1 of the book by Penner and Harer [51] contain all the required proofs (please, read especially pp. 87–101). Much shorter proofs could be also found in the unpublished Ph.D. Thesis by Lok [52]. Since these results are needed only for rigorous mathematical proofs of connections between laminations and train tracks, we hope that our readers will consult these references for better understanding of the obtained results and those which follow in the rest of this paper.

3. From geodesic laminations to 3-manifolds which fiber over the circle

3.1. The Alexander polynomial and surface homeomorphisms

The Alexander polynomial $\Delta_8(t)$ for the figure eight knot is known to be [53]

$$\Delta_8(t) = t^2 - 3t + 1. \tag{3.1}$$

Consider, quite formally (for the time being only!), the zeroes of this polynomial. These are obtained as roots of the equation

$$t^2 - 3t + 1 = 0. \tag{3.2}$$

A simple calculation produces

$$t_{1,2} = \tfrac{1}{2}(3 \pm \sqrt{5}). \tag{3.3}$$

We would like now to compare Eqs. (3.2) and (3.3) with Eqs. (2.61) and (2.62), respectively. Since this comparison yields complete coincidence of results, the rest of this section is devoted to the proof that the above coincidence is not accidental. As a result of such proof, the connection between the dynamics of $2+1$ gravity and 3-manifolds is naturally established. Unlike Witten's [54] treatment of $2+1$ gravity which establish this connection through reformulation of this problem in terms of the Chern–Simons field theory, our treatment does not require field-theoretic arguments at all and is based mainly on Thurston's [18] theory of 3-manifolds. The condensed summary of relevant results of Thurston [18] and McMullen [55] is given in our earlier publications, [15,16].

We would like to begin with reviewing some facts about the Alexander polynomial. Although our earlier published review [56] provides sufficient physical background on knot polynomials and, in particular, on the Alexander polynomial, this background is not sufficient for our current purposes. Hence, we shall avoid references to physics literature when we shall talk about the Alexander polynomial. To simplify matters, we shall treat only the case of the Alexander polynomials for knots. The case of links is considerably more complicated and will be treated in a separate publication. For the case of knots, let V be the Seifert matrix of linking coefficients [17,57], then the Alexander polynomial $\Delta_K(t)$ for a knot K is given by

$$\Delta_K(t) = \det(V^{\mathrm{T}} - tV) \tag{3.4}$$

where V^{T} is the matrix transpose of V. Such defined polynomial has some additional remarkable properties [17,57]. For instance,

$$\Delta_K(1) = \pm 1, \tag{3.5}$$

$$\Delta_K(t) \doteq \Delta_K(t^{-1}), \tag{3.6}$$

where the symbol $\doteq$ denotes an equality up to a constant multiplier (e.g. see Eq. (3.1) for an obvious example). The above properties of Δ_K allow us to write it as polynomial of even degree $2r$ with integer coefficients a_i:

$$\Delta_K(t) = \sum_{i=0}^{2r} a_i t^i, \qquad a_{2r-i} = a_i. \tag{3.7}$$

Using the inversion symmetry property reflected in Eq. (3.6) the following Laurent expansion for Δ_K can be written:

$$\Delta_K(t) \doteq a_r + a_{r+1}(t + t^{-1}) + \cdots + a_{2r}(t^r + t^{-r}). \tag{3.8}$$

We are not interested in all possible knots and their Alexander polynomials since not all knots are of relevance to surface dynamics. As we had discussed earlier [16], only fibered knots are of relevance. The easiest way to talk about fibered knots embedded in S^3 is trough consideration of their complements $S^3 \setminus K$ in S^3. These are just 3-manifolds fibering over the circle S^1. Indeed, let S be the Seifert surface associated with knot K. Incidentally, the punctured torus is the Seifert surface for both the figure eight and the trefoil (right and left) knots [17,57]. There is another knot, the bridge knot b(7, 3), which also has the same Seifert surface but it cannot be fibered over the circle [58]. The Alexander polynomial Δ_{T} for the trefoil is known to be [53]

$$\Delta_{\mathrm{T}}(t) = t^2 - t + 1. \tag{3.9}$$

The zeros of this polynomial are readily obtained

$$t_{1,2} = \tfrac{1}{2}(1 \pm \mathrm{i}\sqrt{3}). \tag{3.10}$$

Apparently, they have nothing to do with the discussion made in the previous section. This, however, is not true as we shall soon demonstrate. To this purpose consider an orientation preserving surface homeomorphism $h : S \to S$. In the case of the punctured (holed) torus, the homeomorphism should respect the presence of a hole. The circumference of this hole is just our base space S^1 (which, as it is not too difficult to guess, is just our knot K since the knot is just a circle embedded into S^3). The Seifert surface itself is a fiber and the 3-manifold is just a fiber bundle constructed in a following way. Begin with products $S \times 0$ (the initial state) and $S_h \times 1$ (the final state) so that for each point $x \in S$ we have $(x, 0)$ and $(h(x), 1)$, respectively. The interval $I = (0, 1)$ can now be closed (to form a circle S^1) by identifying 0 with 1 which causes identification:

$$(x, 0) = (h(x), 1). \tag{3.11}$$

The fiber bundle (also known in the literature as *mapping torus* [55,59])

$$T_h = \frac{(S \times I)}{h} \tag{3.12}$$

is the 3-manifold which fibers over the circle and is complementary to the fibered knot in S^3. The interval $(0, 1)$ can be associated with some local time. The cyclic character of the process leading to formation of 3-manifold(s) is not essential as it will be demonstrated below. Therefore, actually, the time interval can be taken from $-\infty$ to ∞. The periodicity naturally occurs, if homeomorphisms are associated with motion along the Markov geodesics in the Teichmüller space, e.g. see Eq. (2.60). Indeed, the homeomorphisms of surfaces are associated with dynamics of train tracks. If we start with train track dynamics, it is in one-to-one correspondence with dynamics of geodesic laminations and this dynamics, in turn, is associated with motions in Teichmüller space, e.g. along some geodesics in this space. Hence, the periodicity occurs quite naturally. McMüllen [55] and Othal [60] had proved that the situation just described for the punctured torus persist for Riemann surface (with marking and/or boundaries) of any genus g as it was briefly mentioned earlier in our work, [16]. More specifically, they proved the following theorem.

Theorem 3.1. *Let $\psi : S \to S$ be a pseudo-Anosov homeomorphism of compact surface with negative Euler characteristic. Then, in order for the mapping torus T_ψ to have a hyperbolic structure one is looking for the related hyperbolic manifold $M_\psi = S \times R$ on which the homotopy class of ψ is represented by an isometry α. Then, $M_\psi/\langle\alpha\rangle$ is homeomorphic to T_ψ.*

It can be demonstrated, e.g. Proposition 5.10 and the comments which follow in [17], that classification of all fibered knot complements can be formulated in terms of fibering surfaces and maps of such surfaces. Therefore, naturally, this fact is reflected in the associated Alexander polynomials. Unfortunately, the conjecture that all fibered knots are classified one-to-one by their Alexander polynomials happens to be wrong. Morton [61] had demonstrated that for the Seifert surfaces of genus $g > 1$ there are infinitely many different fibered knots for each Alexander polynomial of degree > 2. This fact by no means diminishes the role of the Alexander polynomial in dynamics of $2+1$ gravity, it just makes knot interpretations [21,54] of gravity less convincing. Similar negative conclusions had been reached recently with help of absolutely different set of arguments in [62].

From the knot theory [17] it is known, that instead of expansion, Eq. (3.7), for the Alexander polynomial for any knot, in the case of fibered knots one has to use

$$\Delta_K(t) = \sum_{i=0}^{2g} a_i t^i, \tag{3.13}$$

where g is the genus of the associated Seifert surface. Moreover,

$$\Delta_K(0) = a_0 = a_{2g} = \pm 1, \tag{3.14}$$

i.e. the Alexander polynomial is monic. Using Eq. (3.4), we obtain at once

$$\Delta_K(0) = \det(V^{\mathrm{T}}) = \det(V) = \pm 1 \tag{3.15}$$

to be compared with determinant in Eq. (2.31). Using Eqs. (3.15) and (3.4), we obtain without delay

$$\Delta_K(t) = \det(V^{-1}V^{\mathrm{T}} - tE) \equiv \det(M - tE), \tag{3.16}$$

where E is the unit matrix and M is the monodromy matrix responsible for surface automorphisms [17,57]. This fact easily follows from the observation that $\Delta_K(0) = \det(M) = \pm 1$.

Remark 3.2. Equating the Alexander polynomial to zero and finding the roots of this polynomial is equivalent to solving the eigenvalue problem for surface automorphisms which produces the stretching factors allowing one to distinguish between the hyperbolic, pseudo-Anosov-type (when the largest root is real and greater than one), and the periodic, Seifert-fibered-type (when the modulus of the largest root is equal to one), regimes of surface homeomorphisms.

Let us discuss Eq. (3.16) a bit more. It can be easily shown [63] that, if $M \in \mathrm{GL}(2, Z)$, then

$$\Delta_K(t) = t^2 - (\mathrm{tr}\, M)t + \det M. \tag{3.17}$$

But, as we know already, $\det M = \pm 1$. Hence, in this case we obtain, instead of Eq. (3.17),

$$\Delta_K(1) = t^2 - (\mathrm{tr}\, M)t \pm 1. \tag{3.18}$$

Using Eq. (3.5), we obtain as well

$$\Delta_K(1) = 1 - (\mathrm{tr}\, M) \pm 1 = \pm 1. \tag{3.19}$$

This leaves us with two options: $\mathrm{tr}(M) = 1$ or 3. In the first case, we reobtain the Alexander polynomial for the trefoil, Eq. (3.9), and in the second for the figure eight, Eq. (3.1), knots. No other options are available! Hence we had just proved (in a somewhat different way as compared with [17]) that the trefoil and the figure eight knots are the only two fibered knots associated with the Seifert surfaces of genus 1. Moreover, the conditions $\mathrm{tr}\, M = 3$ and $\det M = 1$ lead us directly to the Markov matrix a, Eq. (2.19). This matrix, Eq. (2.60), and Theorems 5.10 and 5.11 of [17] provide direct connection between the hyperbolic 3-manifold associated with complement of figure eight knot and the pseudo-Anosov surface homeomorphisms associated with a. These conclusions are in accord with results of Thurston [18, Section 4.37], where they had been obtained in a different way.

At the same time, the conditions $\mathrm{tr}\, M = 1$ and $\det M = 1$ are also very interesting since they are associated with the matrix

$$M = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, \tag{3.20}$$

whose eigenvalues we had calculated already in Eq. (3.10). It is easy to check that such surface homeomorphism is not associated with motion along the hyperbolic geodesic since the fixed point equation

$$x = 1 - \frac{1}{x}, \tag{3.21}$$

which is equivalent to Eq. (3.9) (if we require $\Delta_{\mathrm{T}}(t)$ to be zero) does not have real roots. Using the same methods as in Section 2, we easily obtain that projectively the matrix M is equivalent to the combination of LR^{-1}. Such transformations do not fit Theorem 5.11 and the associated 3-manifolds are known as Seifert-fibered spaces. Since the trefoil knot is the simplest representative of torus knots, it can be demonstrated [17], that the complements of all torus knots in S^3 are associated with the Seifert-fibered spaces. According to Thurston's classification of surface homeomorphisms those leading to the Seifert-fibered spaces are known as periodic. The detailed structure of the periodic phase may be very complicated [64–66] and, hence, potentially very interesting from the point of view of physical applications. The discussion of all emerging possibilities, surely, requires separate publications. The geometrical richness of the Seifert-fibered regime provides a likely explanation of different states of order in the case of liquid crystals as we had briefly mentioned in Section 1 and in earlier works. For example, solid and hexatic phases might be associated with the Seifert-like while liquid and/or gas phases may be associated with the pseudo-Anosov-like. In this paper, we shall be concerned only with transition between the pseudo-Anosov and the Seifert-like phases leaving the detailed analysis of possibilities emerging in the Seifert-fibered (periodic) phase for future publications. In order for our results to be consistent with the rest of the literature on phase transitions we need now to introduce several new concepts.

3.2. Mahler measures and topological entropies

Following Ref. [67], let us consider a monic polynomial with integer coefficients

$$F(x) = x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0, \tag{3.22}$$

where $a_0 = \pm 1$. If α_i are the roots of this polynomial, then, equivalently, we can rewrite it as

$$F(x) = \prod_{i=1}^{d}(x - \alpha_i). \tag{3.23}$$

The logarithmic Mahler measure $m(F)$ can be defined now as

$$m(F) = \ln M(F), \tag{3.24}$$

where $M(F)$ is given by

$$M(F) = \prod_{i=1}^{d}\max\{1, |\alpha_i|\}. \tag{3.25}$$

The following theorem is attributed to Walters, [68, Section 8.4]:

Theorem 3.3. *The topological entropy of the transformation M is equal to the logarithmic Mahler measure of the characteristic polynomial of the matrix M.*

To make a connection with statistical mechanics, we would like to notice that the fact the polynomial $F(x)$ is monic is not at all restrictive (we used it only in order to make connection with the previous discussion). Let us recall at this point some facts from Lee and Yang [6] theory of phase transitions. For instance, for a gas of $\mathfrak{M}$ atoms which can be packed into the volume $\mathfrak{v}$ the grand canonical partition function Ξ is given in a usual manner as

$$\Xi = \sum_{n=0}^{\mathfrak{M}} \frac{Q_n}{n!} z^n. \tag{3.26}$$

Hence, the grand partition function is just some polynomial in fugacity z. Surely, the polynomial, Eq. (3.26), must have some zeros. These zeros can be only in the complex plane **C** and they have to come in pairs of complex conjugates for Eq. (3.26) to be real for real z's. Hence, effectively, we can rewrite Ξ as follows:

$$\Xi(z) = \frac{Q_{\mathfrak{M}}}{\mathfrak{M}!} \prod_{n=1}^{\mathfrak{M}/2} (z - \bar{z}_n)(z - z_n). \tag{3.27}$$

Since $\Xi(z = 0) = 1$, the above result can be conveniently rewritten as

$$\Xi(z) = \prod_{n=1}^{\mathfrak{M}} \left(1 - \frac{z}{z_n}\right)\left(1 - \frac{z}{\bar{z}_n}\right). \tag{3.28a}$$

For physical applications one is interested in finding of $\ln \Xi$, i.e.

$$\mathfrak{F}(z) = \ln \Xi = \sum_{n=1}^{\mathfrak{M}/2} \ln(z^2 - 2\cos\theta_n + 1), \tag{3.28b}$$

where use had been made of the fact that the complex zeros should lie on the unit circle in the complex plane **C** (this is demonstrated below). If $g(\theta)$ is the density of these zeros, one can replace the above summation in Eqs. (3.28a) and (3.28b) by the integration with the result

$$\mathfrak{F}(z) = \int_0^{\pi} \mathrm{d}\theta\, g(\theta) \ln|(z^2 - 2\cos\theta + 1)|, \tag{3.29}$$

which is written with account of symmetry of the integrand. To make a connection between this result and the logarithmic Mahler measure it is only sufficient to use.

Lemma 3.4 (Mahler's lemma [67]). *For any nonzero $F \in C[x]$,*

$$m(F) = \int_0^1 \mathrm{d}\theta \ln|F(\exp(2\pi \mathrm{i}\theta))|. \tag{3.30}$$

Proof. This result follows at once if one can prove Jensen's formula: for any $\alpha \in \mathbf{C}$

$$\int_0^1 \ln|\exp(2\pi \mathrm{i}\theta) - \alpha| = \ln \max\{1, |\alpha|\}. \tag{3.31}$$

The detailed proof of this result can be found in [67]. Using Eqs. (3.23)–(3.25), Lemma 3.4 is proved. □

From here, the following theorem follows at once.

Theorem 3.5. *$m(F) = 0$ if and only if $|\alpha_i| = 1$ for all i in Eq.* (3.23), *provided that $F(x)$ is monic polynomial.*

This result coincides with that of Lee and Yang, e.g. see [6, Theorem 3, Eq. (57), p. 418], where it was obtained using different methods. Theorem 3.5. does not imply that for all monic polynomials $m(F) = 0$. This is definitely not the case for the Alexander polynomial, Eq. (3.1), for the figure eight knot. Evidently, for such knot the topological entropy is given by

$$m_8 = \ln \tfrac{1}{2}(3 + \sqrt{5}), \tag{3.32}$$

while for the trefoil knot m_{T}, is indeed zero in view of Eq. (3.10). Hence, the logarithmic Mahler measure can be used for the Alexander polynomial of any fibered knot to distinguish between the pseudo-Anosov and the periodic regimes. We would now like to complicate matters by considering.

3.3. The incompressible surfaces in the once punctured torus bundles over S^1

Although the previous analysis is interesting in its own right, the situations which we had considered so far (with the trefoil and the figure eight knots) are not the only possibilities for the punctured torus automorphims. To go beyond these possibilities means to abandon almost completely the connections with knots and to concentrate attention on solutions of the equation

$$\mathfrak{D}(t) = \det(M - tE) = 0. \tag{3.33}$$

Since we now do not require $\mathfrak{D}(t)$ to be connected with knot polynomials, there is no need for $\mathfrak{D}(1) = \pm 1$. At the same time, Eq. (3.33) is legitimate equation for finding of stretching factors characteristic for pseudo-Anosov homeomorphisms. For an arbitrary matrix $M \in \mathrm{SL}(2, \mathrm{Z})$ we obtain, instead of Eq. (3.18), the equation ($n \in Z$)

$$t^2 - nt \pm 1 = 0, \tag{3.34}$$

which produces the following roots:

$$t_{1,2} = \tfrac{1}{2}(n \pm \sqrt{n^2 \mp 4}). \tag{3.35}$$

From here, we see that only n which obey $|n| \leq 2$ produce the stretching factors characteristic for the periodic (Seifert-fibered) phase. For all $|n| \geq 3$ the stretching factors are characteristic of the pseudo-Anosov phase. These results can be interpreted topologically with help of the notion of incompressible branched surfaces introduced by Oertel [69]. We would like to discuss these surfaces in order to make connections with the results of Section 2 and in order to provide topological interpretation of the results of Section 4.

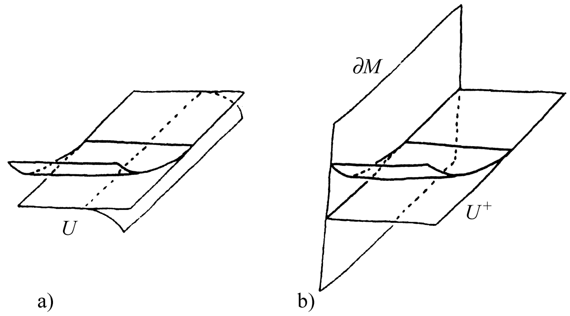

Fig. 5. Branched surfaces associated with dynamics of train tracks.

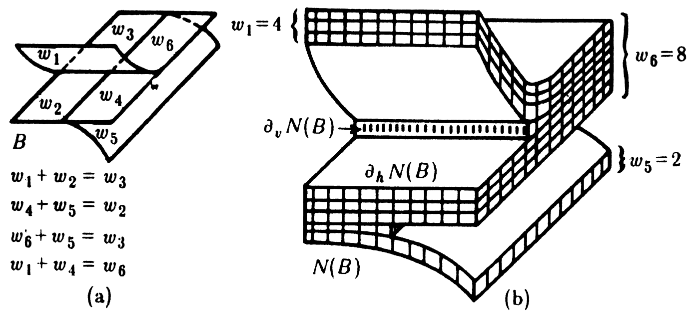

Fig. 6. (a) Kirkhoff-type rules for branched surfaces. (b) Another interpretation of the weights on branches (read the text for details).

The notion of branched surfaces is inseparably connected with the notion of train tracks. As soon as we would like to visualize the dynamics of train tracks in time, we encounter branched surfaces which are perpendicular to the original surface as it is depicted in Fig. 5.

Naturally, these branched surfaces inherit the weights, e.g. see Fig. 1, from the associated with them train tracks and these weights obey the switch conditions as for the train tracks. This is depicted in Fig. 6a. The weights are some positive integers which could be interpreted as number of surfaces in the stack (Fig. 6b). These branched surfaces inherit from the train tracks some topological restrictions as depicted in Fig. 7.

For instance, in the case of train tracks the monogons are forbidden as well as discs of contact, etc. (for details, please, contact Ref. [52]). Floyd and Oertel [70] (see also [69, p. 387]) had proved the following.

Theorem 3.6. *There is a finite collection of incompressible branched surfaces in some 3-manifold M such that every two-sided incompressible, ∂-incompressible surface in M is carried with positive weights by a branched surface $\mathfrak{B}$ of the collection.*

The surface S is called two-*sided* in M if there is a regular neighborhood of S homeomorphic to $S \times I$ where $I = [0, 1]$ [71]. If $S \neq S^2$, RP^2, or a disc D^2 which can be

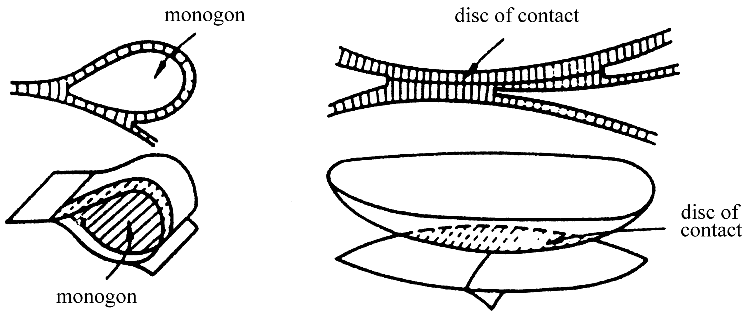

Fig. 7. Topological restrictions which branched surfaces inherit from train tracks.

pushed to ∂M, then S is *incompressible* if every loop (a simple closed curve) on S which bounds an (open) disc in $M \setminus S$ also bounds a disc in S. If S has boundary, then S is also ∂-incompressible if every arc α in S (with $\partial(\alpha) \in \partial S$) which is homotopic to ∂M is also homotopic in S to ∂S [18, Section 4.10].

Fig. 6b shows the local model for the fibered regular neighborhood $N(\mathfrak{B})$ of $\mathfrak{B}$, it also depicts the *horizontal* boundary $\partial_{\mathrm{h}} N(\mathfrak{B})$ and the *vertical* boundary $\partial_{\mathrm{v}} N(\mathfrak{B})$.

The branched surface $\mathfrak{B}$ is incompressible if

1. $\mathfrak{B}$ has no discs of contact;
2. $\partial_{\mathrm{h}} N(\mathfrak{B})$ is incompressible and ∂-incompressible in $M \setminus \mathring{\mathrm{N}}$ where $\mathring{\mathrm{N}}$ is just an interior of $N(\mathfrak{B})$ and ∂-compressing disc for $\partial_{\mathrm{h}} N(\mathfrak{B})$ is assumed to have boundary in $\partial M \cup \partial_{\mathrm{h}} N(\mathfrak{B})$;
3. there are no monogons in $M \setminus \mathring{\mathrm{N}}$.

Being armed with such definitions, we would like to show how actually these surfaces can be built. This task was actually accomplished in [72,73]. Alternative treatment can be found in recently published paper by Minsky [46]. It is essential to contact these references for detailed study of this very involved topic. The theory of incompressible surfaces in connection with general theory of 3-manifolds can be found in the review by Jaco [74]. Here, we restrict ourselves only with very basic facts which are needed for topological interpretation of the results which follow in Section 4.

The mapping torus construction, Eq. (3.12), does produce 3-manifolds which are the complements of Fig. 8 and trefoil knots, respectively. It is clear, based on Eq. (3.19), that nothing much beyond this can be obtained. To get more, one needs to generalize this construction. Following [72], we would like to consider more involved way of constructing torus bundles. For example, let S be some surface and let X_1, X_2 be the topological spaces associated with the mapping $\gamma_i : S \times [0,1] \to X_i$, $i = 1, 2$. For some surface homeomorphism $h : S \to S$ define $X_1/_h X_2$ to be the quotient space obtained from $X_1 \cup X_2$ by identifying $\gamma_1(x, 1)$ with $\gamma_2(h(x), 0)$. In this picture the mapping torus T_h, Eq. (3.12), is just a quotient space obtained from X_1 by identification of $\gamma_1(x, 1)$ with $\gamma_1(h(x), 0)$. Now, however, we can extend this construction by letting i to range from 1 to n. Thus, if $\gamma_i : S \times [0,1] \to X_i$ and $h_i : S \to S$ are homeomorphisms for $i = 1, \ldots, n$, then the fiber

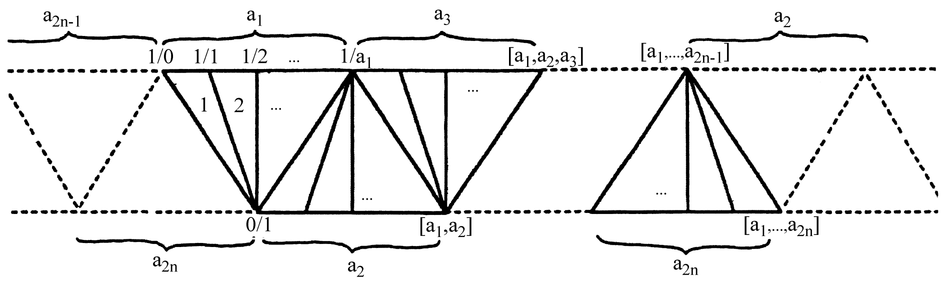

Fig. 8. Bi-infinite strip $\sum_{\varphi}$ of edge-paths associated with hyperbolic transformation φ.

bundle

$$X = X_1/_{h_1}X_2/_{h_2}\cdots/_{h_{n-1}}X_n/_{h_n} \tag{3.36}$$

is associated with some 3-manifold which, actually, can be obtained from the figure eight 3-manifold by means of hyperbolic Dehn surgery [18,75]. Before explaining its meaning, several simpler concepts need to be elucidated. First, since according to Section 2, all toral homeomorphisms can be performed with help of the right R and the left L Dehn twists (for orientation preserving homeomorphisms), the fiber bundle construction just described is equivalent to some transformation φ of the punctured torus given by

$$\varphi = R^{a_1}L^{a_2}\cdots L^{a_{2n}}, \quad a_i \geq 1, \tag{3.37}$$

in accord with [72]. Since each torus is completely determined by the ratio τ of its sides, e.g. $\tau_i = a_i/b_i$, it is clear, in view of Eq. (3.36) that the transformation φ should be such that $\varphi(a_i/b_i) = a_{i+2n}/b_{i+2n}$ for all i and fixed n. This prompts us to discuss the properties of the partition function $Z = \operatorname{tr}\varphi$ which we postpone till Section 4. In this section, we would like to discuss a bit more the topological issues. According to the results of Section 2, the transformation φ is of the same kind as the transformation W in Eq. (2.50). This means that the cumulative result of matrix multiplications in Eq. (3.37) is just some matrix M of the type given by Eq. (2.31). In Section 2, we had already discussed special cases of equations for geodesics. Here, we would like to investigate the most general case. Using Eq. (2.67), we obtain the following roots which provide locations of the beginning and the end of hyperbolic geodesics on S^1_∞:

$$x_{1,2} = \frac{1}{2c}\left(a - d \pm \sqrt{(a+d)^2 - 4}\right). \tag{3.38}$$

Since $a + d$ is just the trace of M we have to impose the usual requirement that $\operatorname{tr}^2 M \geq 4$ for the transformation to be hyperbolic. In addition, however, we have to take into account that $\operatorname{tr}^2 M$ is some nonnegative integer. As we had argued in Section 2, the roots $x_{1,2}$ cannot be rational numbers. Hence, whatever they might be, they will belong to some quadratic irrationalities. From the number theory [44] it is well known that the continued fraction expansions of quadratic irrationalities are periodic as it had been mentioned already after Eq. (2.58) and this fact also explains the periodicity of φ transformation.

Remark 3.7. Since, according to Section 2, the exponents $a_1, \dots, a_{2n}$ in Eq. (3.37) are in one-to-one correspondence with the coefficients of continued fraction expansion, this periodicity is in one-to-one correspondence with the fiber bundles constructed in Eq. (3.36).

That this is indeed the case is shown in great detail in [72,73]. Moreover, since the transformation φ is hyperbolic, it acts by translations along the geodesics in $\mathcal{D}$ whose ends are determined by the roots of Eq. (3.38). This has been illustrated already in Eqs. (2.60) and (2.66). Imagine now that we move along one of such geodesics, say $\mathfrak{G}$, that is we move in the Teichmüller space of the punctured torus. Such motion will be associated with the sequence of crossings of triangles of the Farey tesselation. To understand why this is so the following arguments are helpful. Each torus can be triangulated, i.e. it can be represented as at least two triangles. Obviously, one of this triangles is sufficient for complete characterization of the torus. The triangle is determined by the slopes of its sides. For example, the undistorted triangle is determined by the triple $\Delta(T) = \langle \frac{1}{0}, \frac{0}{1}, \frac{1}{1} \rangle$. It is sufficient to take a look at Figs. 3 and 4 in order to recognize one-to-one correspondence between this triple and the Farey numbers $\mathfrak{F}_1$ for the largest triangle in the Farey tesselation. Hence, motion in the Teichmüller space of the punctured torus is indeed associated with sequence of triangles in the Farey tesselation of $\mathcal{D}$ since $M(\Delta(T)) = \Delta(M(T))$ where $M(z)$ is defined by Eq. (2.37). Topologically, we can visualize this chain of triangles as an infinite strip $\sum_\varphi$, e.g. see Fig. 8, which consists of triangles crossed by $\mathfrak{G}$. The numbers $a_i > 1$ indicate the number of smaller triangles within each larger triangle. They are in one-to-one correspondence with the exponents in Eq. (3.37). The periodicity of the transformation φ is reflected in the periodicity of the triangulation pattern in the strip $\sum_\varphi$. To make all these statements more physical, we need to discuss the meaning of hyperbolic Dehn surgery (actually, Dehn filling!) now.

Let M be a hyperbolic 3-manifold whose boundary ∂M is a torus T. If we write $T = \mathbf{R}^2/\mathbf{Z}^2$ and choose for basis vectors $\mathbf{e}_1, \mathbf{e}_2$ of the lattice $\mathbf{Z}^2$ then, for any coprime pair of integers (p, q) the element $p\mathbf{e}_1 + q\mathbf{e}_2$ determines some simple curve on T. Consider another solid torus $S^1 \times D^2$. This torus we glue into M in such a way that the curve $p\mathbf{e}_1 + q\mathbf{e}_2$ in ∂M is being glued to the meridian (that is to the curve of slope $\frac{1}{0}$) of the solid torus.

Remark 3.8. It is being said that this new 3-manifold $M(p, q)$ is obtained from the old M by the operation of Dehn filling along the curve of slope p/q.

Hatcher has proved the following remarkable theorem (using branched surfaces) [76].

Theorem 3.9. *A 3-manifold with a single torus boundary component has only finitely many boundary slopes.*

Remark 3.10. (a) From the previous discussion it follows that there should be a correspondence between the incompressible surfaces and the boundary slopes. (b) The boundary slopes are also in correspondence with $\mathbf{Q} \cup \{\infty\}$ and, hence, with the Farey tesselation of $\mathcal{D}$.

This correspondence can be rephrased in physically familiar terms as we would like to demonstrate now. To this purpose, let us recall that the continued fraction expansion for α which is quadratic irrational can be written as

$$\alpha = [a_0, \dots, a_r, a_{r+1}, \dots, a_{r+2n}, a_{r+1}, \dots, a_{r+2n}, \dots]. \tag{3.39}$$

The quadratic irrational is purely periodic [77] if we have $a_{m+2n} = a_m$ for all m (i.e. in this case, the terms $a_0, \dots, a_r$ are absent). Evidently, Fig. 8 represents just this case. Let us now associate the boundary slope p/q with continued fraction

$$\frac{p}{q} = [a_1, \dots, a_{2n}], \tag{3.40}$$

(e.g. see the discussion after Eq. (3.37)) and consider directed random walks from the point $\frac{1}{0}$ to the point p/q on the diagram (Fig. 8). The random walk γ is directed if it does not have backtracks, moreover, the walk is minimal if two successive steps never belong to the same triangle. A brief introduction to the properties of such walks can be found in our earlier work [78], where such walk was used for description of the discretized version of the Dirac propagator. Such discretization is useful in some problems relevant to physics of semiflexible polymers.

Following our earlier work [78], let $N = n$ be the total number of steps in the directed (edge-paths) walk γ and let N_+ be the number of right turns while N_- be the number of left turns in such a walk. Associate a system of N Ising spins σ_i with the walk so that $\sigma_i = +1(-1)$ will correspond to the turn to the left (right). With such defined rule for spins, the "magnetization" $\mathbf{M}$ is given by

$$\mathbf{M} = \sum_{i=1}^{N} \sigma_i. \tag{3.41}$$

In the case if eigenvalues of the matrix φ, Eq. (3.37), are positive the boundary slope $m(n)$ is given by [73,79]

$$m(n) = \tfrac{1}{4}\mathbf{M}. \tag{3.42}$$

If the matrix φ has negative eigenvalues then, instead of Eq. (3.42), one has $m(n) = \frac{1}{4}\mathbf{M} + \frac{1}{2}$. In the following section, thus introduced quantities acquire new statistical mechanical meaning. Floyd and Hatcher [73] have proved the following theorem.

Theorem 3.11. *If φ (Eq. (3.37)) is hyperbolic, then a connected, orientable, incompressible, ∂-incompressible surface in M_φ is exactly one of*:

1. *the peripheral torus ∂M_φ,*
2. *the fiber $T^2 - \{0\}$,*
3. *a finite number (≥ 2) of non-closed surfaces S_γ indexed by minimal edge-paths γ.*

From here, it follows, in particular, that for transformation given by the matrix a, Eq. (2.21), which is relevant to 3-manifold associated with the complement of figure eight knot, there are exactly two non-closed orientable incompressible ∂-incompressible surfaces.

According to Thurston [18, Section 4.40], these are just $T^2\backslash 2$ discs surfaces which are oppositely oriented. The above Theorem 3.11 does not include the nonorientable incompressible surfaces. This case was studied by Przytycki [79,80] whose work extends the work by Floyd and Hatcher and is based on the same edge-path methods. For the complement of figure eight knot 3-manifold the incompressible unoriented surface is just the Klein bottle-disc as it was also shown by Thurston in the same Section 4.40. We refer our readers to the original papers for more details. In the meantime, we would like now to provide some physical interpretation of the obtained results using some recent results from statistical mechanics of number-theoretic spin chains.

4. Thermodynamics of the Farey spin chains and statistical mechanics of $2+1$ gravity

In our previous work [15], we have used meanders and Peierls-type arguments for description of phase transitions in liquid crystals and gravity. Such type of approach, although provides some estimate of transition parameters, is not well suited for more refined analysis since even the notion of the order parameter, central to all theories of phase transitions, is not so easy to implement within such an approach. In the previous section, we had introduced the logarithmic Mahler measure, Eq. (3.24), which is ideally suitable for description of dynamical phase transitions and in, addition, perfectly fits Yang and Lee [6] theory of phase transitions. To make use of this measure, we need to provide an equivalent definition of this measure now (incidentally, this alternative definition provides an extra link between the statistical mechanics and the number theory [67]). To this purpose, instead of Eq. (3.23), introduce

$$\Delta_n(F)=\prod_{i=1}^{d}(\alpha_i^n-1). \tag{4.1}$$

Using this defined polynomial Δ_n it can be shown that, provided no zero of $F(x)$, Eq. (3.22), is a root of unity,

$$m(F)=\lim_{n\to\infty}\frac{1}{n}\ln|\Delta_n(F)|. \tag{4.2}$$

In the case if polynomial $F(x)$ has zeros which are roots of unity, $m(F)=0$. Being armed with such results, following Refs. [81,82], let us introduce the spin-like variable $\sigma_i=\{0,1\}$, $i=1-k$, which is related to the Ising spin $\mathrm{s}_i=(-1)^{\sigma_i}$. Define now inductively the matrix D_k via the following set of rules:

$$D_0=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \tag{4.3}$$

$$D_k=L^{1-\sigma_k}R^{\sigma_k}D_{k-1}(\sigma_1,\dots,\sigma_{k-1}). \tag{4.4}$$

The energy E_k can be defined now as

$$E_k=\ln T_k \tag{4.5}$$

with $T_k = \mathrm{Tr}(D_k)$. This allows us to introduce the partition function

$$Z_k(\beta) = \sum_{\{\sigma_i\}} \exp(-\beta E_k). \tag{4.6}$$

with control parameter β playing a role of the inverse temperature. Then, the free energy $\mathfrak{F}(\beta)$ can be defined in the usual way as

$$\mathfrak{F}(\beta) = \lim_{k\to\infty} F_k, \tag{4.7}$$

where

$$F_k = \frac{-1}{k\beta} \ln(Z_k(\beta)). \tag{4.8}$$

To make connection with Eq. (4.2), we notice that

$$m(F) = \lim_{k\to\infty} \frac{1}{k} E_k. \tag{4.9}$$

Since E_k is random variable, more appropriate quantity is the average Mahler logarithmic measure which can be interpreted as an average energy $U(\beta)$, i.e. we have

$$U(\beta) = \lim_{k\to\infty} U_k(\beta) \quad \text{with } U_k(\beta) = \frac{\partial}{\partial\beta}(\beta F_k(\beta)). \tag{4.10}$$

The average magnetization (per site) associated with the average boundary slope, e.g. see Eqs. (3.41) and (3.42), can be defined accordingly as

$$M_k(\beta) = \left\langle \frac{1}{k}\sum_{i=1}^{k} s_i \right\rangle_k, \tag{4.11}$$

where, as usual, for any observable O, the average $\langle\cdots\rangle_k$ is defined by

$$\langle O\rangle_k(\beta) = \frac{\sum_{\{\sigma\}} O(\sigma)\exp(-\beta E_k)}{\sum_{\{\sigma\}} \exp(-\beta E_k)}. \tag{4.12}$$

Detailed calculations performed in [7–9,81,82] allow us to avoid repetitions. Hence, we only provide here the summary of results. In the thermodynamic limit, k$\to\infty$, the partition function, Eq. (4.6), acquires the following form:

$$\hat{Z}(\beta) = \lim_{k\to\infty} Z_k(\beta) = \frac{\zeta(\beta-1)}{\zeta(\beta)} \tag{4.13}$$

announced earlier in Section 1, e.g. see Eq. (1.4).

Remark 4.1. The partition function $\hat{Z}(\beta)$ can be brought into form which coincides with that for lattice gases, e.g. see Eqs. (3.27), (3.28a) and (3.28b). To this purpose let us introduce the notation $\xi(\beta) = \Gamma(\frac{1}{2}\beta)(\beta-1)\pi^{-(\beta/2)}\zeta(\beta)$ where $\Gamma(x)$ is just the Euler's gamma function. Then, following Riemann [2], we notice that

$$\xi(\beta) = \xi(0)\prod_{\rho}\left(1-\frac{\beta}{\rho}\right),$$

where ρ ranges over the roots of the equation $\xi(\rho) = 0$. Hence, the complete statistical mechanics treatment of $2 + 1$ gravity depends crucially on our knowledge of distribution of zeros of Riemann zeta function $\zeta(\rho)$ and, hence, on constructive solution of Riemann hypothesis.

Remark 4.2. Taking into account that the Riemann's zeta function has the following integral presentation [1,2]

$$\zeta(\beta) = \frac{1}{\Gamma(\beta)} \int_0^\infty \mathrm{d}x \, \frac{x^{\beta-1}}{\mathrm{e}^x - 1},$$

one can interpret the obtained results in terms of the thermodynamic properties of some fictitious Bose gas so that the phase transition in such gas to some extent resembles Bose condensation. This topic is discussed further in the Appendix A.

Since solution of the Riemann hypothesis is not yet found (see, however, [5]), we may be content for now by finding some meaningful estimates. For instance, Eq. (4.13) allows one to obtain the critical "temperature" $\beta_{\mathrm{cr}} = 2$ [8]. In the high temperature (pseudo-Anosov) phase, i.e. for $0 < \beta < \beta_{\mathrm{cr}}$, the averaged logarithmic Mahler measure $U(\beta)$ is bounded by the following inequalities [8]:

$$\frac{\ln 2 - \beta \ln \frac{3}{2}}{2 - \beta} \leq U(\beta) \leq \ln \frac{3}{2}. \tag{4.14}$$

Since for $1.7 < \beta < 2$ the l.h.s. of this inequality is negative, more accurate estimate was obtained for $1 \leq \beta < 2$:

$$U(\beta) \geq \tfrac{1}{4}(\beta_{\mathrm{cr}} - \beta). \tag{4.15}$$

With such improved estimate Contucci and Knauf [8] had demonstrated that $U(\beta) \geq 0$ for $0 \leq \beta < 2$. Thus, the high temperature phase is indeed of pseudo-Anosov type. For $\beta \geq \beta_{\mathrm{cr}}$, i.e. in the low temperature (periodic or hexatic [19]) frozen phase, $U(\beta) = 0$, which is physically meaningful. In this phase also $\mathfrak{F}(\beta) = 0$. At the same time, the average magnetization

$$M(\beta) = \lim_{k \to \infty} M_k \tag{4.16}$$

is equal to one in the low temperature phase and is zero in the high temperature phase. This also makes physical sense since, according to the results of Przytycki [80, Proposition 3.2], new $M(p, q)$ manifolds obtained by Dehn filling from 3-manifold M which is complement of the figure eight knot are Seifert-fibered only if $p/q = \pm\frac{1}{1}, \pm\frac{1}{2}$, or $\pm\frac{1}{3}$. In order to connect this result with physics it is important to realize that the "magnetization" defined in Eq. (3.41) is made of spins which are periodically arranged in the bi-infinite strip $\sum_{\varphi}$. The thermodynamic averages are appropriate only if the period $n \to \infty$. In the case of finite n some n-dependent corrections are usually expected [19] for all observables. These corrections naturally should vanish in the thermodynamic limit $n \to \infty$. Hence, identification of the average magnetization with the average slope cannot be done automatically.

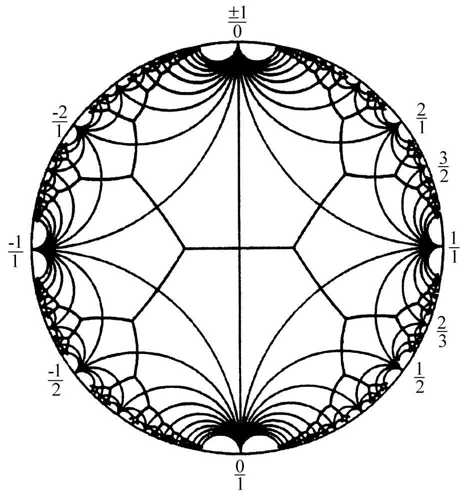

Fig. 9. Graph $\sum$ dual to the Farey tesselation of $\mathcal{D}$.

This casts some doubts on the appropriateness of the above thermodynamic formalism for description of statistical mechanics of $2 + 1$ gravity. Fortunately, there are other ways to arrive at the same conclusions which we are now going to discuss in this and the following section.

First, instead of considering the Farey triangulation of $\mathcal{D}$, we can consider the associated with it dual tree as depicted in Fig. 9. Following Bowditch [34], we notice that, actually, topologically such tree is in one-to-one correspondence with the Markov tree depicted in Fig. 2. This observation can be used for description of random walks on such trees. This walk can be associated naturally with motions in the Teichmüller space of the punctured torus as has been observed already by Penner [83]. Let $\sum$ denote the dual graph depicted in Fig. 9 and let $V(\sum)$, $E(\sum)$ and Ω denote, respectively, the sets of vertices, edges and the regions complementary to $\sum$. Each vertex lies at the boundary of three complementary regions $X, Y, Z \in \Omega$ while each edge meets four complementary regions X, Y, Z, W. The subset $\hat{\Omega}$ of nontrivial nonperipheral closed curves on T introduced in Section 2 can now be identified with Ω because $X, Y, Z, W \in \Omega$ correspond, respectively, to the equivalence classes of generators a, b, ab and ab^{-1} of the free group G as discussed in Section 2. Looking at the Fig. 9, it is easy to recognize that the regions of Ω are in one-to-one correspondence with rationals $\mathbf{Q} \cup \{\infty\}$ of the Farey tesselation. Moreover, in the light of definitions just made, we can further identify the Markov triples, x, y and z with the regions X, Y, Z and W as follows: $x = \phi(X)$, etc., where $\phi(X) = \operatorname{tr} a$, etc. Using Eqs. (2.7) and (2.8), one can easily obtain the following result:

$$z, w = h(x, y) = \frac{xy}{2}\left(1 \pm \sqrt{1 - 4\left(\frac{1}{x^2} + \frac{1}{y^2}\right)}\right). \tag{4.17}$$

This results prompts us to rewrite Eqs. (2.7) and (2.8) in the form

$$\frac{x}{yz}+\frac{y}{xz}+\frac{z}{xy}=1, \tag{4.18}$$

$$\frac{z}{xy}+\frac{w}{xy}=1. \tag{4.19}$$

Written in such form, these equations acquire new geometrical meaning. Given a directed edge $\vec{e}\in E(\sum)$, define $\psi(\vec{e})=z/xy$ so that for each edge e we have, instead of Eq. (4.19),

$$\psi(\vec{e})+\psi(-\vec{e})=1 \tag{4.20a}$$

(edge relation), while instead of Eq. (4.18), we have

$$\psi(\vec{e}_1)+\psi(\vec{e}_2)+\psi(\vec{e}_3)=1 \tag{4.20b}$$

(vertex relation). It could be rather easily shown that $h(x,y)\geq h(x)+h(y)$. Given this fact, we observe that when $2\leq |x|\leq\infty, h(x)$ is real. Accordingly, we expect that $|x|,|y|,|z|\geq 2$. Consider now some vertex $v^*\in V(\sum)$ and consider the set $T_n(v^*)$ which is a tree spanned by all vertices which are at the distance at most n from the “seed”. Let $\Omega_n(v^*)$ be the set of all complementary regions meeting $T_n(v^*)$, while $C_n(v^*)$ be the corresponding subset of edges. Then, it can be shown [84] that

$$\sum_{X\in\Omega_n(v^*)} h(\phi(X))\leq\frac{1}{2}\sum_{\vec{e}\in C_n(v^*)}\psi(\vec{e})=\frac{1}{2}. \tag{4.21}$$

Actually, the inequality above can be replaced with equality and the above equality is known as McShane identity [85]. Since, in spite of its importance, we are not going to use it, we are not going to discuss its significance. The above result is mentioned only because it is of interest to inquire what happens with the convergence of some other functions $f(X)$. Let, for instance, $f:\Omega\to[0,\infty)$ has a lower Fibonacci bound (to be defined shortly below), then, according to Bowdich [34], the series

$$F(f)=\sum_{X\in\Omega}[f(X)]^{-s} \tag{4.22}$$

converges for all $s>2$. A lower Fibonacci bound for function $f(X)$ on $\Omega_n\in\Omega$ exist if there is some constant $k>0$ such that $f(X)\geq kF_e(X)$ for all but finitely many $X\in\Omega_n$. The function $F_e(X)$ can be chosen as the length L of the reduced word W_r [34], see, e.g. Eq. (2.5). Evidently, $L=n$ in view of the results just presented. If this is so, the question arises: how n depends on X? Stated alternatively: is there way to convert the summation over X (that is over Ω_n) into summation over n? Fortunately, the last problem can be easily solved. Since X,Y,Z and W are in one-to-one correspondence with the Farey numbers all these numbers are coprime to each other [34]. Surely, that they are also coprime to n. This means that the length L can take any value n at precisely $2\phi(n)$ regions of Ω_n where $\phi(n)$ has been defined by Eq. (1.5) as number of numbers less than n which are prime to n. The factor of 2 is easy to understand if one puts $n = 1$. Hence, Eq. (4.22) can be rewritten as

$$F(f) = 2\sum_{n=1}^{\infty}\phi(n)n^{-s} = 2\frac{\zeta(s-1)}{\zeta(s)}. \tag{4.23}$$

This result coincides with earlier obtained, Eq. (4.13), thus providing an independent support to earlier developed thermodynamic formalism.

5. Further developments

5.1. The circle packing and the Farey numbers

Rademacher [47] had found alternative geometric representation of the Farey numbers which is worth discussing now since it has physical significance which ultimately goes far beyond the leaky torus model. To facilitate our reader's understanding, we would like to remind at this point few relevant facts from the theory of Möbius transformations. To begin, using Eq. (2.38) we notice that the point ∞ is the fixed point of the Möbius (in our case, modular) transformation as long as $c = 0$. If this is the case and, taking into account that $ad = 1$, we obtain a transformation of the type

$$z' = a^2z + ab, \tag{5.1}$$

which will have another fixed point $z^* = ab/(1 - a^2) \neq \infty$ as long as $a \neq 1$. This cannot happen, however, as we had discussed in connection with Eq. (2.69). Therefore, the only possibility which is left to us is $a = 1$ and, hence, we are left with the parabolic transformation

$$z' = z + b. \tag{5.2}$$

This transformation fixes infinity and makes all integers equivalent (e.g. discussion after Eq. (2.58)). It is directly associated with the presence of a puncture (cusp) in the case of torus as can be easily proven [86]. In H^2 model, realization of hyperbolic space consider the geodesic (or the set of geodesics) which pass through the point at infinity. Surely, these are just semi-infinite rays which are perpendicular to the real axis. Accordingly, the associated horocycles are just the set of lines parallel to the real axis. Consider, in particular, the horocycle located at the vertical distance 1 from the real axis. The parabolic transformation, Eq. (5.2), is not going to change the location of such horocycle. But, in general, it is known, that the Möbius transformations transform lines into lines and circles into circles. Hence, for the above horocycle (actually a circle located at infinity) there must be a modular transformation which transforms a circle at infinity to a circle located at some point on the real axis. In particular, [86], the transformation $z \mapsto -1/z$ maps the horizontal horocycle at unit height to the horocycle resting at the origin and having diameter 1. This example can be easily generalized. To this purpose, it is convenient to rewrite the modular transformation,

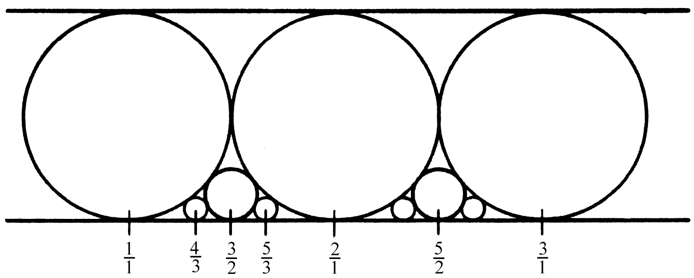

Fig. 10. Circle packing associated with the Farey numbers.

Eq. (2.37), as follows:

$$z' - \frac{a}{c} = -\frac{1}{c^2(z + d/c)}. \tag{5.3}$$

Let now $Z' = z' - a/c$ and $Z = c^2(z + d/c)$ then, we obtain back the transformation $Z' = -1/Z$. Let, now $z^* = Z/c^2$. By this transformation, a circle of radius $\frac{1}{2}$ becomes a circle of radius $1/2c^2$ touching the real axis at zero. Since c can be only an integer number it is clear that all radii are less or equal to $\frac{1}{2}$. Finally, by using the transformation $z = z^* - d/c$, we shift the point of tangency for such circle to the location $z = -d/c$. Consider a very special case first: $c = 1$. Then, our circle is going to touch the real axis at some integer point $-d$ of real axis. It is clear now that for different d's we would have different circles and that all these circles are going to touch each other. Moreover, and this is not difficult to prove [47] for c different from 1 the corresponding circles are all going to touch each other as it is depicted in Fig. 10. Clearly, Fig. 10 represents a very coarse picture since the number of circles is countable infinity between every neighboring pair (unit interval) of integers. Given this, one may think about the distribution of sizes of such circles. Following Sullivan [87], we say that two real numbers have the same ρ-size if they belong to one of the intervals (ρ^{n+1}, ρ^n). Then, we can group the circles into collections whose diameters have the same ρ-size. The number of circles of a given size $s = (\rho^{n+1}, \rho^n)$ within a unit interval is the number of pairs (p, q) with $p \le q$ and p relatively prime to q (so that $s \sim 1/q^2$) which is again the Euler $\phi(q)$ totient function, Eq. (1.5). This fact can be easily understood if we recognize that the horocycles which are located at positions p/q within unit interval are having sizes of order $1/q^2$.

Following Sullivan [87] (and also [88]), let us consider definitions of packing and covering Hausdorff measures. Adopted to our case, we have to consider a (closed) subset Λ of R^2 and to cover this subset by discs of radii $r_1, r_2, \ldots$, all less than some $\varepsilon \ge 0$. Consider now the sum $S_c = \sum_i \psi(r_i)$ where $\psi(r_i) = r_i^\delta$ with δ being some "critical exponent" (Hausdorff dimension) of the set Λ. The *covering* Hausdorff ψ-measure of Λ is the limit (as $\varepsilon \to 0$) of the *infimum* of S_c with the exponent δ being a fractal dimension of Λ. Analogously, one can define the *packing* Hausdorff measure. In this case one should consider an open set Λ' and to cover it by the set of disjoint (that is nontouching) circles. Then, one has to consider the *supremum* for the analogous sum S_p. In general, the packing and the covering Hausdorff dimensions are not the same. Such technicalities are needed for description of limit sets of Fuchsian (Möbius in general) groups with cusps (for a quick introduction to this subject, please consult our earlier work, [24]). Depending upon the rank of the cusp one should use either S_{c} or S_{p} [89]. At this point it is just sufficient to take into account that in our case, in view of the results just obtained,

$$\sum_{\text{circles}} (\text{size})^{\delta/2} = \sum_{n=1}^{\infty} \phi(n) n^{-\delta} = \frac{\zeta(\delta-1)}{\zeta(\delta)}. \tag{5.4}$$

Hence, we have reobtained again the result, Eq. (4.13), which now acquires completely new meaning. It should be apparent at this point, that one still can do much better if one recalls all relevant facts about the Patterson–Sullivan measure of the limit set Λ which we had discussed in our earlier work, [24], in connection with AdS/CFT correspondence.

5.2. The Eisenstein series and the S matrix

If $\rho(x, y)$ is the hyperbolic distance between points x and $y \in H^2$ then, the Patterson–Sullivan measure can be constructed with help of the Poincaré series $g_\delta(x, y)$ defined as

$$g_\delta(x, y) = \sum_{\gamma \in \Gamma} \exp\left(-\frac{\delta}{2}\rho(x, \gamma y)\right) \tag{5.5}$$

for some Fuchsian (or Möbius, in general) group Γ. The factor δ is responsible for convergence/divergence of g_δ and its threshold value is associated with the fractal dimension of the limit set Λ (which is closed (sub)set of the boundary at infinity, in our case, S^1_∞). According to the theorem of Beardon and Maskit [24, Theorem 5.1], the limit set Λ of the discrete group Γ is made of parabolic limit points (i.e. those which are associated with cusps) and conical limit points (i.e. those which are associated with the fixed points of hyperbolic elements of Γ) and the conical limit points always lie outside of the cusps.

This means that motion along the hyperbolic geodesics (e.g. see Sections 2–4) is accompanied with countable infinity of events associated with such geodesic entering and leaving the corresponding horocycle associated with the cusp as it is schematically depicted in Fig. 11. This fact was noticed in the paper of Sullivan [87] who also had estimated the characteristic time of this process (that is for the motion with unit speed along the geodesic (world line) one can estimate how long such motion spends inside the cusp). For additional

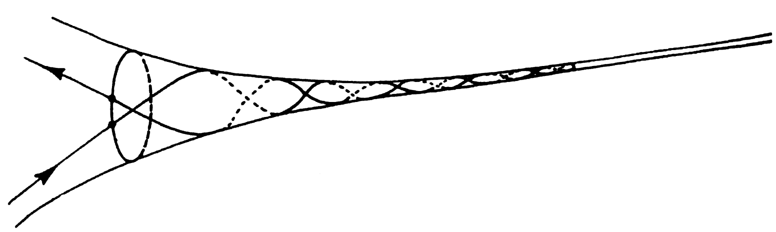

Fig. 11. Entering and leaving the "black hole" (the cusp) while moving along the hyperbolic geodesic.

illustration and more details, please, see Fig. 5 of our earlier work, [24], and the comments associated with it. Let our geodesic begins at the point z^* and ends at the point w^* of real axis. These points are chosen in such a way that it passes through the point z which is the top of horocycle located at the point $-d/c$ of real axis and through the point w of the horocycle located at infinity. Hence, the point $w = \mathrm{i} + x$, the point $z = \mathrm{i}y - d/c$ (and $y = 1/c^2$). It is always possible to find a transformation $g(z)$ such that $g(z^*) = 0$, $g(w^*) = \infty$, $g(z) = \mathrm{i}y$ and $g(w) = \mathrm{i}$ [90]. In view of this, we obtain,

$$\rho(z, w) = \ln\left(\frac{1}{y}\right). \tag{5.6}$$

Using this result in Eq. (5.5), we obtain

$$E(y, \delta) = \sum_{\gamma \in \Gamma} (\gamma y)^{\delta/2}. \tag{5.7}$$

This is just the Eisenstein series. Evidently, the subset Γ must correspond to subset of closed nonperipheral curves (i.e., it does not contain an accidental parabolic elements) on the torus which belong to $\hat{\Omega}$ (defined after Eq. (2.5)). Fortunately, the properties of the Eisenstein series, Eq. (5.7), are well known. This fact allows us not only to provide different interpretation to our main result, Eq. (4.13), but allows, in principle, to generalize the obtained results to the Riemann surfaces of higher genus. To understand why this is so, we would like to reproduce some results from our previous work, [24], at this time.

In the upper-half space model realization of $d + 1$-dimensional hyperbolic space H^{d+1} the hyperbolic Laplacian Δ_{h} acts on some function $f(\mathbf{x}, z)$ according to the following prescription:

$$\Delta_{\mathrm{h}} f(\mathbf{x}, z) = z^2 \left[\Delta f - (d-1)\frac{1}{z}\frac{\partial f}{\partial z}\right], \quad z > 0, \ \ \mathbf{x} \in \mathbf{R}^d. \tag{5.8}$$

In two dimensions, the second-term vanishes. We would like to keep it, nevertheless, since the obtained results can be immediately generalized (see below). For any $d \geq 1$, the eigenvalue equation for the hyperbolic Laplacian reads

$$\Delta_{\mathrm{h}} z^{\delta/2} = \frac{\delta}{2}\left(\frac{\delta}{2} - d\right) z^{\delta/2}. \tag{5.9}$$

In the case of two dimensions, of course, we have to replace z by y. If

$$\Delta_{\mathrm{h}} f(x) \equiv F(x), \tag{5.10}$$

and $x = \{\mathbf{x}, z\}$ then, for any $\gamma \in \Gamma$ where Γ is the group of isometries which leave H^{d+1} invariant, we obtain

$$\Delta_{\mathrm{h}} f(\gamma x) \equiv F(\gamma x). \tag{5.11}$$

This means that not only $z^{\delta/2}$ is an eigenvalue of $\Delta_h$but $(\gamma z)^{\delta/2}$ as well. Of course, the linear combination is also an eigenvalue of Δ_h. Thus, not only for $d = 1$ but **for any** $d \geq 1$

we obtain,

$$\Delta_{\mathrm{h}} E(z,\delta) = \frac{\delta}{2}\left(\frac{\delta}{2} - d\right) E(z,\delta). \tag{5.12}$$

Surely, for $d = 1$ we relabel z as y. In this case, following [91] we obtain,

$$E(y,\delta) = y^{\delta/2} + \varphi_\Gamma(\delta) y^{1-\delta/2} + \mathrm{O}(\mathrm{e}^{-2\pi y}), \tag{5.13}$$

which is sufficient for large y's. Here the scattering S matrix $\varphi_\Gamma(\delta)$ [11] is given by

$$\varphi_\Gamma(\delta) = \sqrt{\pi}\,\frac{\Gamma(\frac{1}{2}\delta - \frac{1}{2})\zeta(\delta-1)}{\Gamma(\frac{1}{2}\delta)\zeta(\delta)}. \tag{5.14}$$

This result should be compared with Eqs. (4.13) and (5.4). In order to generalize these results, we would like to mention several useful properties of the S matrix $\varphi_\Gamma(\delta)$. First, it can be shown [91,92] that

$$E(y,\delta) = \varphi_\Gamma(\delta) E(y, 1-\delta). \tag{5.15}$$

Using this result, we introduce new variable $\delta - \frac{1}{2} = \xi$. Eq. (5.15), when written in terms of this new variable, acquires the following more symmetric form:

$$E(y, \xi + \tfrac{1}{2}) = \varphi_\Gamma(\xi + \tfrac{1}{2}) E(y, \xi - \tfrac{1}{2}). \tag{5.16}$$

Let now $\xi \to -\xi$ in Eq. (5.16). By combining this obtained equation with Eq. (5.16), we obtain very important relation (unitarity condition) [92]:

$$\varphi_\Gamma(\xi + \tfrac{1}{2})\varphi_\Gamma(\xi - \tfrac{1}{2}) = 1. \tag{5.17}$$

Being armed with this results we can now proceed with generalizations.

5.3. Hyperbolic Dehn surgery once again

In this work, we only provide an outline of the relevant results leaving more detailed discussion for future publications. To begin, let us notice that so far we were able to obtain all results of this paper only because the Teichmüller space of the punctured torus happen to coincide with the hyperbolic upper plane Poincaré model H^2, so that the motion in the Teichmüller space coincides with the motion in the hyperbolic space H^2. This is not the case for the Riemann surfaces of higher genus [40] and, hence, at first sight, the results of this paper are not extendable to surfaces of higher genus. Very fortunately, this is not the case as we would like now to argue. Let us recall that we have started our discussion with the figure eight and the trefoil knots in Section 3. Then, we had considered the corresponding 3-manifolds created by the mapping torus construction and, after this, we had considered the incompressible surfaces (in Section 3.3). Three-manifolds associated with these surfaces had been obtained with help of the Dehn surgery (Dehn filling) from the “parent” 3-manifold (e.g. see Remark 3.8) which is just that for the figure eight complement in S^3. Neumann and Zagier [93] in the benchmark paper had developed a sort of perturbative calculations which allow to estimate volumes of 3-manifolds obtained from the parent 3-manifold by the operation of Dehn filling. In particular, for the decendants of the figure eight hyperbolic manifolds, they obtained the following estimate for the volume of $M(p, q)$ 3-manifold:

$$\mathrm{Vol}(M(p,q)) = \mathrm{Vol}(M_8) - \frac{2\sqrt{3}\pi^2}{p^2 + 12q^2} + \frac{4\sqrt{3}(p^4 - 72p^2q^2 + 144q^4)\pi^4}{3(p^2 + 12q^2)^4} + \cdots, \tag{5.18}$$

where $\mathrm{Vol}(M_8)$ is known to be 2.0298832 .... The obtained result is the simplest in the chain of results obtained in this reference. In general, $\mathrm{Vol}(M(p, q)) < \mathrm{Vol}(M)$ where M is parental 3-manifold. This constitutes the essence of Thurston's [18,75] Dehn surgery theorem.

In our previous work [24], we had considered 3-manifolds with cusps, in particular, the 3-manifold for the figure eight knot contains just one $\mathbf{Z} \oplus \mathbf{Z}$ cusp. It can be shown, e.g. see the Appendix C, that all hyperbolic 3-manifolds associated with knots and links are $\mathbf{Z} \oplus \mathbf{Z}$ cusped: one cusp for each embedded circle S^1. Hence, such manifolds are necessarily noncompact but, nevertheless, of finite volume. Surely, there are 3-manifolds without cusps too and these are related to those with cusps. According to Thurston [18, Section 5.33], all 3-manifolds $M(p, q)$ obtained by Dehn filling are without cusps. And, moreover,

$$\lim_{(p,q)\to\infty} \mathrm{Vol}(M(p,q)) = \mathrm{Vol}(M). \tag{5.19}$$

Obtained result is naturally extendable to the case of k-cusped manifolds. By introducing notations

$$M_k = M_{(p_1,q_1,\ldots,p_k,q_k)},$$

the following general result was obtained by Neumann and Zagier [93]

$$\mathrm{Vol}(M_k) = \mathrm{Vol}(M) - \frac{\pi}{2}\sum_{i=1}^{k} L_i + \mathrm{O}(L_i^2), \tag{5.20}$$

where L_i is the length of short geodesic γ_i on M_k which is just the length of the core curve of the solid torus added at ith cusp. The result of major importance for us is the observation by Neumann and Zagier that the difference of volumes depends to a high order only on the geometry of cusps and not on the rest of M [93]. Given this observation, the following question can be asked immediately: is it possible to reobtain the results Eqs. (5.13)–(5.15) for cusped 3-manifolds? The answer is Yes! if these manifolds belong to the arithmetic hyperbolic 3-manifolds. We provide a condensed summary of results related to such manifolds in the Appendices B and C while in the main text we discuss the scattering theory for such manifolds.

5.4. *Scattering theory for arithmetic 3-manifolds*

Let us begin with observation that Eq. (5.12) is valid for any $d \geq 1$. Accordingly, the asymptotic expansion, Eq. (5.13), also survives and acquires the following form (for H^3):

$$E(y,\delta)=y^{\delta/2}+\Phi_{\Gamma}(\delta)y^{2-d/2}+\mathrm{O}(\mathrm{e}^{-cy}) \tag{5.21}$$

with c being some constant. The question immediately arises: will the S matrix $\Phi_{\Gamma}(\delta)$ have the same form as in Eq. (5.14)? In general, the answer is “NO” but for the arithmetic 3-manifolds the answer is “Yes” as had been demonstrated by Sarnak [25]. It is instructive to restore some of his calculations now. In order to do so, we have to think about the ways the group $\mathrm{PSL}(2,\mathrm{Z})$ can be extended in order to be used for description of 3-manifolds associated with the punctured torus bundle.

To this purpose, we notice that the group $\mathrm{PSL}(2,\mathrm{Z})$ is just a subgroup of $\mathrm{PSL}(2,\mathrm{R})$ which is group of isometries of H^2. The group of isometries of H^3 is $\mathrm{PSL}(2,\mathrm{C})$ as is well known [24]. Hence, we have to figure out what kind of subgroup of $\mathrm{PSL}(2,\mathrm{C})$ is analogous to $\mathrm{PSL}(2,\mathrm{Z})$. This problem was actually solved for the torus bundles in [94] and for surfaces of higher genus in very important paper by Margulis [29] to be discussed further in Appendix C. For the torus bundles, naturally, one should try to find solutions of Eq. (2.7) in the complex domain. In [94], it was shown that the complex counterpart of the triple (3, 3, 3) is

$$\left(\frac{3+\sqrt{-3}}{2},\frac{3+\sqrt{-3}}{2},\frac{3-\sqrt{-3}}{2}\right), \tag{5.22}$$

and the rest of triples can also be obtained. The numbers above are integers which belong to the ring of integers $\mathcal{O}_3$ in the quadratic number field $\mathbf{Q}(\sqrt{-3})$, e.g. read Appendix B. Hence, for the punctured torus bundle the analog of $\mathrm{PSL}(2,\mathrm{Z})$ is $\mathrm{PSL}(2,\mathcal{O}_3)$ and the punctured torus fiber bundle is 3-orbifold $H^3/\mathrm{PSL}(2,\mathcal{O}_3)$ as was proven rigorously in [94]. The question remains: can this result be generalized to the noncompact (i.e. having punctures (or cusps)) 3-manifolds originating from automorphisms of Riemann surfaces of genus higher than one? The answer is “Yes”, provided that discrete group of isometries is Bianchi group $\mathrm{PSL}(2,\mathcal{O}_d)$, where

$$-d=1,2,3,5,6,7,10,11,14,15,19,23,31,35,39,47,55,71,95,119. \tag{5.23}$$

This remarkable result can be found in [95]. What is even more remarkable is that it is just a refinement of the very comprehensive work by Bianchi [27] which was completed already in 1892! Reid [96, Proposition 1], had proved the following theorem.

Theorem 5.1. *Every non-compact arithmetic Kleinian group is conjugate in* $\mathrm{PSL}(2,\mathrm{C})$ *to a group commensurable with some Bianchi group* $\mathrm{PSL}(2,\mathcal{O}_d)$.

Remark 5.2. (a) The definition of arithmeticity is rather involved and is discussed in the Appendices B and C. The above theorem is complementary to the theorem by Margulis (e.g. Theorem C.11 of Appendix C), which is much more comprehensive. (b) The noncompactness implies existence of the parabolic peripheral subgroups (e.g. see Section 2). (c) The notion of commensurability between groups can be found, e.g. in [97], and is associated with set theoretical definition of overlap (i.e. $\cap$) between sets.

Corollary 5.3. *Every cusped arithmetic hyperbolic* 3-*orbifold is commensurable with Bianchi orbifold* $H^3/\mathrm{PSL}(2, \mathcal{O}_d)$, *where* $\mathcal{O}_d$ *is ring of integers in the quadratic number field* $\mathbf{Q}(\sqrt{-d})$.

Remark 5.4. (a) There is one-to-one correspondence between the rings S^1 (in some (arithmetic) link) and the cusps, e.g. read Appendix C for more details. (b) The non-compactness is caused by cusps. (c) Cusped 3-manifolds do have finite volume (e.g. read Section 5.3 again). (d) The arguments of Section 5.1 strongly suggest that 3-manifold associated with the figure eight knot is arithmetic. This is rigorously proven by Reid [96], who also proved that the figure eight is the only knot which is arithmetic. (e) Corollary 5.3 and Eq. (5.23) suggest that there is a countable infinity of arithmetic links.

Being armed with these results, we can now discuss the results of Sarnak [25]. To this purpose, we have to rewrite Eq. (5.7) in a more convenient (for the present purposes) form. We have ($z = x + \mathrm{i}y$)

$$E(y, \delta) = y^{\delta/2} + \sum_{c=1}^{\infty} \sum_{(c,d)=1} \sum_{m=-\infty}^{\infty} \frac{1}{[(c(x+m)+d_0)^2 + c^2 y^2]^{\delta/2}}, \tag{5.24}$$

[11, pp. 171–172]. Here in the second sum summation is over d provided that c and d are relative primes, i.e. $(c, d) = 1$, $d = d_0 + mc$ and $0 \leq d_0 < c$. The summation procedure is explained in great detail in the same reference. In the Sarnak's case one needs only to replace x by z and y by t ($t > 0$) (in the H^3 model realization of hyperbolic space). The details of calculations can be found in [25] so that the final result, indeed, has the form given by Eq. (5.21) where for the quadratic number field $\mathbf{Q}(\sqrt{-d})$ of class number one (e.g. read Appendix B for explanation of terminology) $\Phi_\Gamma(\delta)$ is given by

$$\Phi_\Gamma(\delta) = \frac{\pi}{V(F_\mathrm{L})(\frac{1}{2}\delta - 1)} \frac{\zeta_d(\frac{1}{2}\delta - 1)}{\zeta_d(\frac{1}{2}\delta)} \tag{5.25}$$

with $V(F_\mathrm{L})$ being defined as

$$V(F_\mathrm{L}) = \begin{cases} \frac{1}{2}\sqrt{D} & \text{if } D = 3(\mathrm{mod}\, 4), \\ \sqrt{D} & \text{if } D \neq 3(\mathrm{mod}\, 4), \end{cases} \tag{5.26}$$

provided that

$$d = \begin{cases} -D & \text{if } D = 3(\mathrm{mod}\, 4), \\ -4D & \text{if } D \neq 3(\mathrm{mod}\, 4). \end{cases}$$

Since, according to Appendix B, Eq. (B.12), the Dedekind zeta function $\zeta_d(s)$ has the same pole as the ordinary zeta function, the result, Eq. (5.21), can be used instead of earlier obtained Eq. (5.14). This time, however, one can get much more. In particular, one can get the volume $V(F_\mathrm{D})$ of the associated cusped hyperbolic 3-manifold. The expression for the volume is known to be [18,25,97]

$$V(F_{\mathrm{D}}) = \frac{|d|^{3/2}\,\zeta_d(2)}{4\pi^2}. \tag{5.27}$$

This result can be obtained directly from the Eisenstein series, Eq. (5.24). It comes as the residue at the pole $\frac{1}{2}\delta = 2$ of the Dedekind zeta function in the numerator of Eq. (5.25) [25]. That is

$$\mathrm{Re}\, s\left[E(y,\delta), \frac{\delta}{2} = 2\right] = \frac{V(F_{\mathrm{L}})}{V(F_D)}. \tag{5.28}$$

Since both the residue and $V(F_{\mathrm{L}})$ are known, the volume, Eq. (5.27), is obtained using Eq. (5.28).

Consider now generalization of the obtained results to the case of multiple cusps, i.e. to the arithmetic links (e.g. see Appendix C). Following Efrat and Sarnak [26], and also [97], consider a complete set $\{\kappa_1 = \infty, \kappa_2, \dots, \kappa_n\}$ of the nonequivalent cusps and let Γ_i be a subgroup of $\Gamma \subset \mathrm{PSL}(2, \mathrm{C})$ for which κ_i is the fixed point, i.e. Γ_i is stabilizer in Γ of the ith cusp. It is always possible to select transformation $\rho_i(\kappa_i) = \infty$ where $\rho_i \in \Gamma$. Then,

$$\rho_i \Gamma_i \rho_i^{-1} = \begin{pmatrix} 1 & \omega \\ 0 & 1 \end{pmatrix}, \tag{5.29}$$

which coincides with the matrix U_1 defined in Eq. (C.1) with ω being a unit of quadratic number field. For each cusp one can define its own Eisenstein series analogous to Eq. (5.7). Also, by analogy with Eq. (5.21) we write now

$$E_i(y,\delta) = \delta_{i,j}\, y^{\delta/2} + \Phi_{i,j}(\delta)\, y^{2-\delta/2}, \quad i, j = 1, \dots, n. \tag{5.30}$$

It can be shown that the S matrix $\Phi_{i,j}$ obeys the matrix equation

$$\mathbf{E}(y,\delta) = \mathbf{\Phi}(\delta)\mathbf{E}(y, 2-\delta), \tag{5.31}$$

which is analogous to Eq. (5.15) discussed earlier. Finally, the analog of Eq. (5.14) is the determinant of the matrix $\mathbf{\Phi}(\delta)$ given by

$$\det \mathbf{\Phi}(\delta) = \mathrm{const}\,\frac{\xi_{\mathrm{H}}(\frac{1}{2}\delta - 1)}{\xi_{\mathrm{H}}(\frac{1}{2}\delta)} \tag{5.32}$$

to be compared with Eq. (5.25). $\xi_{\mathrm{H}}(s)$ for the quadratic number field of class number one is given by

$$\xi_{\mathrm{H}}(s) = \left(\frac{\sqrt{d}}{2\pi}\right)^s \Gamma(s)\zeta_{\mathrm{H}}(s) \tag{5.33}$$

with $\zeta_{\mathrm{H}}(s)$ being defined by

$$\zeta_{\mathrm{H}}(s) = \prod_{\chi} L(s,\chi). \tag{5.34}$$

It should be noted that the above product over characters (which are specific for each quadratic field) always starts with $\chi = 1$. This produces $L(s, 1) = \zeta(s)$ where, again,

$\zeta(s)$ is usual Riemann zeta function. Hence, again, the singularities of the determinant are just those for the ratio of zeta functions as in Eq. (5.4). Hence, the exact partition function for $2+1$ gravity is given by the determinant $\det \mathbf{\Phi}$, Eq. (5.32), of the scattering S matrix $\mathbf{\Phi}$.

Acknowledgements

The author would like to thank Jeff Weeks, Geometry Center, University of Minnesota, for encouragement and many useful references, Josef Przytycki, George Washington University, for sending his papers and the unpublished manuscript of T. Jorgensen, Bill Floyd, Virginia Tech, for many clarifying remarks related to the incompressible surfaces, Alan Reid, University of Texas at Austin, for sending chapters of his unpublished yet book on arithmetic hyperbolic manifolds, Scot Adams, University of Minnesota, for sending his papers related to Lorentz manifolds and arithmeticity and sharing the information about his planned book on this subject.

Appendix A. Phase transitions in $2 + 1$ gravity: analogy with Bose condensation

Although in Section 4 the limiting value of the partition function, Eq. (4.13), is obtained in a systematic way [8,9], the extraction of useful information from this result is plagued by some serious problems as was noticed already in earlier work of Cvitanovic [14]. The result, Eq. (4.13), is obtained as limiting value of the Farey spin chain partition function whose size tends to infinity. Since in Section 5 we generalize the punctured torus results to surfaces of higher genus, the analogy with spin chain cannot be straightforwardly extended. Hence, we need to develop methods of extraction of useful information without being dependent on spin chain analogy.

The main difficulty with Eq. (4.13) lies in the fact that it contains singularities. This can be easily demonstrated if we notice that for large values of n the Euler totient function $\phi(n) \sim n$ [44]. This observation allows us to write

$$\hat{Z}(\beta) \leq \zeta(\beta - 1). \tag{A.1}$$

Since $\zeta(\beta - 1)$ is singular at $\beta_{\mathrm{cr}} = 2$ one is faced with the problem of removing this singularity in physically acceptable way. There are many ways of doing so but it is not our purpose here to provide a complete list of possibilities. Rather, we would like to notice the following. First, by direct numerical calculation one observes that for large β's the partition function $\hat{Z}(\beta)$ approaches 1. So that, indeed, for low "temperatures" the free energy is zero in accord with earlier results [8,9]. The authors of these papers claim that $\hat{Z}(\beta)$ is zero as soon as $\beta > 2$. This can be understood only if one considers finite number terms in the corresponding zeta functions and takes the thermodynamic limit using Eq. (4.8). To be consistent, we need to do the same for $\beta < 2$. This is not a simple task, however, since according to Knauf [9, p. 428], "for high temperatures ($\beta < 2$), the analytic continuation of $\hat{Z}(\beta)$ cannot be directly interpreted as the partition function of the infinite chain". This statement is rather pessimistic, especially, if one is thinking about extension of obtained results to surfaces of higher genus. Therefore, we offer here a somewhat different interpretation of the obtained results.

To this purpose, let us introduce new function $g_\beta(\alpha)$ defined by the following equation:

$$g_\beta(\alpha) = \sum_{n=1}^{\infty} \frac{\alpha^n}{n^\beta} = \frac{1}{\Gamma(\beta)} \int_0^\infty \mathrm{d}x \, \frac{x^{\beta-1}}{\alpha^{-1}\,\mathrm{e}^x - 1}. \tag{A.2}$$

Such type of functions are well known from the theory of Bose–Einstein (B–E) condensation [98]. The analogy, unfortunately, is not complete since in the case of B–E condensation the exponent $\beta - 1$ is never zero or less than zero. Evidently, for $\alpha = 1$ we obtain back the Riemann's zeta function. The parameter α is related to the chemical potential μ in a usual way. For $\mu = 0$ we have $\alpha = 1$. This value of α is associated with singularity which has physical interpretation. In properly chosen system of units equation which connects the chemical potential with the particle density $\rho = N/V$ in B–E case reads [98]

$$\rho = g_{3/2}(\alpha) + \frac{1}{V}\frac{\alpha}{1-\alpha}, \tag{A.3}$$

where V is volume of the system and N the total number of particles. The second term in the r.h.s. represents the density of Bose condensate associated with zero translational mode. For finite volume this density is infinite for $\alpha = 1$. Therefore, as usual, it is assumed that both the volume and the density tend to infinity in such a way that the ratio is finite number. In our case, we do not have the luxury of having volume term but we can extract the singularity in a manner similar to the B–E gas. Taking into account Eqs. (A.1)–(A.3) we can subdivide the domain of integration into two parts: (a) $0 \le x \le 1$ and (b) $1 \le x \le \infty$. In the first case, we obtain

$$I_1 = \int_0^1 \mathrm{d}x \, x^{\beta-3} = \frac{1}{\beta - 2}, \tag{A.4}$$

which is expected singularity of the zeta function at already known critical value $\beta_{\mathrm{cr}} = 2$. As for the second part, we obtain (close to criticality)

$$I_2 = \int_1^\infty \mathrm{d}x \, \frac{x^{\beta-2}}{\mathrm{e}^x - 1} \simeq \int_1^\infty \mathrm{d}x \, \frac{1}{\mathrm{e}^x - 1} + (\beta - 2)\int_1^\infty \mathrm{d}x \, \frac{\ln x}{\mathrm{e}^x - 1} + \cdots . \tag{A.5}$$

Combining Eqs. (A.4) and (A.5) we obtain as well,

$$I = \frac{1}{\varepsilon}(1 + c_1\varepsilon)(1 + c_2(\beta)\varepsilon^2 + \cdots), \tag{A.6}$$

where $\varepsilon = \beta - 2$ and $c_1, c_2(\beta)$ are known constants. In particular, $0.373 < c_2 < 0.5438$ for $1 \le \beta \le 2$. Taking into account that $\Gamma(\beta)$ is nonsingular in the range of β's which is physically interesting, we can subtract the singular part along with well behaving regular part in order to get

$$-\beta\mathfrak{F}(\beta) = \ln(1 + c_2(\beta)\varepsilon^2 + \cdots) \ge c_2(\beta)\varepsilon^2. \tag{A.7}$$

This result coincides almost exactly with the estimate, Eq. (2.5), obtained in [8] (the constant c_2 is ≈ 0.25 in [8]) for the free energy in the high temperature "disordered" (or pseudo-Anosov) phase. Hence, the rest of arguments of [8] now can be used. In particular, we have

$$-\beta\mathfrak{F}(\beta) = \int_\beta^2 U(x)\,\mathrm{d}x \le U(\beta)(2-\beta). \tag{A.8}$$

Using this result, we obtain

$$U(\beta) \ge -\frac{\beta\mathfrak{F}(\beta)}{(2-\beta)} \ge c_2(\beta)(2-\beta), \quad 1 \le \beta \le 2, \tag{A.9}$$

in complete accord with Eq. (4.15) of the main text.

Remark A.1. (a) The analogy with Bose gas condensation can be strengthened, perhaps, even more if one notices that the criticality condition $\beta_{\mathrm{cr}} = 2$ coincides with the simple estimate made in famous Kosterlitz and Thouless paper [22]. In this paper, the Coulomb gas model is used for description of phase transitions in two-dimensional liquid crystals, liquid helium, etc. This, in part, explains the success of such type of models prevailing so far in physics literature [19]. (b) In the case of liquid helium Feynman [99] had obtained equation very similar to Eq. (A.3) and Landau and Lifshitz [100, Section 27] also resort to a simple subtraction of the undesirable singularity at zero temperatures. Hence, intuitive physical arguments may provide a reliable guidance needed for further refinements of the results just presented.

Appendix B. Some results from the algebraic number theory

Theory of arithmetic hyperbolic manifolds has strong connections with the algebraic number theory [101,102]. In this section, we provide a condensed summary of ideas on which such theory is based. Selection of topics is, naturally, subjective and is meant only to encourage interested reader to read much more comprehensive texts.

Development of number theory is associated with the desire to extend a simple idea, known to nonprofessionals, that every nonnegative integer n can be uniquely represented as

$$n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}, \tag{B.1}$$

where p_i is some prime number and $\alpha_i \ge 0$ is some integer. This result becomes much less obvious if one is willing to extend it, say, into domain of complex or algebraic numbers. For instance,

$$6 = 2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5}) = (4+\sqrt{10})(4-\sqrt{10}). \tag{B.2}$$

Such nonuniqueness is highly undesirable since many theorems of usual arithmetic breakdown. Solution of the nonuniqueness problem is one of the major tasks of the modern number theory.

Let us define a rational number $\xi = p/q$ as a root of the equation

$$q\xi - p = 0 \tag{B.3}$$

with p and q being some integers. An integer is then a root of an equation with first coefficient $q = 1$. Analogously, an algebraic number ξ is any root of an algebraic equation

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0 = 0, \tag{B.4}$$

where the coefficients $a_n, \dots, a_0$ are rational integers. Accordingly, an algebraic integer is any root of the monic polynomial, e.g. see Eq. (3.22).

Remark B.1. Since only fibered knots and links are relevant for dynamics of $2+1$ gravity, e.g. read Section 3, and since such knots/links are associated with monic Alexander polynomials, we conclude that the arithmeticity is an intrinsic property of $2+1$ gravity. Surely, much more is associated with this observation as we shall demonstrate shortly.

In particular, let us consider the module (i.e. the set which is closed under operations of addition and subtraction) of quadratic integers, i.e. those, originating as roots of quadratic equation with $a_2 = 1$. We had discussed the roots of such equations already, e.g. see Eqs. (3.2), (3.9), (3.18) and (3.35). Now we know that all these roots are quadratic integers. As with usual (rational) integers, it is important to define the units, especially if we would like to go from the module to a ring (which is module where the multiplication operation is defined which keeps all the members of the set within the set) or to a field (where, in addition, the operation of division is defined). It is clear that the typical representative of a quadratic integer ξ should look like

$$\xi = \frac{a + b\sqrt{d}}{c}, \tag{B.5}$$

where a, b, c and d are some integers. Moreover, more detailed analysis shows that d can be chosen as square-free. Let us notice, that in the case if $d = -1$, we obtain more familiar field of complex numbers. Geometrically, the integer lattice in such field can be built using the basis $[1, i]$. Therefore, in general case we can think of a basis $[1, \omega]$ where

$$\omega = \tfrac{1}{2}(d + \sqrt{d}) \tag{B.6}$$

with $d = D$ if $D = 2, 3$ or $d = 4D$ if $D \equiv 1 \pmod 4$ e.g. see Eq. (B.8) for explanation of notations. The square-free number d is called *discriminant* of the field. This result confirms, in particular, that the Markov triple in Eq. (5.22) is indeed made of integers since $\frac{1}{2}(3-\sqrt{-3})$ is just the conjugate of $\frac{1}{2}(3+\sqrt{-3})$. For different d's one will have different rings of quadratic integers. To make rings, once the integers are defined, one should also define the units. This can be done again by analogy with complex analysis. Using Eq. (B.5) we define a norm $N(\xi)$ as

$$N(\xi) = \xi \cdot \bar{\xi}, \tag{B.7}$$

where $\bar{\xi}$ is conjugate of ξ, i.e. the same as ξ in Eq. (B.5) but with $b \to -b$. In the number theory the norm can be both positive and negative. For quadratic imaginary field it is positive, however. By definition, units are such numbers whose norm is unity. In the case of usual complex numbers, one has therefore four units: ± 1 and $\pm$i, since in both cases the norm is one. In the case of $d = -3$ ring the units are $\pm 1, \pm\rho, \pm(1+\rho)$ and ρ^2 where $\rho = \frac{1}{2}(-1+\sqrt{-3})$, in all other cases the number of units is two [101, p. 156]. The role of units can be easily understood by analogy with more familiar case of ordinary integers. That is if a and b are integers then, $a = 1 \cdot b = b \cdot 1$. Units play the same role for the ring of quadratic numbers. These numbers are called *associates* if they differ by unit factor. Next, one defines *primes* as those quadratic numbers whose norm is a rational prime. In particular, we notice, that the Markov triple, Eq. (5.22), is made of primes. Once the primes are defined, one is naturally interested in finding an analog of Eq. (B.1) for quadratic integers. This happens to be not a simple task: not all d's will lead to unique decomposition into primes. For reasons which will become clear upon reading of Appendix C, in the case of gravity only negative d's are relevant. Among those which allow unique prime decomposition are: $-d = 1, 2, 3, 7, 11, 19, 43, 67$ and 163 [101,102] to be compared with Eq. (5.23). Such comparison indicates that some d's listed in Eq. (5.23) suffer from nonuniqueness. Such nonuniqueness had lead number theoreticians to introduce the concepts of ideals and class numbers. Class numbers reflect the extent to which integers of the field deviate from uniqueness of decomposition. Naturally, if the decomposition is unique, the class number is one. The ideals can be defined set theoretically and geometrically as follows. Suppose, we have some set I and a larger set L, then $\forall\,\alpha, \beta \in I$ and $\forall \xi \in L$, we have

$$\alpha + \beta \in I \quad (\text{module property}), \qquad \alpha\xi \in I \quad (\text{ideal property}).$$

Geometrically, this can be understood as follows. Consider, without loss of generality, two-dimensional lattice $\mathfrak{L}$ built on vectors $[1, \omega]$ then, for any rational integer n the point $[n, n\omega]$ belongs to $\mathfrak{L}$. Now one can think of different classes of rational integers and, hence, of different sublattices $\mathfrak{M} \subset \mathfrak{L}$. This is easily accomplished by introducing the residue classes modulo m. Recall, that notation (equivalence relation)

$$x \equiv y \pmod m \tag{B.8}$$

means that $x - y = mk$ where k is some integer. The above notation is just the statement that when both x and y are divided by m they both have the same residues. In particular, the statement $x \equiv 0 (\bmod\, m)$ means that m divides x. Let m have the same presentation as n in Eq. (B.1), then any number x relatively prime to m may be determined $(\bmod\, m)$ by the equations $x \equiv x_i (\bmod\, p_i^{a_i})$, $(x_i, p_i) = 1$, $i = 1, 2, \dots, k$. The number of residue classes is just the Euler ϕ-function $\phi(m)$. Now, if for a given sublattice $\mathfrak{M}$ and vectors $\mathbf{x}$ and $\mathbf{y} \in \mathfrak{L}$ we have equivalence relation

$$\mathbf{x} \equiv \mathbf{y} \pmod{\mathfrak{M}}, \tag{B.9}$$

this means only that $\mathbf{x} - \mathbf{y} \in \mathfrak{M}$. The number of different sublattices within given lattice (i.e. the number of different residue classes for ideals) is called index j. Evidently, $j = [\mathfrak{L}/\mathfrak{M}]$.

Equivalently, it is also called a norm of an ideal $\mathfrak{L}$ so that $j = \mathbf{N}[\mathfrak{L}]$. It happens, that in the case of quadratic fields calculation of the norm is relatively easy [102]. If a is quadratic or rational integer, then $\mathbf{N}[\mathfrak{L}] = a^2 = |N(a)|$. For both ideals and integers the respective norms obey the following composition law:

$$\mathbf{N}(\mathfrak{a})\mathbf{N}(\mathfrak{b}) = \mathbf{N}(\mathfrak{a}\mathfrak{b}), \tag{B.10}$$

$$N(a)N(b) = N(ab), \tag{B.11}$$

where the Gothic letters stand for the ideals and the italics for numbers. The question arises: if there are prime numbers are there prime ideals? The answer is "Yes" and since prime numbers play the central role in theory of rational integers, one should expect that the prime ideals play the same role in the number theory. This is indeed the case, but, unlike numbers, every ideal is decomposable into prime ideals uniquely. This is the main reason why these objects were introduced.

Remark B.2. In the case of "quadratic fields" there is one-to-one correspondence between the "quadratic ideals" and the quadratic forms [101]. We deliberately omit discussion of this fact to save the space referring interested readers to literature [101–103].

Without going into details of such decomposition, we would like to take advantage of the composition laws, Eqs. (B.10) and (B.11). To this purpose we need the following theorem.

Theorem B.3. *The rational prime p factors in the quadratic number field* $\mathbf{Q}(\sqrt{d})$ (*irrespective to the sign of d*), *with accuracy up to multiplication by unit, as follows*:

1. *p does not factor* (*p is inert in the field d*) *if and only if* $(d/p) = -1$,
2. *p splits in d into two different factors* (*e.g. z and* z' *which belong to* $\mathbf{Q}(\sqrt{d})$) *if and only if* $(d/p) = -1$,
3. *p is ramified in d* (*i.e.* $p = z^2$) *if and only if* $(d/p) = 0$.

In the above cases (d/p) is the Kronecker symbol which sometimes is denoted as $\chi(p)$ for fixed d. $\chi(p)$ is actually a character (the Dirichlet's character to be exact) of an Abelian group, e.g. of $\mathfrak{L}/\mathfrak{M}$. This symbol should not be confused with δ_{ij} known in physics. At the same time, the above theorem may serve as a definition of such a symbol. We shall adopt this point of view in this work. Then, by analogy with the theorem just stated, one can formulate the analogous theorem for the ideals [102].

Theorem B.4. *The quadratic prime ideals* $(\mathfrak{p})$ *are related to integers p in the rational field in the following possible ways*:

1. $(\mathfrak{p}) = (\mathfrak{p})$, *or* $(\mathfrak{p})$ *does not factor, i.e. inert, then* $\mathbf{N}(\mathfrak{p}) = p^2$;
2. $(\mathfrak{p}) = \mathfrak{p}_1\mathfrak{p}_2$, *or* $(\mathfrak{p})$ *splits, then* $\mathbf{N}(\mathfrak{p}_1) = \mathbf{N}(\mathfrak{p}_2) = p$;
3. $(\mathfrak{p}) = \mathfrak{p}_1^2$, *or* $(\mathfrak{p})$ *ramifies, then* $\mathbf{N}(\mathfrak{p}_1) = p$.

Being armed with such results, we are ready to introduce the zeta function for the field whose discriminant is d:

$$\zeta_d(s) = \sum_{\mathfrak{a}} [\mathbf{N}(\mathfrak{a})]^{-s} = \prod_{\mathfrak{p}} (1 - [\mathbf{N}(\mathfrak{p})]^{-s})^{-1}, \tag{B.12}$$

where the last product is over all prime ideals. In view of Theorem B.4, this result can be rewritten in more familiar terms as follows:

$$\zeta_d(s) = \zeta(s) \sum_{n=1}^{\infty} \frac{(d/n)}{n^s} \equiv \zeta(s) L(s; d), \tag{B.13}$$

where $L(s, d)$ is known in the literature as the Dirichlet L-series. It can be shown [101–103] that such defined zeta function is having a pole singularity only at $s = 1$, i.e. the singularity only comes from the usual zeta function. The residue of the pole at $s = 1$ is much more interesting in the present case as it is explained in the main text in Section 5. Finally, we need an ideal analog of the Euler totient function $\phi(n)$. To this purpose, we take into account that an analog of Eq. (B.1) is the same equation in which all numbers are replaced by the ideals. Then, for an ideal $\mathfrak{a}$ we have

$$\mathbf{\Phi}(\mathfrak{a}) = \mathbf{N}(\mathfrak{a}) \prod_{\mathfrak{p}/\mathfrak{a}} \left(1 - \frac{1}{\mathbf{N}(\mathfrak{p})}\right), \tag{B.14}$$

where $\mathfrak{p}$ runs through the distinct prime divisors of $\mathfrak{a}$.

Appendix C. Connections between the knot theory, theory of hyperbolic spaces and the algebraic number theory: implications for $2 + 1$ and $3 + 1$ gravity

For the sake of saving space and to avoid repetitions, we expect our readers to be familiar with our earlier work [24], while reading this section. Let us begin with the group $\mathrm{PSL}(2, \mathrm{C})$ of isometries of hyperbolic space H^3. Only discrete subgroups of $\mathrm{PSL}(2, \mathrm{C})$ known as Kleinian groups are of interest [97]. The group $\mathrm{PSL}(2, \mathrm{C})$ is just the projectivized version of $\mathrm{SL}(2, \mathrm{C})$. The following theorem for $\mathrm{SL}(2, \mathrm{C})$ is of particular interest [97, Chapter 1].

Theorem C.1. *The group SL(2,C) is generated by two elements*

$$U_1 = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \qquad U_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad a \in \mathbf{C}. \tag{C.1}$$

Remark C.2. This theorem is an analog of the well-known fact [104] that the group $\mathrm{PSL}(2, \mathrm{Z})$ can be generated by the same two elements (upon projectivization) with $a = 1$. Hence, the parabolic elements play a very special role in $\mathrm{PSL}(2, \mathrm{C})$ too.

The parabolic subgroups of $\mathrm{PSL}(2, \mathrm{C})$ are Abelian subgroups associated with $\mathbf{Z}$ and $\mathbf{Z} \oplus \mathbf{Z}$ cusps. If $\Gamma \in \mathrm{PSL}(2, \mathrm{C})$ contains an Abelian subgroup of rank one, i.e. $\mathbf{Z}$-cusp, then the associated fundamental domain for Γ is non-compact and of infinite volume [28]. Hence, of no interest to us (additional reasons are given below). If, however, $\Gamma \in \mathrm{PSL}(2, \mathrm{C})$ contains an Abelian subgroup of rank two, i.e. $\mathbf{Z} \oplus \mathbf{Z}$ cusp, then the group Γ is not cocompact but of finite covolume. The following theorem was used in our earlier work, [24], and can be stated now as follows:

Theorem C.3. *Let $\Gamma \in \mathrm{PSL}(2, \mathrm{C})$ be a discrete group of finite covolume, then Γ has only finitely many classes of cusps.*

The proof can be found in [97, Chapter 2, Proposition 3.8]. To insure that the subgroup Γ is discrete the following theorem is the most helpful [97].

Theorem C.4 (Jorgensen's inequality). *If $\langle A, B\rangle$ is non-elementary discrete subgroup of* $\mathrm{PSL}(2, \mathrm{C})$*, then, the following inequality holds*:

$$|\mathrm{tr}^2 A - 4| + |\mathrm{tr}[A, B] - 2| \geq 1. \tag{C.2}$$

Remark C.5. (a) In the case of punctured torus extension of Markov triples Eqs. (2.7) and (2.8) to the field of quadratic integers was discussed in Appendix B and in the original works [34,94]. Such an extension is in accord with Jorgensen's inequality. (b) This observation provides us with an easy way of defining the arithmetic Kleinian groups and, hence, the arithmetic 3-manifolds. To this purpose, we need still one more theorem [94].

Theorem C.6. *Let $A_1, A_2, \ldots, A_n \in \mathrm{SL}(2, \mathrm{C})$ be such that A_1 is not parabolic and $\langle A_1, A_2\rangle$ is not elementary. Then, the $A_1, A_2, \ldots, A_n$ are determined up to conjugacy in $\mathrm{SL}(2, \mathrm{C})$ by $\mathrm{Tr}\, A_i$, $\mathrm{Tr}\, A_i A_j$ and $\mathrm{Tr}\, A_i A_j A_k$ with $i, j, k \in \{1, 2, \ldots, n\}$.*

Remark C.7. The arithmetic Kleinian groups can be regarded as images in $\mathrm{PSL}(2, \mathrm{C})$ of subgroups of $\mathrm{SL}(2, \mathrm{C})$ so that the trace of an element in the Kleinian group is defined up to a sign. The traces $\mathrm{Tr}\, A_i$ belong to the field of quadratic integers with negative discriminant d. The negativity of d is explained shortly below.

Following [105, Theorem E.4.4, p. 192], and also Riley [28] consider complement of a link $\mathcal{L}$ in S^3 (which is just a union of knots $\mathcal{L}_1, \ldots, \mathcal{L}_k$). To this purpose, let T_i be an open tubular neighborhood of the knot $\mathcal{L}_i$ (that is T_i is a solid torus with boundary removed). Then, the following theorem holds.

Theorem C.8. *Given a compact 3-manifold N there exists a link $\mathcal{L}$ in S^3 which is union of knots $\mathcal{L}_1, \ldots, \mathcal{L}_k$, such that N is obtained from the manifold*

$$M = S^3 \setminus \left(\bigcup_{i=1,\ldots,k} T_i \right),$$

where T_i's are pairwise disjoint open tubular neighborhoods of the knots $\mathcal{L}_i$, by the operation of Dehn filling.

Corollary C.9 (Riley and Thurston [18,28]). *If M is to be hyperbolic, then every non-cyclic Abelian subgroup (of rank two) of $\pi_1 M$ must be peripheral and correspond to the component $\mathcal{L}_i$ of link $\mathcal{L}$.*

Remark C.10. Since the fundamental group $\pi_1 M$ can be embedded into $\mathrm{PSL}(2, \mathrm{C})$ and since the non-cyclic peripheral (see e.g. Section 2 for definition) Abelian subgroup of rank two corresponds to $\mathbf{Z} \oplus \mathbf{Z}$ cusp there is a one-to-one correspondence between the cusps and the link components. E.g. in the case of figure eight knot we have hyperbolic 3-manifold with just one cusp [18,28]. Moreover, and this is the most important, the following theorem of Margulis [29] requires such cusped 3-manifolds to be arithmetic.

Theorem C.11 (G. Margulis). *Let S be Riemannian symmetric noncompact space of rank greater than one and Λ is irreducible discrete group of motions of S such that S/Λ has finite volume but noncompact. Then, Λ is the arithmetic subgroup in the group of motions of space S.*

Definition C.12 (Helgason [30], Chapter 5). Let S be the Riemannian globally symmetric space. The rank of S is the maximal dimension of a flat, totally geodesic subspace M. The subspace M is geodesic at $p \in M$ if each S-geodesic which is tangent to M at p is a curve in M. The submanifold M is totally geodesic if it is geodesic at each of its points.

Remark C.13. To use the above definition, the following alternative interpretation of rank is helpful. Following Besse [31], let $\mathfrak{g} = \mathfrak{h} + \mathfrak{m}$ be decomposition (into irreducible parts) of the Lie algebra of the group of motions of space S where $\mathfrak{h}$ is non and $\mathfrak{m}$ is Abelian parts. Then, the rank is the total dimension of the Abelian subalgebra $\mathfrak{m}$ of $\mathfrak{g}$.

Remark C.14. Since H^3 is symmetric space, Margulis theorem implies arithmeticity of hyperbolic 3-manifolds with $\mathbf{Z} \oplus \mathbf{Z}$ cusps. In the case of one cusp the complement of figure eight knot hyperbolic 3-manifold is indeed arithmetic as it was shown independently by Reid [96]. The arithmeticity can actually be anticipated based on the results of our Section 3. Indeed, since surface automorphisms are associated with the roots of the monic Alexander polynomial whose coefficients are rational integers, these roots lie necessarily in the field of quadratic integers. The Seifert-fibered phase is associated with units of the field while the pseudo-Anosov with nonunit integers. Since these stretch factors are related to the traces of the corresponding matrices, e.g. see Eq. (2.15), connection with Theorem C.6 is clear.

Remark C.15. Since Margulis theorem is not limited to groups of motions of H^3, its application can be extended to $3 + 1$ gravity. According to Besse [31, Chapter 10] every irreducible symmetric space is Einstein space. The classification of irreducible symmetric spaces was made by Helgason [30] and is also cited in Besse's book. From this classification it follows that most of such spaces have rank higher than 1 and, hence, should be arithmetic.

It would be interesting to know if there is an analog of Theorem C.3 for Einstein spaces. This is of interest because, from the discussion we had in Section 5, it follows that the cusps act very much like black holes, e.g. see Fig. 11 (except in $2+1$ case the "particle" cannot be trapped by the "black hole") and, therefore, the role of black holes in the Universe acquires new meaning: they make our Universe arithmetic. If this is the case, then the Universe becomes something like a crystalline solid which potentially can undergo phase transitions between lattices of different symmetry. The symmetry of the lattice is determined by the discriminant of the quadratic number field.

Remark C.16. The connection between the cusps and the black holes is not purely visual, e.g. Fig. 11, or formal. It is real, at least for $2+1$ Euclidean black hole! Indeed, following Carlip [21, p. 50] by the appropriate choice of coordinates and units the metric of the Euclidean black hole becomes that of H^3, i.e.

$$(\mathrm{d}s)^2 = \frac{(\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2}{z^2},$$

provided that in addition the following identification is made:

$$(x, y, z) \sim \mathrm{e}^{2\pi r_+}(x \cosh 2\pi r_+ - y \sinh 2\pi r_+, y \cosh 2\pi r_+ - x \sin nh 2\pi r_+, z),$$

which is easily recognizable as $\mathbf{Z} \oplus \mathbf{Z}$ cusp [24].

Remark C.17. The reasons for existence of black holes in nature were discussed in the past [106]. Ironically, the same author had came to the opposite conclusions in [107], i.e. the black hole singularity is fictitious. The same conclusion is reached in [108] which claims that two-dimensional quantum black holes are nonsingular.

Now we have to explain why this discriminant should be negative. For one thing, if it would be positive the norm of the quadratic field would be negative. This is unphysical in view of Remark C.14. There are other reasons in addition which we would like now to explain. Let us begin with famous example by Riley [109] who considered the mapping of the fundamental group π_8 of the complement of figure eight knot into $\mathrm{PSL}(2, \mathrm{C})$. The fundamental group π_8 is free non-Abelian group composed of two generators which has the following presentation:

$$\pi_8 = \langle x_1, x_2; w x_1 = x_2 w\rangle, \tag{C.3}$$

where $w = x_1^{-1} x_2 x_1 x_2^{-1}$. The homomorphism of π_8 into $\mathrm{PSL}(2, \mathrm{C})$ can be achieved with help of the Theorem C.1 as follows:

$$x_1 \to A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad x_2 \to B = \begin{pmatrix} 1 & 0 \\ -\omega & 1 \end{pmatrix}, \tag{C.4}$$

where ω is some number, different from 1 and to be determined below. The matrix B is just a conjugate of U_2 in Eq. (C.1), i.e. $B = U_1 U_2 U_1^{-1}$ with $a \leftrightarrow \omega$. Since these matrices should obey the relation $wx_1 = x_2 w$, replacements $x_1 \to A$ and $x_2 \to B$ produce the following equation for ω:

$$\omega^2 + \omega + 1 = 0 \tag{C.5}$$

with solution $\omega = \frac{1}{2}(-1 + \sqrt{-3})$. One can easily recognize the unit ρ of the quadratic field with discriminant 3 (from Appendix B). Hence, figure eight knot is indeed arithmetic. From this example, it should be clear, that the arithmeticity is associated with our choice of ω.

Consider now the subgroup of $\mathrm{PSL}(2, \mathrm{C})$ which fixes j (i.e. in the upper space model H^3 with coordinates $z + \mathrm{j}t$ ($z = x + \mathrm{i}y$, $t > 0$) we are looking for the Möbius transformation $M(j) = j$). It is easy to prove [97] that such transformation is associated with $\mathrm{SL}(2, \mathrm{C})$ matrix of the type

$$M = \begin{pmatrix} x & y \\ -\bar{y} & \bar{x} \end{pmatrix}, \tag{C.6}$$

such that $\det M = |x|^2 + |y|^2 = 1$ where $|x|^2 = x\bar{x}$ and $\bar{x}$ is a complex conjugate of x, etc. Such matrix M is in one-to-one correspondence with the field $\mathcal{H}$ of quaternions q known in physics literature as Hamiltonian quaternions [110]. Such quaternions usually are represented as $q = a + b\mathrm{i} + c\mathrm{j} + d\mathrm{k}$ where a, b, c, d are real numbers and $\mathrm{k} = \mathrm{ij}$. In terms of quaternions Eq. (C.6) can be written as

$$M(q, \mathcal{H}) = \begin{pmatrix} a + b\mathrm{i} & c + d\mathrm{i} \\ -c + d\mathrm{i} & a - b\mathrm{i} \end{pmatrix}. \tag{C.7}$$

Surely, if we write $\mathrm{i} = \sqrt{-1}$, then the above matrix represents the most general matrix associated with the unit of the imaginary quadratic field with discriminant $d = -1$. Naturally, one can therefore think of the most general matrix associated with units of imaginary quadratic fields with discriminant d other than -1. To this purpose, one needs to extend the notion of quaternions known in physics.

Let K be the number field of characteristic different from two (the characteristic p of the field K is associated with subdivision of the field into residue classes Z_p, so $p \neq 2$ and p is prime) and let $a, b \in K$ be two non-zero elements. A quaternion algebra $\mathcal{H}(a, b; K)$ over K is generated by elements i and j satisfying $\mathrm{i}^2 = a$, $\mathrm{j}^2 = b$ and $\mathrm{ij} = -\mathrm{ij}$. The elements 1, i, j, ij form a K-basis of $\mathcal{H}(a, b; K)$ as a vector space. The pair (a, b) is called the Hilbert symbol for $\mathcal{H}$. In particular, for Hamiltonian quaternions the field is $\mathbf{R}$ and the Hilbert symbol is $(-1, -1)$. If L is the field extension of K with $\sqrt{a}$ and $\sqrt{b} \in L$ then, repeating all the steps leading to Eq. (C.7), we obtain

$$M(2, L) = \begin{pmatrix} x_0 + x_1\sqrt{a} & x_2\sqrt{b} + x_3\sqrt{ab} \\ x_2\sqrt{b} - x_3\sqrt{ab} & x_0 - x_1\sqrt{a} \end{pmatrix}. \tag{C.8}$$

Newman and Reid [111] proved the following major theorem.

Theorem C.18. *Let Γ be a non-compact Kleinian group of finite covolume. Then*

1. *Γ is arithmetic if and only if the trace field* (*e.g. see Theorem* C.6) $K = \mathbf{Q}(\sqrt{-d})$ *for some square-free* $d \in N$ (*e.g. see Eq.* (5.23)) *and* $\mathrm{tr}\,\Gamma$ *consists of algebraic integers*.
2. *Γ is derived from the quaternion algebra* $\mathcal{H}(a, b; K)$ *if and only if* $\mathrm{tr}\,\Gamma \subset \mathcal{O}_d$ *for some* d.

References

[1] E. Titchmarsh, The Theory of the Riemann Zeta Function, Clarendon Press, Oxford, 1986.
[2] H. Edwards, Riemann's Zeta Function, Academic Press, New York, 1974.
[3] S. Patterson, An Introduction to the Theory of the Riemann Zeta Function, Cambridge University Press, Cambridge, 1988.
[4] M. Berry, J. Keating, The Riemann zeros and eigenvalue asymptotics, SIAM Rev. 41 (1999) 236–266.
[5] A. Connes, Trace formula in noncommutative geometry and zeros of the Riemann zeta function, Sel. Math. New Ser. 5 (1999) 23–106.
[6] C. Yang, T. Lee, Statistical theory of equations of state and phase transitions, Phys. Rev. 87 (1952) 404–419.
[7] A. Knauf, On a ferromagnetic spin chain, Commun. Math. Phys. 153 (1993) 77–115.
[8] A. Kontucchi, A. Knauf, The phase transition of the number-theoretic spin chain, Forum Mathematicum 9 (1997) 547–567.
[9] A. Knauf, Phases of the number-theoretic spin chain, J. Statist. Phys. 73 (1993) 423–431.
[10] B. Pavlov, L. Faddeev, Scattering theory and automorphic functions, J. Sov. Math. 3 (1975) 522–548.
[11] P. Lax, R. Phillips, Scattering Theory for Automorphic Functions, Princeton University Press, Princeton, NJ, 1976.
[12] M. Gutzwiller, Stochastic behavior in quantum scattering, Physica D 7 (1983) 341–355.
[13] R. Artuso, P. Cvitanovic, B. Kenny, Phase transitions on strange irrational sets, Phys. Rev. A 39 (1989) 268–281.
[14] P. Cvitanovic, Circle maps: irrationally winding, in: M. Waldschmidt, P. Moussa, J.-M. Luck, C. Itzykson (Eds.), From Number Theory to Physics, Springer, Berlin, 1992.
[15] A. Kholodenko, Use of meanders and train tracks for description of defects and textures in liquid crystals and $2 + 1$ gravity, J. Geom. Phys. 33 (2000) 23–58.
[16] A. Kholodenko, Use of quadratic differentials for description of defects and textures in liquid crystals and $2 + 1$ gravity, J. Geom. Phys. 33 (2000) 59–102.
[17] G. Brude, H. Ziechang, Knots, Walter de Gruyter, Berlin, 1985.
[18] W. Thurston, Geometry and Topology of 3-manifolds, Princeton University Lecture Notes, 1979. http://www.msri.org/gt3m/.
[19] P. Chaikin, T. Lubensky, Principles of Condensed Matter Physics, Cambridge University Press, Cambridge, 1995.
[20] C. Godbillon, Feuilletages, Birkhauser, Boston, MA, 1991.
[21] S. Carlip, Quantum Gravity in $2 + 1$ Dimensions, Cambridge University Press, Cambridge, 1998.
[22] J. Kosterlitz, D. Thouless, Metastability and phase transitions in two-dimensional systems, J. Phys. C 6 (1973) 1181–1203.
[23] J. Peyriere, Trace maps, in: F. Axel, D. Gratias (Eds.), Beyond Quasicrystals, Springer, Berlin, 1995, pp. 465–480.
[24] A. Kholodenko, Boundary conformal field theories, limit sets of Kleinian groups and holography, J. Geom. Phys. 35 (2000) 193–238.
[25] P. Sarnak, The arithmetic and geometry of some hyperbolic 3-manifolds, Acta Math. 151 (1983) 253–295.
[26] I. Efrat, P. Sarnak, The determinant of the Eisenstein matrix and Hilbert class fields, AMS Trans. 290 (1985) 815–824.
[27] L. Bianchi, Sui gruppi de sostituzioni lineari con coefficienti appartenenti a corpi quadratici immaginari, Math. Ann. 40 (1892) 332–412.
[28] R. Riley, An elliptical path from parabolic representations to hyperbolic structures, Lect. Notes Math. 722 (1979) 99–133.
[29] G. Margulis, Arithmetic properties of discrete groups, Russ. Math. Surv. 29 (1974) 49–98.
[30] S. Helgason, Differential Geometry, Lie Groups and Symmetric Spaces, Academic Press, New York, 1978.
[31] A. Besse, Einstein Manifolds, Springer, Berlin, 1987.

[32] J. Libre, R. Mackey, Pseudo-Anosov homeomorphisms on a sphere with four punctures have all periods, Math. Proc. Camb. Phil. Soc. 112 (1992) 539–549.
[33] R. Penner, An introduction to train tracks, LMS Lect. Notes Ser. 112 (1986) 77–90.
[34] B. Bowdich, Markov triples and quasi-Fuchsian groups, Proc. London Math. Soc. 77 (1998) 697–736.
[35] T. Cusick, M. Flahive, The Markov and Lagrange Spectra, AMS, Providence, RI, 1989.
[36] B. Bowdich, A variation of McShane's identity for once punctures torus bundles, Topology 36 (1997) 325–334.
[37] T. Jorgensen, On pairs of punctured tori, Columbia University, New York, unpublished.
[38] H. Cohn, Approach to Markoff's minimal forms through modular functions, Ann. Math. 61 (1955) 1–12.
[39] J. Roberts, Escaping orbits and trace maps, Physica A 228 (1996) 295–325.
[40] Y. Imayoshi, M. Taniguchi, An Introduction to Teichmüller Spaces, Springer, Berlin, 1992.
[41] R. Penner, Bounds on the least dilatations, AMS Proc. 113 (1991) 443–450.
[42] A. Haas, Diophantine approximation on hyperbolic Riemann surfaces, Acta Math. 156 (1986) 33–82.
[43] R. Alperin, W. Dicks, J. Porti, The boundary of the Giseking tree in hyperbolic three space, Topol. Appl. 93 (1999) 219–259.
[44] G. Hardy, E. Wright, An Introduction to the Theory of Numbers, Clarendon Press, Oxford, 1962.
[45] C. Series, The modular surface and continued fractions, J. London Math. Soc. 31 (1985) 69–80.
[46] Y. Minsky, The classification of punctured torus groups, Ann. Math. 149 (1999) 559–626.
[47] H. Rademacher, Higher Mathematics From an Elementary Point of View, Birkhauser, Boston, MA, 1983.
[48] H. Cohn, Mathematical microcosm of geodesics, free groups and Markoff forms, LNPAM 149 (1999) 559–626.
[49] A. Beardon, J. Lehner, M. Sheingorn, AMS Trans. 295 (1986) 635–647.
[50] R. Courant, H. Robbins, What is Mathematics?, Oxford University Press, Oxford, 1996.
[51] R. Penner, J. Harer, Combinatorics of Train Tracks, Princeton University Press, Princeton, NJ, 1992.
[52] W. Lok, Deformation of Locally Homogenous Spaces and Kleinian Groups, Ph.D. Thesis, Columbia University, New York, 1984.
[53] N. Gilbert, T. Porter, Knots and Surfaces, Oxford University Press, Oxford, 1994.
[54] E. Witten, $2+1$-dimensional gravity as exactly soluble problem, Nucl. Phys. B 311 (1988) 46–78.
[55] C. McMullen, Renormalization and 3-manifolds which Fiber Over the Circle, Princeton University Press, Princeton, NJ, 1996.
[56] A. Kholodenko, Th. Vilgis, Some geometrical and topological problems in polymer physics, Phys. Rep. 298 (1998) 251–370.
[57] D. Rolfsen, Knots and Links, Publish or Perish, Houston, MA, 1990.
[58] D. Collins, R. Grigorchuk, P. Kurchanov, H. Zieschang, Combinatorial Group Theory and Applications to Geometry, Springer, Berlin, 1998.
[59] W. Thurston, Three-dimensional Geometry and Topology, Princeton University Press, Princeton, NJ, 1997.
[60] J.-P. Otal, Le theoreme d'hyperbolization pour les varietes fibres de dimension 3, Asterisque 235 (1996) 1–159.
[61] H. Morton, Fibered knots with given Alexander polynomial, Enseignment Math. 31 (1983) 205–222.
[62] H. Matschul, On the relation between $2+1$ Einstein gravity and Chern–Simons theory, Class. Quant. Gravity 16 (1999) 2599–2609.
[63] E. Rykken, Markov partitions and the expanding factor for pseudo-Anosov homeomorphisms, Ph.D. Thesis, Northwestern University, Evanston, IL, 1993.
[64] W. Jaco, P. Shalen, Seifert fibered spaces in 3-manifolds, AMS Memoirs 21 (1979) 1–292.
[65] P. Orlik, Seifert Manifolds, Lect. Notes Math. 291 (1972) 1–155.
[66] R. Kulkarni, K. Lee, F. Raymond, Deformation spaces for Seifert manifolds, Lect. Notes Math. 1167 (1985) 180–216.
[67] G. Everest, Th. Ward, Heights of Polynomials and Entropy in Algebraic Dynamics, Springer, Berlin, 1999.
[68] P. Walters, An Introduction to Ergodic Theory, Springer, Berlin, 1982.
[69] U. Oertel, Incompressible branched surfaces, Invent. Math. 76 (1984) 385–410.
[70] W. Floyd, U. Oertel, Topology 23 (1984) 117–125.
[71] A. Kawauchi, A Survey of Knot Theory, Birkhauser, Boston, MA, 1996.
[72] M. Culler, W. Jaco, H. Rubinstein, Incompressible surfaces in once punctured torus bundles, Proc. London Math. Soc. 45 (1982) 385–419.
[73] W. Floyd, A. Hatcher, Incompressible surfaces in punctured torus bundles, Topol. Appl. 13 (1982) 263–282.
[74] W. Jaco, Lectures on 3-manifold Topology, AMS, Providence, RI, 1980.

[75] W. Neumann, Notes on geometry and 3-manifolds in Low-dimensional Topology, Journal of Bolyai Mathematical Society, Budapest, 1999.
[76] A. Hatcher, On the boundary curves of incompressible surfaces, Pacific J. Math. 99 (1982) 373–377.
[77] S. Lang, Introduction to Diophantine Approximations, Springer, Berlin, 1995.
[78] A. Kholodenko, Fermi–Bose transmutation: from semiflexible polymers to superstrings, Ann. Phys. 202 (1990) 186–225 .
[79] J. Przytycki, Incompressible surfaces in 3-manifolds, Ph.D. Thesis, Columbia University, New York, 1981.
[80] J. Przytycki, Incompressibility of surfaces after Dehn surgery, Michigan Math. J. 30 (1983) 289–308.
[81] P. Contucci, P. Kleban, A. Knauf, A fully magnetizing phase transition, J. Statist. Phys. 97 (1999) 523–539.
[82] P. Kleban, A. Qzluk, A Farey fraction spin chain, Commun. Math. Phys. 203 (1999) 635–647.
[83] R. Penner, The universal Ptolemy group and its completions, London Math. Soc. Lect. Notes Ser. 243 (1997) 293–312.
[84] B. Bowditch, A proof of McShanes identity via Markoff triples, Bull. London Math. Soc. 28 (1996) 73–78.
[85] G. McShane, A remarkable identity for lengths of curves, Ph.D. Thesis, University of Warwick, 1991.
[86] S. Katok, Fuchsian Groups, The University of Chicago Press, Chicago, 1992.
[87] D. Sullivan, Disjoint spheres, approximation by imaginary quadratic numbers, and the logarithm law for geodesics, Acta Math. 149 (1982) 215–237.
[88] S. Cosentino, Equidistribution of parabolic fixed points in the limit set of Kleinian groups, Ergod. Theory Dyn. Syst. 19 (1999) 1437–1484.
[89] D. Sullivan, Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups, Acta Math. 153 (1984) 259–277.
[90] A. Beardon, The Geometry of Discrete Groups, Springer, New York, 1983.
[91] Y. Motohashi, Spectral Theory of the Riemann Zeta Function, Cambridge University Press, Cambridge, 1997.
[92] A. Venkov, Spectral theory of automorphic functions, Selberg’s Zeta function and some problems of analytical number theory and mathematical physics, Russ. Math. Surv. 34 (1979) 69–135.
[93] W. Neumann, D. Zagier, Volumes of hyperbolic three manifolds, Topology 24 (1985) 307–332.
[94] B. Bowditch, C. McLachlan, A. Reid, Arithmetic hyperbolic surface bundles, Math. Ann. 302 (1995) 31–60.
[95] F. Grunewald, U. Hirsch, Link complements arising from arithmetic group actions, Int. J. Math. 6 (1995) 337–370.
[96] A. Reid, Arithmeticity of knot complements, J. London Math. Soc. 43 (1991) 171–184.
[97] J. Elstrodt, F. Grunewald, J. Mennicke, Groups Acting on Hyperbolic Space, Springer, Berlin, 1998.
[98] K. Huang, Statistical Physics, Wiley, New York, 1963.
[99] R. Feynman, Statistical Mechanics, Addison-Wesley, Reading, MA, 1990.
[100] L. Landau, L. Livshits, Statistical Mechanics, Nauka, Moscow, 1976.
[101] E. Hecke, Lectures on the Theory of Algebraic Numbers, Springer, Berlin, 1981.
[102] H. Cohn, Advanced Number Theory, Dover, New York, 1980.
[103] H. Weil, Algebraic Theory of Numbers, Princeton University Press, Princeton, NJ, 1940.
[104] H. McKean, V. Moll, Elliptic Curves, Cambridge University Press, Cambridge, 1999.
[105] R. Benedetti, C. Petronio, Lectures on Hyperbolic Geometry, Springer, Berlin, 1992.
[106] G. Horowitz, M. Perry, Black holes and the stability of gravitation, Gen. Relativity and Gravitation 15 (1983) 1–5.
[107] V. Kostelesky, M. Perry, No more space–time singularities, Gen. Relativity and Gravitation 26 (1994) 7–12.
[108] E. Teo, What have we learned from two-dimensional models of quantum black holes? Gen. Relativity and Gravitation 26 (1994) 13–19.
[109] A. Riley, A quadratic parabolic group, Math. Soc. Camb. Phil. Soc. 77 (1975) 281–288.
[110] C. Misner, K. Thorne, J. Wheeler, Gravitation, Freeman, San Francisco, CA, 1973.
[111] W. Neumann, A. Reid, Arithmetic of hyperbolic manifolds in Topology 90, De Gyiter, Berlin, 1992, pp. 273–310.