
ELSEVIER

Journal of Geometry and Physics 33 (2000) 23–58

JOURNAL OF GEOMETRY AND PHYSICS

Use of meanders and train tracks for description of defects and textures in liquid crystals and 2 + 1 gravity

Arkady L. Kholodenko

375 H.L. Hunter Laboratories, Clemson University, Clemson, SC 29634-1905, USA

Received 28 January 1999; received in revised form 25 April 1999

Abstract

In this work the qualitative analysis of statics and dynamics of defects and textures in liquid crystals is performed with help of meanders and train tracks. It is argued that similar analysis can be applied to 2 + 1 gravity. More rigorous mathematical justifications are presented in the companion paper (Part II) on quadratic differentials and measured foliations. Meanders were recently introduced by Arnold (Siberian J. Math. 29 (1988) 36) and are used originally in the combinatorial problem of finding the number of distinct ways given curve can intersect another curve in prescribed number of points fixed along this auxiliary curve. Train tracks were introduced by Thurston (Geometry and Topology of three-Manifolds, Princeton U. Lecture Notes, 1979) in connection with the description of homeomorphisms of two-dimensional surfaces. Train tracks alone are sufficient for the description of statics and dynamics of liquid crystals and gravity. Using train tracks the master equation is obtained which could be used alternatively to the Wheeler–DeWitt equation for 2 + 1 gravity. Since solution of this equation is possible but requires large scale numerical work, in this paper we resort to the approximation of train tracks by the meanditic labyrinths. This then allows us to analyze possible phases (and phase transitions) of gravity and liquid crystals using Peierls-like arguments.

Subj. Class.: Quantum gravity; Quantum mechanics
1991 MSC: 83C45;
Keywords: Meanders; Train tracks; Defects in liquid crystals; 2 + 1 gravity

1. Introduction

1.1. Motivations and background

The role of topology in solving physical problems is steadily increasing [1,2]. In some instances it has become difficult to decide whether physical arguments are helping to solve

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

0393-0440/00/$ – see front matter
PII: S0393-0440(99)00039-X

topological problems or topological arguments are helping to solve physical problems [2]. Here we would like to provide yet another example of this sort.

In 3 + 1 dimensions the laws of Newton's gravity and Coulombic electrostatics look suspiciously similar: both are being described by the Poisson-type equation:

$$\nabla^2\varphi = -\frac{4\pi}{\varepsilon}\rho, \tag{1.1}$$

where ε is the dielectric constant in the case of electrostatics while $\varepsilon^{-1} = G$ is the gravitational constant in the case of gravity. The density of charges $\rho(\mathbf{r})$ can be both positive and negative for electrostatics while only positive in the case of gravity. The solution of Eq. (1.1) for the potential $\varphi(\mathbf{r})$ is given by

$$\varphi(\mathbf{r}) = \frac{1}{\varepsilon}\int_V \mathrm{d}^3 r \frac{\rho(\mathbf{r}')}{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}}, \tag{1.2}$$

where $\mathbf{r} = \{x, y, z\}$ and V is the volume which encloses the charges. It is expected, that if the charges are confined within the domain V, then $\varphi(\mathbf{r}) \to 0$ when $|\mathbf{r}| \to \infty$. Indeed, let $\rho(\mathbf{r}) = q\delta(\mathbf{r})$, then we obtain trivially

$$\varphi(\mathbf{r}) = \frac{q}{\varepsilon}\frac{1}{|\mathbf{r}|}, \tag{1.3}$$

where q is the magnitude of charge placed at the origin. Solution given by Eq. (1.2) is valid only in three-dimensions, however. In two dimensions it should be replaced by

$$\varphi(\mathbf{r}) = \frac{1}{\varepsilon}\int_A \mathrm{d}^2 r' \rho(\mathbf{r}') \ln\left(\frac{1}{\sqrt{(x-x')^2+(y-y')^2}}\right), \tag{1.4}$$

(with volume V being replaced by area A) so that, instead of (1.3), we obtain now

$$\varphi(\mathbf{r}) = -\frac{q}{\varepsilon}\ln|\mathbf{r}|. \tag{1.5}$$

This time, for $|\mathbf{r}| \to \infty$ we obtain $\varphi(\mathbf{r}) \to \pm\infty$ (depending upon the sign of q). The result (1.5) can be obtained *directly* from Eq. (1.2) for a special case of charges uniformly distributed along the infinitely long and infinitely thin rod placed perpendicular to the x–y plane, i.e.

$$-\frac{1}{2}\ln\left(\sqrt{(x-x')^2+(y-y')^2}\right) = \lim_{L\to\infty}\left(\frac{1}{2}\int_{-L}^{L}\mathrm{d}z'\sqrt{\frac{1}{(x-x')^2+(y-y')^2+(z-z')^2}} - \ln 2\right). \tag{1.6}$$

Instead of a rod we can imagine a particle "living" in 2 + 1 dimensions (i.e. particle which evolves in time while moving in x–y plane).

The difference of behaviors of potentials at infinity in two and three dimensions leads to the profound difference in the underlying physics. Moreover, this difference forbids smooth dimensional continuation of the expression for the potential, given in d-dimensions by

$$\varphi_d(\mathbf{r}) = \frac{q}{\varepsilon}\frac{1}{|\mathbf{r}|^{d-2}}, \tag{1.7}$$

sometimes used in statistical mechanics in connection with the renormalization group methods. Let us explain why this continuation is illegitimate. Since the potential φ is an auxiliary quantity in both gravity and electrostatics, we shall use instead the forces $\mathbf{F}$ obtained in a usual way through

$$\mathbf{F} = -\vec{\nabla}\varphi. \tag{1.8}$$

The forces create the vector fields, and in the case of two-dimensions, we have to consider the vector fields on surfaces. In particular, let us consider the plane $\mathbf{R}^2$, the sphere $S^2 = \mathbf{R}^2 \cup \{\infty\}$ and the disc D^2 of some radius R. Evidently, physically all three possibilities are almost indistinguishable: the sphere S^2 can be obtained from the plane by one point compactification, the plane can be obtained from the disc D^2 by taking the limit $R \to \infty$. Topologically, however, the above tree surfaces are completely different: the sphere S^2 has the Euler characteristic $\chi(S^2) = 2$, the disc $\chi(D^2) = 1$ and the plane $\mathbf{R}^2$ cannot be given any value of χ [3]. For the time being we shall work with S^2 and D^2. The choice between these two surfaces is determined by the boundary conditions.

On the surface (manifold M) vector fields obey the Poincare–Hopf (P–H) index theorem [4]:

$$\chi(M) = \sum_i I(x_i), \tag{1.9}$$

where the index $I(x_i)$ of isolated singularity is defined as follows. Consider the vector field $\mathbf{v}(x, y) = \{\mathrm{v}_x(x, y), \mathrm{v}_y(x, y)\}$ around just one of the singularities and let $U \in M$ be some (say, circular) domain around this singularity. In this domain we can construct a unit vector $\mathbf{n}(x, y)$ via

$$\mathbf{n}(x, y) = \frac{\mathbf{v}(x, y)}{|\mathbf{v}(x, y)|}. \tag{1.10}$$

This vector provides the Gauss map from M to S^1. The degree of this mapping is $I(x_i)$. More explicitly [5,6],

$$I(x_i) = \frac{1}{2\pi}\oint_{C_i}\frac{\mathrm{v}_x\,\mathrm{dv}_y - \mathrm{v}_y\,\mathrm{dv}_x}{\mathrm{v}_x^2 + \mathrm{v}_y^2}. \tag{1.11}$$

In cases of both Newtonian gravity and Coulombic electrostatics the index of an isolated singularity is equal to one as can be easily seen by introducing the complex variable $z = x + y = re^{i\varphi}$ and by combining Eqs. (1.5), (1.8) and (1.11). This means that the *minimal* number of charges which one can put on the sphere is 2 while that on the disc is 1. The P–H theorem, Eq. (1.9), does not automatically leads to the requirement of electroneutrality: one can place two charges of the same sign on S^2 without violating this theorem. The complications arise, however, if one would like to place more than two charges on the sphere. In this case if initially we would have, say, two pluses, we would be unable to place

the additional charges while if we would have initially "+" and "−" charges, then such placement becomes possible *if we add charges pairwise* while keeping the electroneutrality. This becomes possible only because the presence of two identical charges creates a saddle (e.g. see Fig. 1(a)) which has an index −1 [4]. Hence, in the case of electrostatics the only possible singularities are the sources/sinks (Fig. 1(d))with index 1 and the *induced* saddles (Fig. 1(a)) with index −1. The existence of multiple charges in electrostatics and the electroneutrality are directly related to the topology of the underlying manifold and to the emergence of the induced saddles.

Consider now the case of Newton's gravity. Let us place the same two charges (masses) on the sphere and let, say, one of the charges be "+" so that another is "−", and since this is gravity, the masses are naturally attracting each other. Let us try now to add one additional mass (charge). We immediately run into problem: since all masses attract each other, there cannot be saddles. If this is the case, we are unable to place an additional mass on S^2. Hence, in such spherical universe we would have just two masses! The situation becomes even more dramatic if, instead of S^2, we would consider D^2. In this case we would not be able to put more than one mass while for the case of an annulus (i.e. the disc with a hole) we would not be able to put on it even a single mass (!) since the Euler characteristic of an annulus is 0, [3]. At the same time, we can easily put e.g. three charges of the same sign on the annulus in the case of electrostatics. These very simple arguments lead us to the conclusion that there are some profound differences between the Coulombic electrostatics and the Newtonian gravity and that the above differences alone are sufficient in order to arrive at the correct Einstein formulation of gravity. This is going to be demonstrated mainly in the accompanying paper [7] while here we provide only the qualitative arguments. For this purpose we also need to discuss systems other than electrostatic and gravitational (in Newton's sense). These are naturally occurring as defects and textures in liquid crystals [8,9].

To study these defects it is helpful to recall some basic facts from the qualitative theory of dynamical systems on surfaces [5,6]. These are also described in terms of flows of the vector fields. In two dimensions we are usually dealing with the system of two equations:

$$\begin{aligned} \frac{dx}{dt} &= v_x(x, y), \\ \frac{dy}{dt} &= v_y(x, y). \end{aligned} \tag{1.12}$$

Irrespective to the explicit form of the r.h.s. of Eq. (1.12), it is known [5,6] that the singularities of the flow could only be of the type depicted on Fig. 1.[1] These can be easily obtained by linearization of the Eq. (1.12) around the singularities of the vector field. The whole phase portrait can then be built out of these local pictures by gluing together the local pieces in some consistent way. Moreover, the topological considerations allow us to restore the phase portrait with help of only *partial* knowledge of the existing flow around singularities. This principle is very helpful for our case too but the situation is complicated by the fact that, in the case of liquid crystals as well as in 2 + 1 gravity, instead of *vector*

[1] Clearly, the flow directions on some of the figures could be reversed.

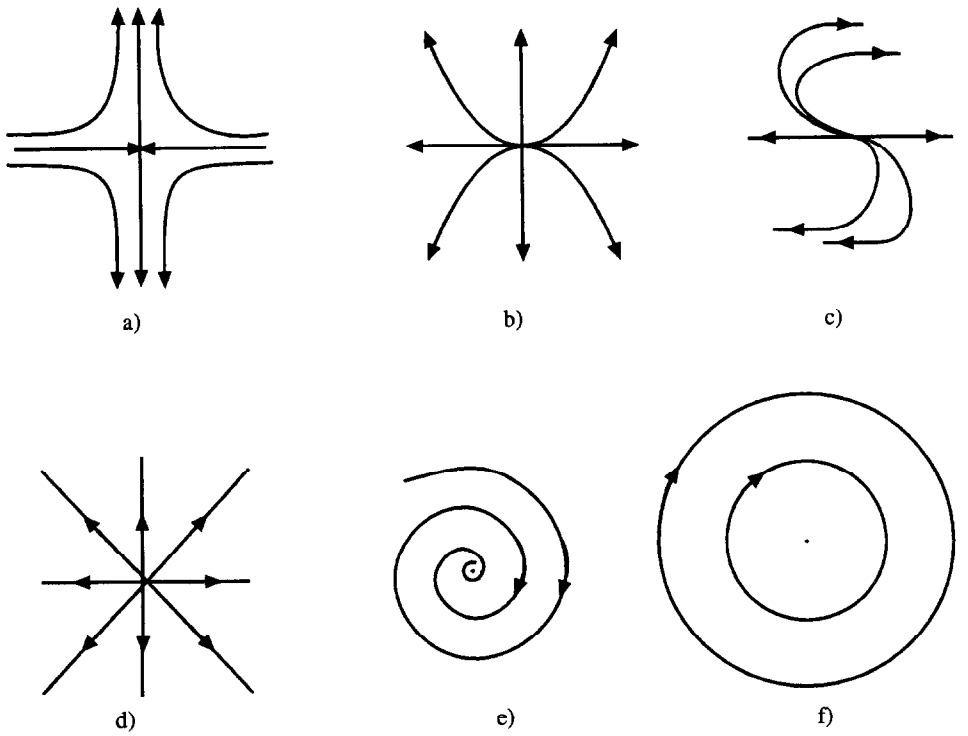


Fig. 1. Typical singularities of the vector fields in the plane.

fields on surfaces one has to deal with *line* fields. This was demonstrated for liquid crystals in [10,11] and will be demonstrated in [7] for the case of gravity. The case of line fields has been also studied in the theory of differential equations on surfaces [5,6]. In this case, instead of the system of Eqs. (1.12), we have to consider

$$\frac{\mathrm{d}x}{\mathrm{v}_x(x,y)} = \frac{\mathrm{d}y}{\mathrm{v}_y(x,y)}. \tag{1.13}$$

Typical singularities of the line fields are depicted in Fig. 2 with the values of the corresponding indices.[2] For example, in the case of the line fields with indices $\pm 1/2$, the differential equation which describes these fields is given (in polar r, φ coordinates) by [5]

$$\frac{\mathrm{d}r}{\mathrm{d}\varphi} = r\tan\left(n\frac{\varphi}{2}\right). \tag{1.14}$$

For $n = 1$ we obtain the line field with index $-1/2$. This field is described analytically by

$$r = \frac{a}{\left(\cos\frac{3\varphi}{2}\right)^{2/3}}$$

[2] Eq. (1.11) cannot be used for the calculation of indices of line fields. The analogous formula is provided in [7], e.g. see Eq. (4.20).

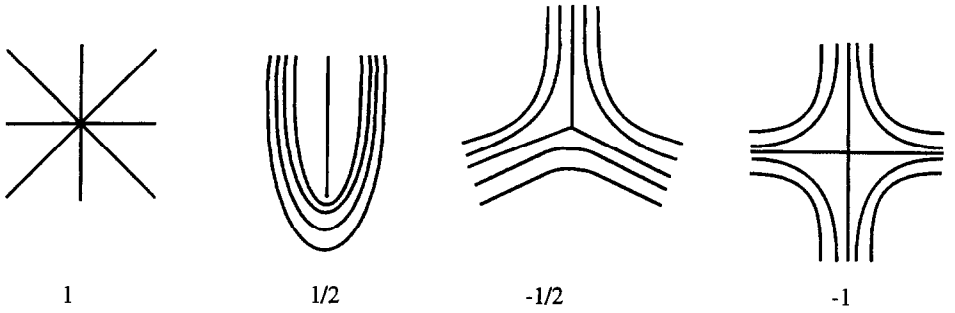


Fig. 2. Typical singularities of the line field and the associated indices.

with a being an arbitrary positive constant and φ lying in the following sectors:
(a) $-\pi/3 < \varphi < \pi/3$,
(b) $\pi/3 < \varphi < \pi$,
(c) $\pi < \varphi < 5\pi/3$.

Similar analysis can be performed for $n = -1$ which produces the index $+1/2$ as depicted in Fig. 2. Although for the line fields it is impossible to introduce consistently the global orientation [4–6], nevertheless, it is possible to introduce the indices for these fields [4] by analogy with the vector fields. The absence of orientation for the line fields leads to the absence of "forces" between the singularities (defects). This property is not unusual for the theories of $2+1$ gravity where it is impossible to introduce the interaction forces between massive particles [12]. In Sections 4, and 5 of Part II [7] we shall demonstrate that both $2+1$ gravity and the textures in liquid crystals are described by the *line* fields. In this part (which we call Part I) we shall rely mostly on the intuitive arguments.

Although it is impossible to introduce the interaction forces between defects (or masses) it is possible, nevertheless, to talk about the total energy in both $2+1$ gravity [13] and in the theory of liquid crystals (e.g. see Part II and Section 5 below). Moreover, because of this fact, it makes also sense to talk about the different phases (orders) at least for the case of defects in liquid crystals [14]. Although the connections between the defects and textures in liquid crystals and $2+1$ gravity were discussed before [15], here we provide completely different treatment of these connections.

1.2. Organization of the rest of this paper

In Section 2 we consider the role of topology in phase transitions in two-dimensions. We argue, that in cases of the vector (Coulomb-like) and line fields on surfaces topological considerations alone are sufficient for predicting the Kosterlitz–Thouless type of phase transition [14] from gas to dipole phase while in case of the line fields in liquid crystals, the existence of hexatic phase [16] can be easily established. Clearly, topological considerations alone provide only sufficient conditions for existence of ordered phases. The necessary conditions require us to study the nature of the disordered phase in some detail. This is accomplished in Sections 3 and 4. In Section 3 we provide an explicit construction of

the disordered phase using some ideas from the qualitative theory of ordinary differential equations. Superposition of these ideas with topological arguments produces notions of "labyrinths" and "meanders". Since neither labyrinths nor meanders are widely known in mathematical physics literature, to avoid repetitions, we refer the reader directly to Section 3 for precise definitions. Here, we only would like to mention that the description of labyrinths is closely related to the description of the maps of the circle [17,18] which should be more familiar to physically trained reader. The notion of a meander has its origin in the work by Arnold [19] who actually invented this word. In the simplest case one can think about the meander as a representative of the class of curves which intersect another curve in the prescribed number of points. Although from the above vague definition it may not be clear that the meanders and the labyrinths may have many things in common, nevertheless, this is the case as we demonstrate in Section 3.

In Section 4 we discuss yet another way of looking at the line fields through the notion of "train tracks". This way of looking at line fields was proposed originally by Thurston [20] in connection with his studies of three-manifolds. We argue in this section that the train tracks can be used for description of dynamics of textures in liquid crystals and $2+1$ gravity. The master equation which describes the evolution of the train tracks could be used as an alternative to more familiar Wheeler–DeWitt equation (in the case of gravity). Since the actual calculations which involve this equation resemble that of the Heisenberg-type quantum mechanics, they require a large scale numerical work which we try to avoid (at this stage of research) by restricting ourself to the approximation of train tracks by meandritic labyrinths. The legitimacy of this approximation is discussed in Appendix A.

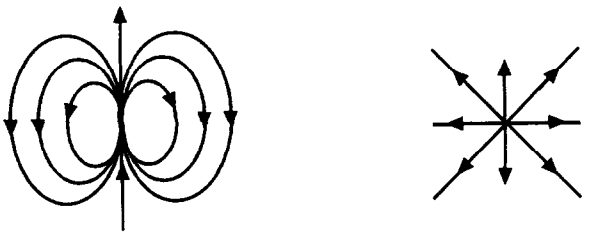
Use of the meandritic labyrinth approximation allows us to obtain additional quantitative information about the stability of the disordered phase and the parameters of the order–disorder phase transition in the system of meandritic labyrinths. This is accomplished in Section 5.

All results of Part I are provided without serious proofs (with few exceptions). Part II [7] serves to some extent to correct this deficiency by providing the mathematical justifications to the emerging picture. Sufficient details (and references) are provided in both parts of this work to make this presentation self-contained and accessible not only to the experts on gravity but to the interested condensed matter researchers as well.

2. Role of topology in phase transitions in two-dimensions

We had already demonstrated (in Section 1.1) that the property of electroneutrality is directly connected with the topology of the underlying manifold, at least in the case of Coulombic-like systems. Here, we would like to demonstrate that, in addition, the topological considerations *alone* determine the nature of phase transitions in such systems. For the line fields the sequence of arguments, unfortunately, is considerably more complicated and will be discussed in the rest of this paper and in Part II.

In the case of Coulombic-like system, let us assume that under some favorable conditions the incoming "+" and "−" have "decided" to stay together thus forming a dipole. The vector



a) b)

Fig. 3. The field lines around the dipole (a) and the source (b) (in the case of a sink the directions of lines are changed into the opposite).

field in this case is depicted in Fig. 3 along with a separate source (+) (or sink (−)). The index of the dipole is 2 so that P–H theorem is not violated if on the sphere we would have only two charges which have "decided" to form a dipole. Let us place a source/sink in the vicinity of such dipole. By direct inspection we can notice that topologically such source/sink cannot fit the vector field flow pattern coming from the dipole unless it is combined with yet another charge thus forming yet another dipole. Formation of the second dipole will cause the formation of two-additional saddles so that P–H theorem is not violated. Evidently, this procedure can be extended to other charges. From here several conclusions could be drawn. First, the transition to the dipole phase should be sharp. Second, since in the above arguments the density of charges ρ was not present, it means that such transition cannot be considered in traditional thermodynamical sense, that is for some critical temperature T_c and critical pressure P_c there is no true critical density ρ_c. These conclusions are in complete accord with the results of Hague and Hemmer [21] who had provided more traditional statistical mechanics treatment of this type of transition. Obtained results are also in complete accord with the results of more sophisticated treatment performed by Kosterlitz and Thouless [22].

The above simple picture breaks down immediately if we are willing to analyze possible transitions in the case of defects which produce the line fields, e.g. liquid crystals or $2+1$ gravity. Unlike the Coulombic case where the vector field picture is rather simple: or we have an independent charges or we have dipoles, in the line fields case there are many more possibilities. Since the local and global analysis of the line fields had been (and still is!) the subject of intensive research in mathematics, it is impossible to squeeze the enormous amount of results accumulated to date in this work. For a good summary, please, consult [23–25]. Without trying to present a summary even of these results, we shall, nevertheless, select those which, we feel, are of immediate physical relevance.

Let us begin with the observation that in case of the Coulombic-like fields the regrouping of charges had not produced a violation of the P–H theorem. Evidently, we have to require the same in the case of line fields. The moves which do not violate P–H theorem are depicted in Fig. 4 and are known in the literature as Whitehead moves [23]. In addition, it is possible to imagine the situations depicted in Fig. 5. Clearly, by looking at Fig. 4 we observe the case of *creation* of "new" defects with index ± 1 from the "old" ones with index $\pm 1/2$, while looking at Fig. 5 we effectively observe the case of *destruction* of defects. The indices of

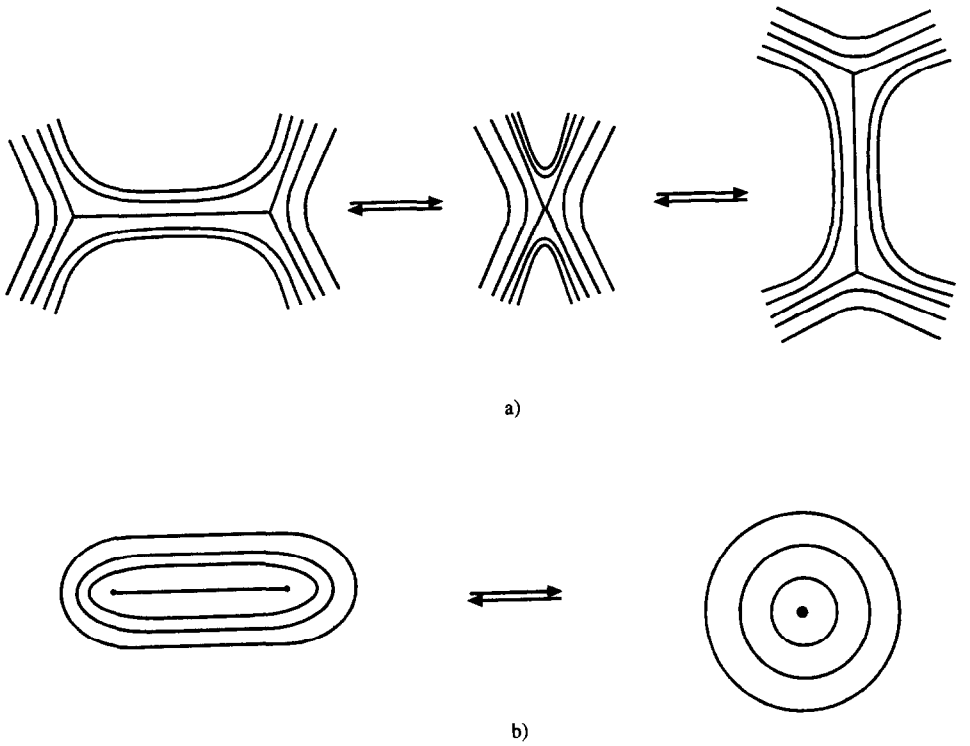


Fig. 4. The allowed Whitehead moves for the line fields ("creation").

(a) and (b) are $+1/2$ while the index of (c) is zero. In principle, other situations are also possible but, for reasons which will be explained in Section 4 and Part II, we shall not be concerned here with more complicated situations.

Given these *local* moves, what one can say about the "phase transitions" (as compared to the Coulombic-like vector fields) in the system of such defects? For example, can we expect, by analogy with Coulombic case, that the transition(s) is (are) going to be density independent? For this purpose let us consider the case of a disc $\chi(D^2) = 1$. To be in accord with the P–H theorem, Eq. (1.9), we need to have at least four defects with half integer indices (since only half integer types of defects are topologically stable [26]): one with index $-1/2$ and 3 with the index $+1/2$. The resulting "stable phase" is depicted in Fig. 6. This does not look like a dipole, and because of the P–H theorem, we cannot just form another dipole and so on as in the Coulombic case and we cannot use the electroneutrality condition as well. So, what else could we possibly imagine? For example, should we have *exactly* 14 defects, we could have an orderly structure which resembles that depicted in Fig. 7. We use the word "periodic" to describe this phase following Thurston's classification of surface homeomorphisms [25] as will be explained in more details in Section 4.

Is it possible to obtain something else? The answer depends upon how many defects are at our disposal. Consider the case of thermodynamic limit, i.e. the case when the size R of our disc D^2 is allowed to approach infinity along with the number of defects so that their ratio

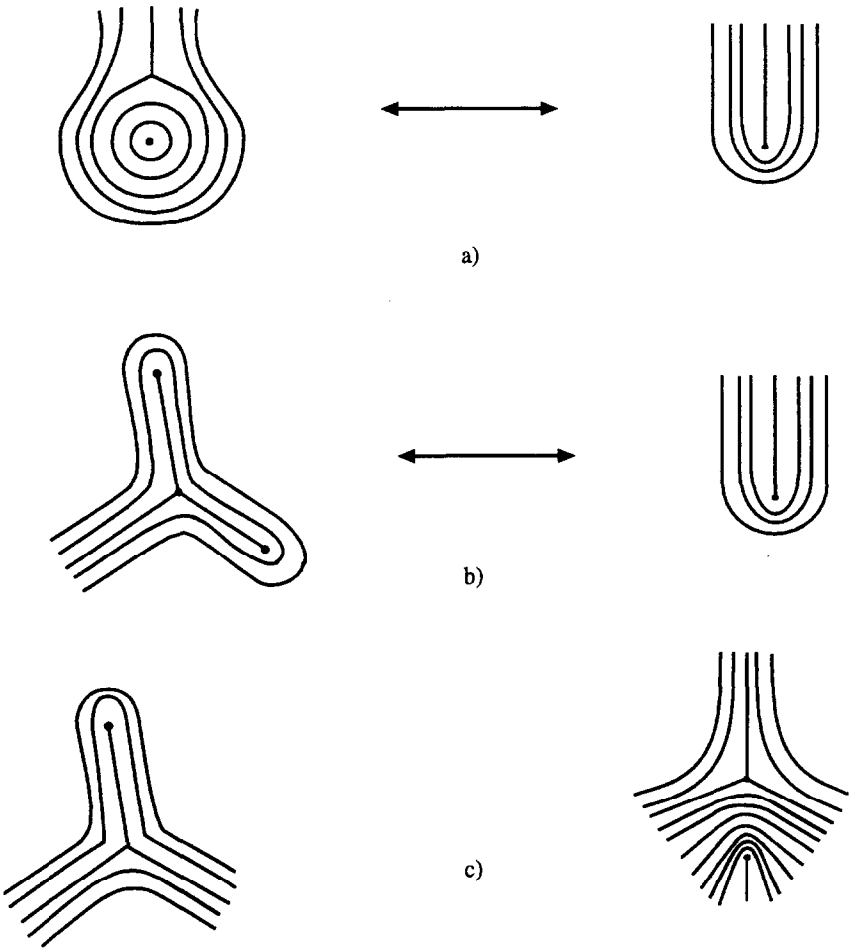


Fig. 5. The additional configurations leading to "destruction" of defects.

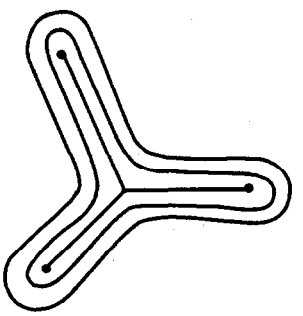


Fig. 6. Possible "periodic" phase in the case of minimal number of line field defects in the disc

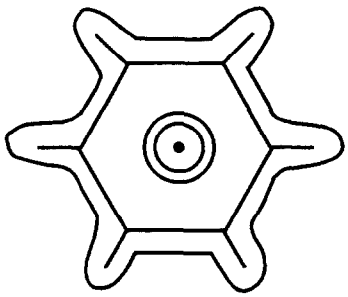


Fig. 7. Possible "periodic" phase made up of 14 defects in the disc.

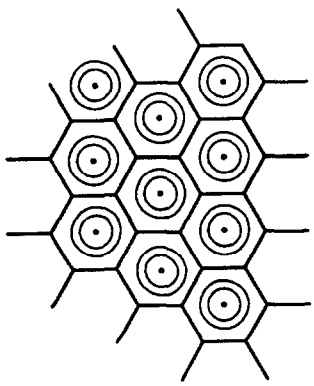


Fig. 8. Possible hexatic-type phase obtained in the thermodynamic limit.

remains fixed. In this situation we can imagine the structure depicted in Fig. 8. How can we check that this structure is actually in agreement with the P–H theorem? This can be accomplished following the original arguments by Hopf [4] (see also [27]). Let S be closed orientable surface of genus g. Let a_0 be the number of vertices, a_1 be the number of edges and a_2 be the number of two-cells (e.g. hexagons). Then, according to the Euler theorem, we obtain,

$$a_0 - a_1 + a_2 = 2 - 2g. \tag{2.1}$$

Looking at Fig. 8 and following the same line of arguments as were made by Hopf, let us replace the original hexagonal lattice with the elementary cell depicted in Fig. 9.

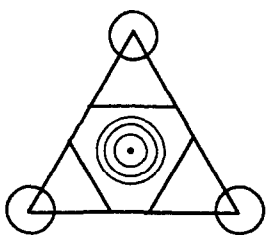


Fig. 9. The elementary triangle used in P–H theorem calculations related to the hexatic-type phase depicted in Fig. 8.

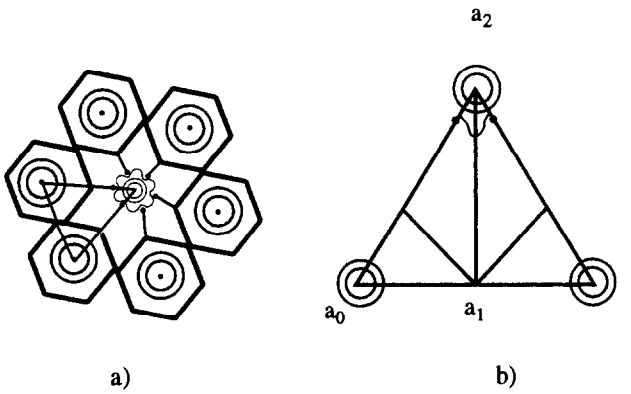


Fig. 10. (a) Another possible hexatic phase which is permitted by the Euler theorem; (b) The elementary triangle used in Hopf-like calculations of the Euler characteristic.

The face of each cell has an index $+1$, the edge has an index $2(-1/2) = -1$ and each vertex has an index 1. Hence, such construction holds for an arbitrary g, and therefore, the hexagonal phase is consistent with the Euler theorem, Eq. (1.2), and, accordingly, with the P–H theorem, Eq. (1.9). As in the case of the Kosterlitz and Thouless transition earlier described, the existence of the above hexatic phase is already known in physics literature, e.g. see [28], where this phase is obtained with the help of completely different set of arguments. In Section 4 of Part II we shall provide yet another arguments in support of the existence of this phase.

The questions arise:

(a) is such obtained hexatic phase unique?

(b) can such phase occur for an arbitrary concentration of defects (as in the Coulombic phase case)?

The answer to the first question is negative as can be easily seen from the situation depicted in Fig. 10(a) and the corresponding elementary triangle given in Fig. 10(b). Taking into account that, in view of Fig. 5(c), the effective total index of singularities at the edge (a_0a_2) is zero, the present case coincides with that described by Hopf [4]. That is, we have the index of two-cell the same as a_2 vertex (i.e. $+1$), the index of the edge the same as the index at a_1 (i.e. -1) and the index of the vertex is the same as the index at a_0 (i.e. $+1$). Hence, we are again in accord with the Euler theorem, Eq. (2.1).

Since we have at least two types of hexagonal structures we may think about some sort of phase transition between them. In addition, we have not actually proved that only two hexagonal structures are possible. A similar but much simpler problem of packing of hard hexagons on the triangular lattice was considered by Baxter [29] (and, more recently, by Monasson and Pouliquen [30]) with partial success (since the solution of the hard hexagon model is related to the solution of $2d$ Ising model in the magnetic field which is not known in general).

Because of some similarities between the hard hexagons and the present liquid crystal problem, it is useful to furnish some details since they will eventually be used to provide an

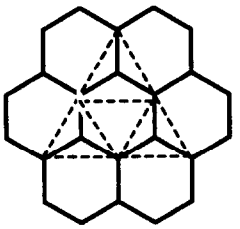


Fig. 11. Two-dimensional plane can be covered by hexagons. At the same time, one can cover the plane by the triangular lattice so that for the complete coverage three hexagons should fit together around each vertex.

answer to the second question (about the concentration dependence) posed above. To this purpose, in the case of hard hexagons, let us introduce the surface density ρ via

$$\rho = \frac{n}{N}, \tag{2.2}$$

(a similar problem which involves the one component plasma was considered recently in [31,32]), where n is the total number of hexagons and N is the total number of lattice sites (triangles). More useful quantity, however, is the packing fraction $\eta = 3\rho$. To understand the emerging factor of 3, the inspection of Fig. 11 is helpful (see also [33]). By construction, the complete packing corresponds to $\eta = 1$. Looking at Fig. 10, we easily can conclude that, in the case of the packing depicted in Fig. 10(a), the packing fraction is 1/2. Even in the case of hard hexagons the packing problem is not completely under control. That is the description of transition from one mode of partial packing to another (e.g. loose $\eta < 1$ versus dense $\eta = 1$) is still lacking [29,30]. At the same time, in the present case we are not even dealing with *hard* hexagons. Our "hexagons" can be easily destroyed as it will be discussed in Section 6 of Part II. To analyze this more complicated situation, quantities other than ρ may be helpful. Following Ref. [31], let us introduce some cut-off radius r_0 (the size of the disclination core [14,26]) as well as the surface density $\hat{\rho}$ via

$$\hat{\rho} = \frac{\hat{n}}{A}. \tag{2.3}$$

where A is the surface area and $\hat{n}$ is the total number of surface defects (charges). Next, we can introduce the Wigner-Seitz radius r_{w-s} via

$$\pi r_{w-s}^2 = \frac{1}{\hat{\rho}}. \tag{2.4}$$

Finally, the filling fraction ν can now be introduced via

$$\nu = \left(\frac{r_0}{r_{w-s}}\right)^2 = \pi r_0^2 \hat{\rho}. \tag{2.5}$$

By construction, $0 \leq \nu \leq 1$ so that ν can be used (instead of η) to characterize the possible phase transitions. These results will be used in Section 5.

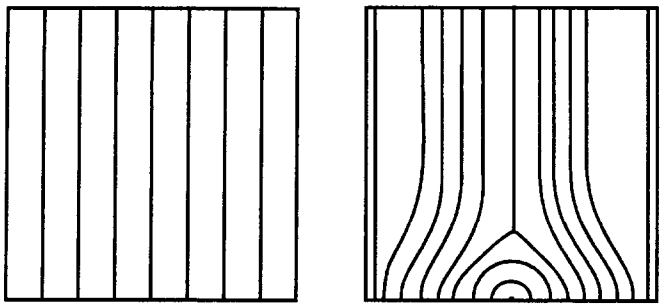


Fig. 12. Simplest "building blocks" for measured foliations.

3. Labyrinths and meanders

The disordered phase of line field defects is no less interesting than the ordered phase. Let us consider the simplest possible case of just four defects in D^2. The "ordered" phase is depicted in Fig. 6. To get a feeling of the disordered phase, following [34], let us consider the "foliation box" $\mathcal{F}_0$ depicted in Fig. 12(a). The lines, known in mathematical literature as *foliations* (measured foliations to be exact [10,23–25]) are free of singularities in the box a) and contain two singularities in the box (b) (with account of the results depicted in Fig. 5(c))). Following [34], we call these singularities "Y" and "thorn", respectively. Since the total index in the box b) is zero, we can, evidently, place in the line field as many as we wish such Y-thorn doubles without violating of the P–H theorem, Eq. (1.9). This gives us a certain freedom of moving such objects around. In particular, if for the moment we would like to forget about the "tail " of Y-singularity, then, we can concentrate our attention at the foliations in semicircle(s). We shall restore the tail afterwards (to account for the P–H theorem). In the meantime, let us consider collection of three foliated semicircles as depicted in Fig. 13.

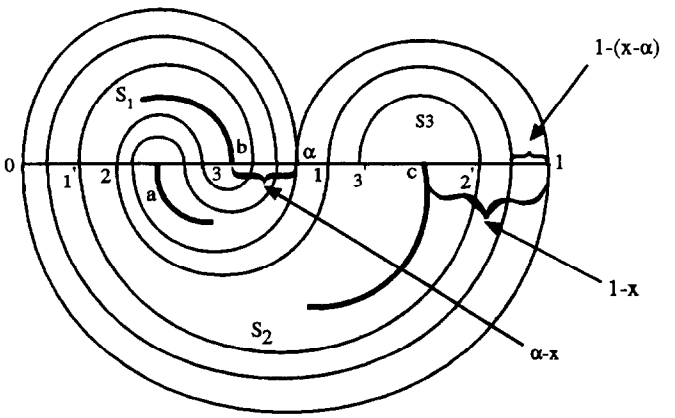


Fig. 13. Simplest nontrivial global foliation pattern made of simple building blocks.

The foliations for S_1 are centered at a, for S_2 at b, and for S_3 at c, respectively. The ratio of diameters $D_{S_3}/D_{S_1} = \alpha$ is chosen to be some irrational number. Let x denote the location of an intersection of a given foliation leave with the axis [0, 1]. Then, depending upon the direction from which the foliation hits the x-axis, one can distinguish between the following three possibilities: (1) if the foliation hits [0, 1] by coming from S_1 (e.g. see point 2), then let $R_1(x) = \alpha - x, 0 \leq x \leq \alpha$; (2) if the foliation hits [0, 1] by coming from S_2 (e.g. see point c), then let $R_2(x) = 1 - x, 0 \leq x \leq 1$; (3) if the foliation hits [0, 1] by coming from S_3 (e.g. see point 2'), then let $R_3(x) = 1 - (x - \alpha), \alpha \leq x \leq 1$. The meaning of thus introduced functions R_1, R_2, and R_3 becomes clear if we consider their compositions, e.g. $[R_1 \circ R_2](x)$ and $[R_3 \circ R_2](x)$. Looking at Fig. 13, we may notice, that in both cases we are dealing with the second return map of the closed interval [0.1], e.g. see the paths 1–2–3 (for $R_1 \circ R_2$) and 1'–2'–3' (for $R_3 \circ R_2$). Explicitly, we obtain

$$[R_1 \circ R_2](x) = \alpha - (1 - x), \tag{3.1}$$

$$[R_3 \circ R_2](x) = \alpha + 1 - (1 - x). \tag{3.2}$$

If now we identify the ends of the interval, i.e. 0 and 1, then both Eq. (3.1) and (3.2) are *translations* by α along the circle. The theory of circle homeomorphisms is well documented [17,18], and for the sake of continuity, we would like to remind to our readers several important facts which will help us to clarify the results just obtained and those which will follow.

Let us notice that the circle S^1 is the quotient: $S^1 = \mathbf{R}/\mathbf{Z}$, i.e. the circle is the factor group of real numbers $\mathbf{R}$ modulo integers $\mathbf{Z}$. Let $f : S^1 \to S^1$ be some circle homeomorphism and let $F : \mathbf{R} \to \mathbf{R}$ be a lift of f (e.g. $f(S^1) = \exp\{2\pi i x\}$ and $F : x' = x + \alpha$, that is in the simplest case F are just translations in $\mathbf{R}$). The rotation number $\tau(F)$ can be defined via equation:

$$\tau(F) = \lim_{|n|\to\infty} \frac{1}{n}(F^n(x) - x). \tag{3.3}$$

In general, $F(x)$ can have the form given by

$$F(x) = x + a(x), \qquad a(x + 1) = a(x), \qquad 0 \leq x < 1. \tag{3.4}$$

In terms of $a(x)$ the rotation number can be rewritten as

$$\tau(F) = \lim_{|n|\to\infty} \frac{1}{n}(a(x) + a(F(x)) + \cdots + a(F^{n-1}(x))) \equiv \lim_{|n|\to\infty} \frac{1}{n} a_n(x). \tag{3.5}$$

Let the mapping F has a fixed point x^* for some q:

$$F^q(x^*) = x^* + a_q(x^*). \tag{3.6}$$

It can be shown that,

$$a_q(x^*) = p, \tag{3.7}$$

and

$$\tau(F) = \frac{p}{q}. \tag{3.8}$$

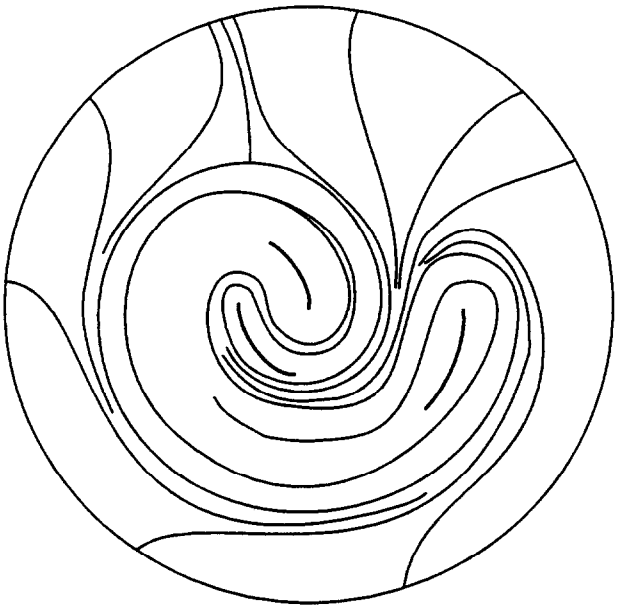


Fig. 14. The structure of disordered phase (labyrinth) in the case of minimal number of line defects (to be compared with the ordered phase depicted in Fig. 6).

That is q translations in the lifting space are equivalent to "pure" rotation of the circle by the angle $\vartheta = 2\pi p/q$. Accordingly, if Eq. (3.6) does not have fixed point(s) for any q, then $\tau(F) = \alpha$ where α is irrational. With respect to the situation depicted in Fig. 13 we have to conclude, that the "motion" along the line 1–2–3, etc. will never stop or cross another line, i.e. we have obtained the foliation which is actually a labyrinth. Let us take now into account the P–H theorem and make the situation more realistic. To this purpose we attach the "tail" to S^1 semicircle and gently open the foliation leaf starting at α in Fig. 13. The result of such opening procedure is depicted in Fig. 14 which represents the simplest example of a labyrinth: all incoming leaves are being trapped inside the labyrinth forever.

The picture just described can be wasty generalized with help of the following auxiliary observation. Consider some nonintersecting line $\mathcal{L}$ (closed or not) and place on the top of it a finite collection of foliated half circles (semicircles) provided that none of these half circles are touching each other (unless the otherwise is specified). Let us place the remaining half discs on the bottom of $\mathcal{L}$ in the way depicted in Fig. 15. We have obtained in this way an example of a "meander" labyrinth. Let us explain now what actually the word "meander" means. According to [35], a meander of order n is a closed nonselfintersecting curve which intersects another straight line in exactly $2n$ preassigned points (more exactly, an equivalence class of such closed curves which leave the straight line $\mathcal{L}$ fixed). Evidently, the line $\mathcal{L}$ need not be straight in general. A few examples of the meanders of lower order and the corresponding meandritic labyrinths are depicted in Fig. 16. Although the four last meandritic labyrinths are degenerate (two half discs are touching each other), this may be still permissible (if other discs are not touching each other), e.g. see Example 2 of Ref. [34].

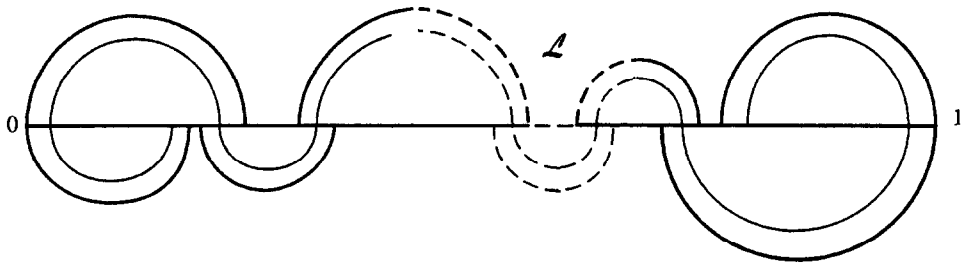


Fig. 15. The simplest example of the meander labyrinth.

This is so because we still have to attach "tails" to some of the discs (e.g. see Fig. 14) to make them Y-type in order to be in accord with P–H theorem. Trough the tails the external lines penetrate into the labyrinth and become trapped forever. Moreover, because of the tails, the symmetry between different labyrinths depicted in Fig. 16 becomes broken so they all have to be considered, just like the underlying meanders.

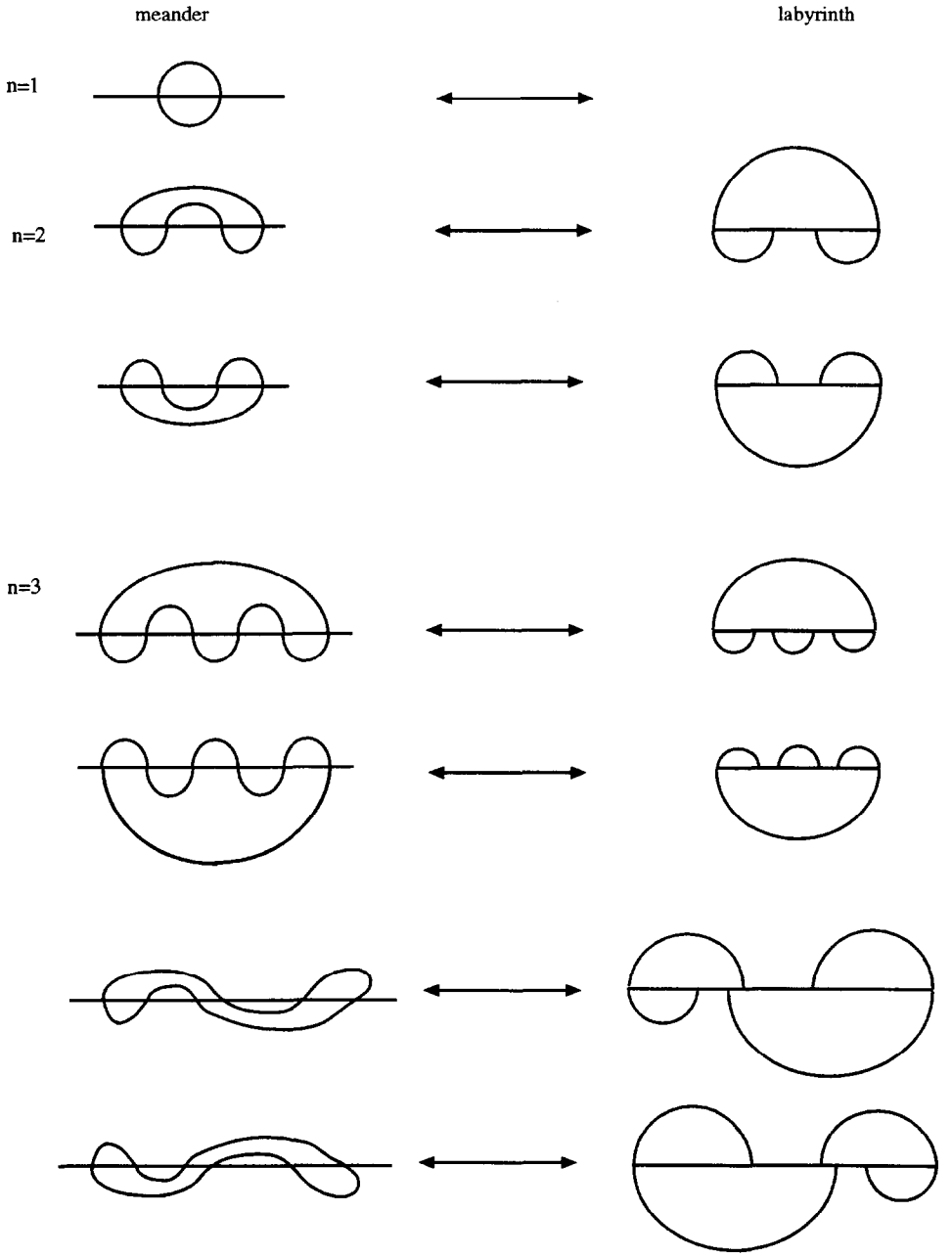
The meandritic number M_n (that is the total number of meanders of order n) provides the degeneracy factor for the partition function of defects to be discussed in Section 5. Its value is essential in the description of the diclination binding–unbinding "melting" transition [14] in the non-Coulombic two-dimensional "plasma" of such defects. Actually, the above description is still incomplete since, as it is well known from statistical mechanics of two-dimensional plasma [21], the clustering effects should also be taken into account. This, in particular, means that, instead of just one connected meander, we may as well consider a collection of nonintersecting meanders. Specifically, for the multicomponent meander of order n we may choose k nonintersecting lines ($1 \leq k \leq n$) so that the total number of crossings on all lines is $2n$. Accordingly, one can introduce the meandritic number $M_n^{(k)}$ to account for this fact [36,37]. This number will be estimated and used in Section 5. In the meantime, it is useful to introduce, yet another, set of meanders, the projective meanders [19,38]. These are defined as follows. Consider $2n$ points on the circle S^1. Divide the points into n pairs so that the chords connecting points in each pair will not intersect. Identify the diametrically opposite points on the circle, thus turning the disc into the projective plane (or sphere). Such formed set of curves in the disc is called the projective meander of order n. Evidently, the number of possible projective meanders of order n is equal to the nth Catalan number

$$C_n = \frac{(2n)!}{(n+1)!n!}, \tag{3.9}$$

which obviously reflects the combinatorics of the problem. We shall need the notion of the projective meanders in order to demonstrate that the above labyrinth construction is not artificial but, actually, is intrinsic for the whole description of the measured foliations on surfaces. In addition, there is yet another way to look at the whole problem for the case of line fields. Because of its potential usefulness for the problems which involve $2+1$ gravity, we discuss this other approach in the next section.

4. Train tracks and pseudo-Anosov homeomorphisms

The logical development of dynamical systems on surfaces goes from the detailed treatment of the circle maps through the torus maps and then, to consideration of flows (foliations) on the Riemann surfaces of genus higher than one. Nevertheless, already the circle maps



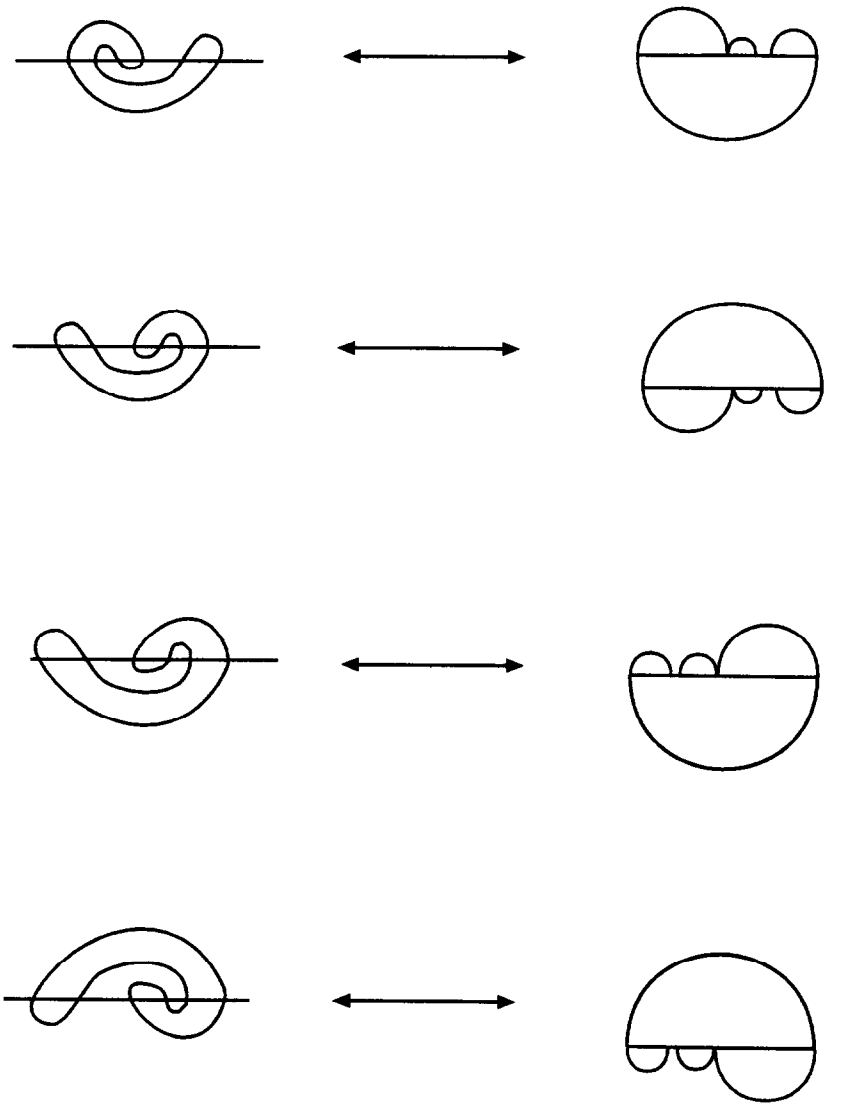


Fig. 16. The simplest meanders of order $n = 1$–3 and the associated with them meandritic labyrinths.

exhibit all qualitative features of more complicated situations. Moreover, as it was shown by Thurston [39], all surface homeomorphisms can be actually related to the circle maps through the construction which he calls the "earthquake". [3]

Let us briefly explain why this is so. (In the Appendix of Part II these arguments will be extended to three-dimensional hyperbolic manifolds.) As in the case of the circle maps discussed in the previous section, the surface homeomorphism $f : R \to R'$ (where both R

[3] More recent discussion of earthquakes could be found in paper by F. Bonahon [AMS Transactions 330 (1992) 69–95].

and R' are some Riemann surfaces) can be lifted to the universal covering space which is either the Poincare upper halfplane H^2 defined by

$$H^2 = \{z = x + \mathrm{i}y \in C \mid y > 0\} \tag{4.1}$$

or the open unit disc D^2

$$D^2 = \{w = u + \mathrm{i}v \in C \mid u^2 + v^2 < 1\} \tag{4.2}$$

which is related to H^2 via mapping

$$\mathrm{w} = \frac{z - \mathrm{i}}{z + \mathrm{i}}, \quad z \in H^2. \tag{4.3}$$

Since every Riemann surface R of genus g greater than one can be obtained as the quotient $R = H^2/\Gamma$ where Γ is some Fuchsian group, i.e. the group of linear fractional transformations of the type

$$\gamma(z) = \frac{az + b}{cz + d}, \quad ad - bc = 1 \tag{4.4}$$

with a, b, c, d being *real* numbers, we can always lift the flow on surface R to the flow on D^2 where by means of earthquakes the description of flows becomes connected with the description of circle maps. Let now C be some closed curve on R which is homotopically nontrivial (that is it cannot be shrunk to the point). It can be proven [40] that,

Theorem 4.1. *C is freely homotopic to a unique closed geodesic l.*

Also, it is known [41] that,

Theorem 4.2. *For a closed Riemann surface of genus g there are $3g - 3$ independent closed geodesics (please, see Fig. 5 of Part II).*

Finally, following Thurston [25] and Casson and Bleiler [42], we introduce

Definition 4.3. Disjoint union of geodesics is called lamination ($\mathcal{L}$); the geodesics contained in $\mathcal{L}$ are called leaves of $\mathcal{L}$.

It can be very easily shown [17,18] that, when lifted to H, the geodesics look like that depicted in Fig. 17. That is they are either halfcircles with center on the real axis or the

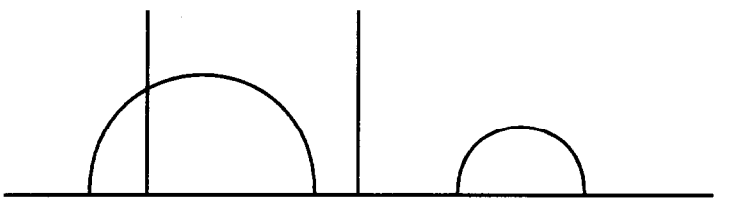


Fig. 17. Geodesics on the Poincare upper H-plane.

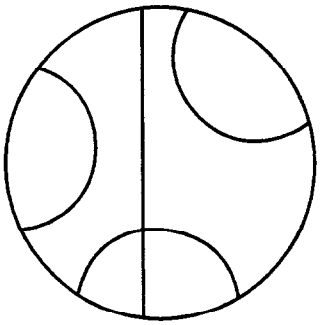


Fig. 18. Geodesics on the unit open disc D^2.

lines perpendicular to the real axis. When transformed from H to D^2, the halfcircles go to halfcircles and the vertical lines to the diameters of D^2 as depicted in Fig. 18. If we consider the case of lamination $\mathcal{L}$ then, by construction, the geodesics are nonintersecting on R, and whence on D^2 [43]. This means that in D^2 they look like the projective meanders (before identification of points on S^1 is made) described in the previous section. A simple minded permutation of these geodesics in D^2 produces the Catalan number C_n, Eq. (3.9). In general, the situation is considerably more complicated [23,25,42].

If for a given Riemann surface of genus g with n punctures we have the inequality $2g + n - 2 > 0$, then the mapping class group $M(R)$ is defined through its action on $\mathcal{L}$. If $f \in M(R)$ and $C = \{c_1, \ldots, c_r\} \in \mathcal{L}$, then the mapping f is *reducible* if

$$f(C) = C. \tag{4.5}$$

Alternatively, it is being said that f is *reduced* by C.

If, in accord with the theory of braids [44], we assume that Eq. (4.5) holds for any permutation of the set C (and this is indeed the case [42]), then the Catalan number, Eq. (3.9), naturally emerges. The reducible case is not the only possibility however. If, say, f is reduced by C, then let $S = \{s_1, \ldots, s_m\}$ be the set of components of the complement of $R - C$, e.g. see Fig. 5 of Part II. Since f permutes s_i, let n_i be the smallest power of f such that

$$f^{n_i}(s_i) = s_i. \tag{4.6}$$

For any $n_i > 0$ such mapping is called *periodic*. It could as well be that Eq. (4.6) does not have fixed point(s) for any finite n_i. In this case, if Eq. (4.5) still holds, then the mapping is still reducible, but if neither (4.5) or (4.6) hold, then the mapping is called *pseudo-Anosov*. This term will be explained below. Before doing so, we notice that Eq. (4.5) can be lifted to D^2 so that with respect to the ends of geodesics lying on S^1 we have some sort of a mapping of a circle analogous to that discussed in the previous section. According to Thurston [39], "every homeomorphism of the circle S^1 extends to the homeomorphism of the disc D^2". (In the Appendix to Part II we shall discuss the extension of this result to the case of three-manifolds). Evidently, the reverse should be true as well. In which case, the lift of Eq. (4.6) to D^2 has its analogue in Eq. (3.6) and the case when Eq. (3.6) does not have

solution for any q corresponds to the pseudo-Anosov type of mapping. Let us now explain the meaning of the word "pseudo-Anosov". According to [17,18], the following theorem can be proven.

Theorem 4.4. *The geodesic flow on H^2/Γ is the Anosov flow.*

The best way to explain this is again pictorial. We lift the geodesics from R to H in order to obtain the picture given in Fig. 19. Accordingly, we can also look at the same picture by using the disc D^2 model. In this case we have the situation depicted in Fig. 20. The circles tangent to x-axis at some point x_0 (in H-plane) are called *horocycles*. It can be shown, that the set of horocycles (touching x-axis at x_0) and the set of geodesics (emanating from x_0) are mutually orthogonal (thus forming stable and unstable manifolds, see below). For the geodesics directed as in Fig. 19 it is intuitively clear that the flow is expanding (unstable). If we would change the direction of geodesics (i.e. if we change the time evolution from t to $-t$), then the flow will be contracting. It can be shown [18] that the rate of expansion is e while the rate of contraction is e^{-1} (e = exp). Now, we are ready for a general definition of the Anosov flow.

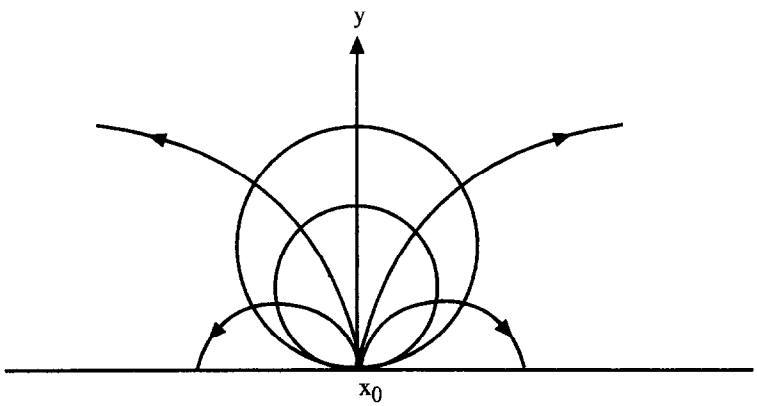


Fig. 19. Anosov flow on H-plane.

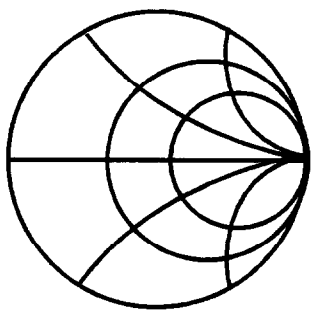


Fig. 20. Anosov flow on the disc D^2.

Definition 4.5. If $A : M \to M$ is some diffeomorphism of compact manifold M such that the tangent bundle $T_x M$ for any point $x \in M$ is decomposable into direct sum $\oplus$ of the type

$$T_x M = X_x \oplus Y_x, \tag{4.7}$$

so that in some (Riemannian) metric, and for some number $\lambda > 1$ the following inequalities hold

$$\|A_* \xi\| \geq \lambda \|\xi\| \quad \forall \xi \in X_x \quad \text{and } \|A_* \eta\| \leq \lambda^{-1} \|\eta\| \quad \forall \eta \in Y_x, \tag{4.8}$$

where A_* is the differential of the operator A (acting in tangent space) and $\|\cdots\|$ is the usual square of the length in some (Riemannian) metric space, then the flow $T_x M$ is Anosov.

Remark 4.6. *If, in addition, the flow contains some singularities, then it is called pseudo-Anosov* [45]. *The sets X_x and Y_x are called (un)stable foliations respectively.*

Remark 4.7. *It can be shown, that the geodesic flow on H is ergodic [18]. That is there is some measure $\mu(x)$ such that*

$$\int \mathrm{d}\mu(x)\varphi(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathrm{d}t \varphi(A^t x(t)), \tag{4.9}$$

where A^t is some evolution operator (continuous analog of A_* in (4.8)).

In the case of pseudo-Anosov flows it is customary to introduce the transverse (vertical) measure which can be intuitively understood using the following picture, e.g. see Fig. 21. That is, if *locally* the foliation looks like the set of horizontal lines (away from singularities), Fig. 21(a)), then, when we try to restore the global picture, the local charts should be glued

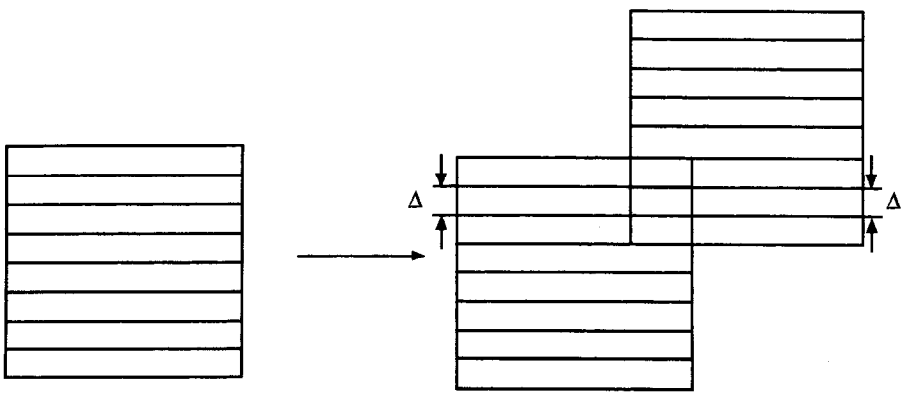


Fig. 21. For the measured foliation the vertical distance between the leaves on two adjacent patches (local charts) must be the same.

in such a way that the vertical distance Δ between the leaves on one chart is expected to be in agreement with the vertical distance Δ on another. If this is the case, the foliation is called *measured*. We shall be dealing only with measured foliations [10,25] in the rest of this work in accord with Poenaru [10], Langevin [11] and Ref. [25].

The above nice picture leaves us with no clues about the mechanisms by which the flows can be changed from the reducible to periodic or from the reducible to pseudo-Anosov, etc. As in the case of circle homeomorphisms (Section 3), there must be some *physical* reasons of changing from one regime to another. The best way to get some physical feeling of the processes which cause "phase transitions" (i.e. changes in the flow regime) is through the notion of "*train tracks*" introduced by Thurston [20] and subsequently developed by many other mathematicians, e.g. see Refs. [46,47]. The idea behind the train track is simple but very powerful. To appreciate it, it is helpful to recognize that there is some similarity between the way knots (or links) are described and the train tracks. In the case of knots (links) one usually studies a "shadow" of a knot obtained by projecting it onto some arbitrary plane [48]. The projection is a four-valent planar graph K without the dangling ("dead") ends. Evidently, different knots may have the same projection for a plane which orientation is fixed. To distinguish between different knots (links) one has to resolve each 4-valent vertex (i.e. to decide which strand is "over" and which is "under"), and in addition, to use the set of isotopy moves (the Reidemeister moves) to bring one projection in accord with the other. It is believed, that, at least in principle, one can decide if two knots are equivalent by performing some finite (but could be very large!) sequence of the Reidemeister moves. In all these operations the physical nature of a knot is not playing any role. In the case of train tracks one is also dealing with a graph T without dead ends and also there are the isotopy moves (analogous to the Reidemeister moves), but in addition, and this is the most important, there are moves which do not respect isotopy. These moves change the topology of the graph T and could be associated with some physical processes as we shall explain below and in Part II (Section 6 and Appendix A). The graph T has two basic building blocks which can be easily recognized, e.g. see Fig. 22. One can see from this figure that, instead of having all leaves and precise angles, one can actually "survive" only with topologically equivalent objects which are smooth at joints. With the basic building blocks just defined, we can construct our first train track. In particular, instead of having rather complicated foliation pattern depicted in Figs. 13 and 14, we can draw the following graph depicted in Fig. 23. Of course, the r.h.s. of Fig. 23 provides still only a coarse grained foliation pattern thus creating an illusion that the line is closed (and this is indeed the case for the "rational" labyrinth). The numbers on the graph are quite arbitrary but are subject to one restriction: at every switch the sum of "entering" numbers is equal to the sum of the "exiting" numbers. The numbers are associated with the invariant transverse measures which were formally introduced earlier. Since one can construct from the collection of 3-valent vertices the vertex of more complicated nature, e.g. 4-vertex, 5-vertex, etc., it makes sense to talk about given vertex v_k ($k = 1 - M$) in general. If $\{a_i^{\text{in}}\}$ is the set of input branches and $\{a_j^{\text{out}}\}$ is the set of output branches, then for a given vertex v_k the *switch condition* can be written as

$$\sum_i \mu_{v_k}(a_i^{\text{in}}) = \sum_i \mu_{v_k}(a_i^{\text{out}}). \tag{4.10}$$

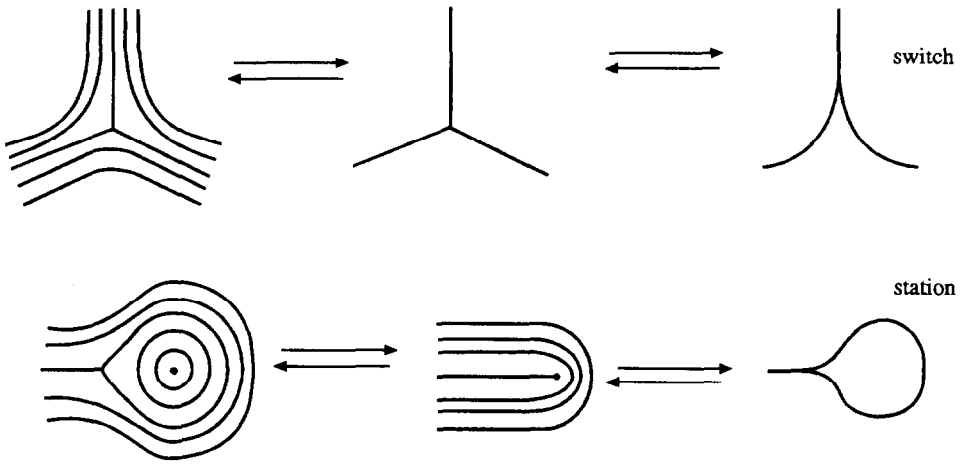


Fig. 22. Basic building blocks of the train track.

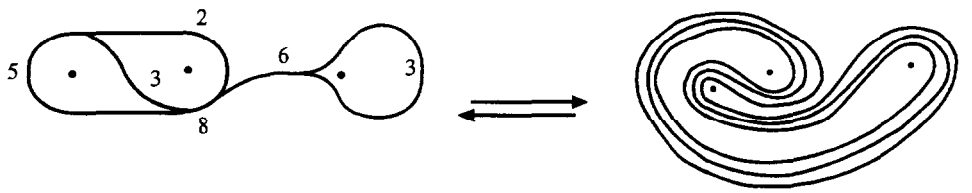


Fig. 23. A simple train track and the foliation pattern which is associated with it.

The switch condition is analogous to the Kirkhoff rule for currents known in physics [49]. It could be proven [20,45], that the assignment of weights (transverse measures) $\mu(a_i)$ for each branch of the train track which is subject to the switch conditions, Eq. (4.10), at each vertex is equivalent to the reconstruction of the entire measured foliation (up to isotopies and Whitehead moves, Fig. 4) [23]. The new feature which makes the train tracks more complicated than knots (links) lies in *additional moves* which, unlike the isotopy, do change the topological type of the track. These moves have the major physical significance as it will be demonstrated shortly. Before doing so, we would like to provide a list of topology changing moves. They are depicted in Fig. 24. The shift operation can be performed without reference to weights but, of course, the switch condition, Eq. (4.10), should be obeyed before and after the shift. The split (or collapse) is dependent, however, upon the particular distribution of weights. The shift does not destroy (or create) the existing singularities while the split (collapse) may destroy (create) the singularities thus apparently violating the P–H theorem. This deficiency can be easily corrected if some physics is taken into consideration. To do so, several steps are still required. *First*, following Refs. [46,47], let us consider the process of collapse as the matrix operation. That is, we form the column vector made of weights $a', b', \ldots, e'$ and look for the transition matrix $\mathbf{S}^{-1}$ such that $\mathbf{x} = \mathbf{S}^{-1}\mathbf{x}'$, or more

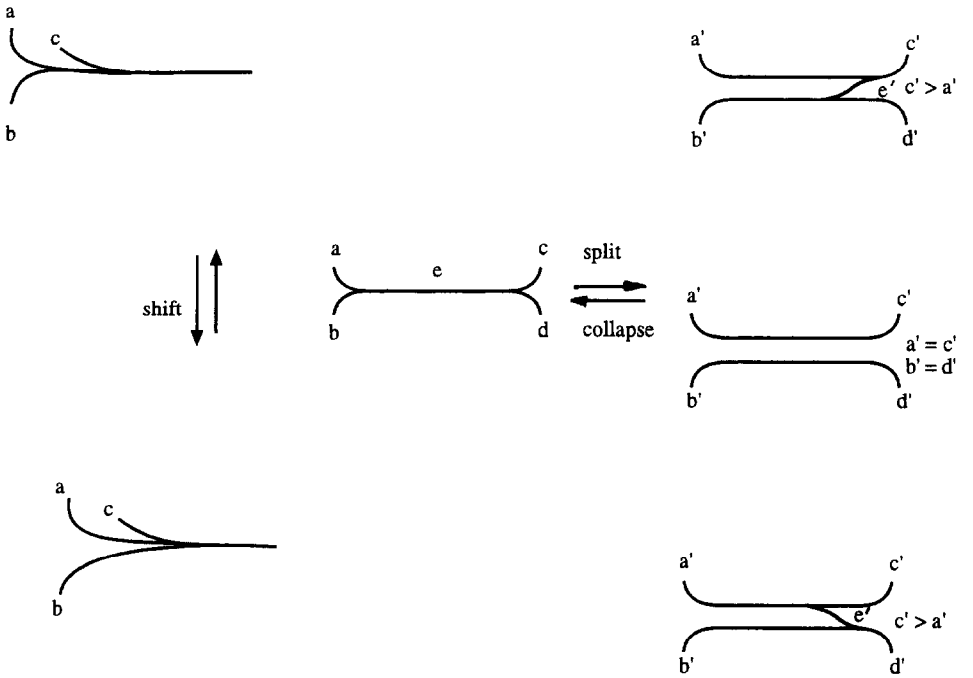


Fig. 24. Major topology nonpreserving moves for the train tracks.

explicitly, when $c' < a'$, we obtain,

$$\begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} a' \\ b' \\ c' \\ d' \\ e' \end{pmatrix}. \tag{4.11}$$

This result can be easily understood based on Eq. (4.10) and Fig. 24. Indeed, according to Eq. (4.10) we have $a' + e' = c'$ $(c' > a')$ and also $e' + d' = b'$ for the split diagram while $a + b = e$ and $c + d = e$ for the collapsed diagram. At the same time, using Eq. (4.11), we obtain: $a = a'$; $b = b'$; $c = c'$; $d = d'$ but $e = a' + d' + e'$. The last result becomes an identity if the Kirkhoff sum rules, just obtained, are utilized. Evidently, the corresponding matrices can be constructed for all elementary processes depicted in Fig. 24, and furthermore, if, for example, $\mathbf{S}$ is the split matrix, then $\mathbf{S}^{-1}$ is the corresponding collapse matrix, etc. *Second*, instead of considering the individual (local) processes, we can consider the state vector of the *entire* graph T which is just a column vector of all transverse measures which we denote as $\vec{\mu}$. The evolution of the entire graph is determined then by some transition matrix $\mathbf{T}$ so that

$$\vec{\mu}' = \mathbf{T}\vec{\mu}. \tag{4.12}$$

Third, this equation is still unphysical. To make it physical, we have to ascribe certain statistical weights to each of the diagrams depicted in Fig. 24. The weights could be chosen either on the basis of statistical mechanics arguments (e.g. in the case of liquid crystals) or on the basis of quantum mechanical arguments (e.g. in the case of gravity). For instance, in the case of liquid crystals one can think of energy of the defect formation and destruction (for more details, please, see the next section and Sections 4 and 6 of Part II) as well as about the average energy (per defect) as a function of ν (defined by Eq. (2.5)). Irrespective of the specific form of the weights, it is clear, that the time evolution of the graph T can be described in terms of the master equation (as it is usually done in statistical mechanics)

$$\frac{\partial \mu_i}{\partial t} = \sum_{i \neq j} (W_{ij}\mu_j - W_{ji}\mu_i). \tag{4.13}$$

The matrix W_{ij} is *not* symmetric unfortunately. This can be seen already from the Eq. (4.11). In the case of $2+1$ gravity the above equation could, in principle, provide an alternative to the Wheeler–DeWitt equation [50]. Eq. (4.13) may have time -independent solution which (in terms of discrete maps, e.g. see Section 3), by analogy with Eq. (4.5), we may call *reducible*. It may as well have a periodic (in time) solution which we may (or may not) associate with the previously defined *periodic* (in space) case. Finally, it may not have any stationary or periodic solutions. This then will be indicative of the pseudo-Anosov type of evolution. The actual analysis of Eq. (4.13) would require a large scale numerical work, and evidently, some drastic approximations will be required. Therefore, for the time being, we would like to explore yet another possibility. Before doing so, several remarks should be made.

Remark 4.8. *Since in the case of train tracks there are no dead ends, this concept should be applied to open surfaces, e.g. to the open disc D^2, with some caution, e.g. see* [51].

Remark 4.9. *Since the surface of nonnegative Euler characteristic contains no train tracks* [20,46,47] *some care should be taken to by-pass this difficulty. This is discussed to some extent in the Appendix A (see also* [34,51]).

Remark 4.10. *Without use of the probabilities leading to the master Eq.* (4.13), *different outcomes of iteration had been studied already* [46,47,51,52]. *Clearly, the process is reducible, periodic or pseudo-Anosov depending upon what sequence of moves, Fig.* 24, *had been used. For example, if only direct processes are being considered, such as the right (R) and the left (L) splits and the shift (S), then the matrix sequence could be associated with some words in the alphabet made up of three letters* [52] *so that, in general, Eq.* (4.13) *describes a random walk on the mapping class group $M(R)$ as could be realized upon reading of Ref.* [23].

Incidentally, the three letters alphabet is also being used for the meanders [35]. Hence, we would like now to study how one can use the meanders in resolving the types of dynamical behaviour. This is accomplished to some extent in the next section while in the Appendix A we provide some evidence that the meandritic labyrinths are indeed associated with the pseudo-Anosov homeomorphisms.

5. Phase transitions in the system of meandritic labyrinths

Phase transitions in the system of meanders was recently considered in [36] in connection with the problem of folding of polymers. At the same time, the arguments used in [36] could be traced back to the arguments made by Peierls in connection with phase transintion in two-dimensional Ising model [53]. To facilitate reader's understanding of the development which follows, we would like to provide here a summary of the arguments made by Peierls.

In the case of planar Ising model one is expecting to have droplets of spins of the "wrong" sign embedded among the "bulk" spins which have the "correct" sign (e.g. "+" for the correct and "−" for the wrong). The interface between the correct and the wrong spins determines the interfacial energy E_L of the droplet so that for the droplet of perimeter L the energy is given by $E_L = LJ$ with J being known, in principle, constant. The partition function Z_I for the Ising model can now be written as follows:

$$Z_I = 2\sum_{L=0}^{\infty} G_L \exp\{-\beta J L\}, \quad \beta^{-1} = k_B T, \tag{5.1}$$

where $k_B T$ is the usual temperature factor. Since the Ising model is defined on the lattice, L is necessarily discrete and G_L is the combinatorial entropic factor which determines the number of ways the droplets of the total perimeter length L can be arranged on the lattice.

The difficulty in calculating Z_I lies in determining the factor G_L correctly. If, following Peierls, we make a drastic approximation: $G_L \simeq 4^L$ (since for the square lattice the coordination number is 4), then Z_I can be calculated at once with the result

$$Z_I \simeq (1 - 4\lambda)^{-1}, \tag{5.2}$$

where $\lambda = \exp\{-\beta J\}$. The above expression makes sense only for $\lambda < 0.25$ so that 0.25 determines the critical temperature of the Ising model through equation: $4\lambda^* = 1$. This result is interesting to compare against the exact result [54]: $\lambda^* = \sqrt{\sqrt{2} - 1} \simeq 0.64$ (for $J = 1$). Obviously, the value 0.25 appears to be too low even if compared with the Bethe approximation [55] which yields $3\,(\lambda^*)^2 = 1$. This was noticed already by Peierls [53]. Nevertheless, the above crude estimate can be systematically improved and the Peierls arguments are the most powerful tool in general study of phase transitions in discrete spin systems as is well known [55]. The exact result, $(\lambda^*)^2 = \sqrt{2} - 1$, is valid only for the square lattice. In the case of triangular lattice the exact result is different: $(\lambda^*)^2 = (\sqrt{3})^{-1}$, while for the hexagonal lattice, it is different again: $(\lambda^*)^2 = (2 + \sqrt{3})^{-1}$, e.g. see [54]. Hence, the critical temperature is actually a function of two parameters: J and z, where z is the coordination number of the lattice. Something similar happens in our (meandritic) case too.

In the case of Ising model we can effectively define the notion of an order parameter. This is less obvious in the case which we are going to consider. What is important for us, however, is the fact that the partition function Z_I diverges. We shall, by analogy with the Ising model, associate such divergence with the phase transition, in our case from the

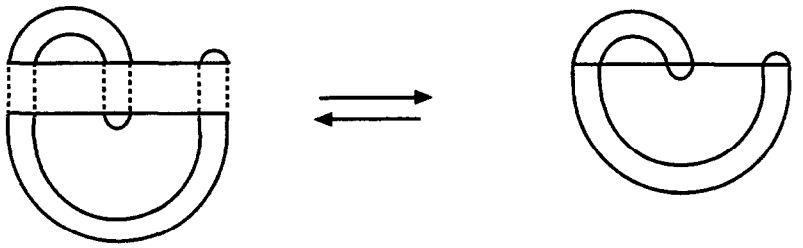


Fig. 25. Construction of a typical meander.

meandritic (pseudo-Anosov, according to Refs. [23,34] and Appendix A) to the periodic and/or reduced phases (since these are the only existing possibilities).

Using the results of Section 2, Eq. (2.5) and Section 4 of Part II, Eq. (4.49), the total surface energy of n defects can be estimated as

$$E_n \geq \frac{\tilde{K}}{4} n q^2 \ln\left(\frac{1}{\nu}\right) \equiv n\hat{J}, \tag{5.3}$$

where the charge q is equal to the index of the singularity and $\tilde{K}$ is some known constant (defined in Section 4, Part II) related to the surface tension. In arriving at Eq. (5.3) we had assumed that only Y's and thorns are present (so that the magnitude of charges are all the same). To estimate the combinatorial factor (analogous to G_L in Eq. (5.1)) the following arguments are helpful.

According to [37], any meander can be built by superposition of two arc configurations of the same order: one is considered to be the top while another is the bottom as depicted in Fig. 25. Since both the top and the bottom are configurations of the same order, then by concentrating our attention, say, on the top configuration, we can obtain a projective meander (e.g. see Section 3) by identifying the beginning and the end of the horizontal straight line. In this case we already know that the number of possible configurations is C_n. Evidently, we can do the same for the bottom arc system thus obtaining another C_n. If we try all possible permutations, then, a priory, there is no guarantee that if the top and the bottom are connected together we shall obtain just one connected meander. More likely, we shall obtain C_n^2 meanders some of them connected and some not. More exactly, let $M(x)$ be the meander generating function, i.e.

$$M(x) = \sum_{n=0}^{\infty} \mathrm{M}_n x^n, \tag{5.4}$$

where M_n was defined in Section 3. We have the following theorem.

Theorem 5.1. $C_n \leq M_n \leq C_n^2$.

Proof. The upper bound was just established. The lower bound can be established if we associate the system of meanders with the three letter alphabet (as was briefly mentioned in Section 4). For more details, please, consult Ref. [35]. □

Among all of C_n^2 meanders of order n not all are topologically different. Let N_n denote the number of topologically distinct configurations of meanders which altogether pass through $2n$ points (which actually may belong to different lines), then the generating function for these numbers could be defined as

$$N(x) = \sum_{n=0}^{\infty} \mathrm{N}_n x^n. \tag{5.5}$$

If we disrespect the topological differences, then the generating function for the system of meanders can be written as

$$B(x) = \sum_{n=0}^{\infty} C_n^2 x^n. \tag{5.6}$$

Theorem 5.2. *Functions $B(x)$ and $N(x)$ are connected with each other through the functional equation*

$$B(x) = N(xB^2(\mathrm{x})) \tag{5.7}$$

Proof. Please, consult [35]. □

Corollary 5.3. *The meandritic numbers $M_n^{(k)}$ introduced in Section 3 are generated via the following generating function* [36,37]*:*

$$m_n(g) = \sum_{k=0}^{n} M_n^{(k)} g^k \tag{5.8}$$

provided that $m_n\ (g = 1) = C_n^2$.

Based on the results just described, we are ready now to write down the partition function for the system of meandritic labyrinths. It is given by

$$Z_g(x) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} M_n^{(k)} g^k \mathrm{x}^n, \quad x = \exp\{-\beta \hat{J}\}, \tag{5.9}$$

where $\hat{J}$ is defined by Eq. (5.3). This equation, in principle, is exact but, it requires knowlege of g. Evidently, g is the fugacity which determines the average number of meanders in the cluster $\langle k \rangle$ which is given by

$$\langle k \rangle = g \frac{\partial}{\partial g} \ln Z_g(x). \tag{5.10}$$

This quantity is hard to estimate (see, however, some attempts in this direction below). Even if we would succeed, still, we would need to invert the infinite power series in order to write the fugacity g in terms of $\langle k \rangle$. Therefore, we would like to make an approximation

based on the exact result coming from Eq. (5.8) for $g = 1$. Thus, we obtain the following approximation for the partition function of the system of meanders:

$$Z(x) = \sum_{n=0}^{\infty} C_n^2 \mathrm{x}^n, \quad x = \exp\{-\beta \hat{J}\}. \tag{5.11}$$

Remarkably enough, this partition function admits the *exact* resummation [35] with the result ($t^2 = x$):

$$Z(t^2) = \frac{1}{4t^2}\left(-1 + \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\varphi \sqrt{1 - 8t\cos\varphi + 16t^2}\right). \tag{5.12}$$

At the same time, one can study the convergence of $Z(x)$ based on known asymptotic value for C_n:

$$C_n \simeq \mathrm{const}\frac{4^n}{n^{3/2}}, \quad n \to \infty. \tag{5.13}$$

From here, we obtain,

$$\frac{C_{n+1}^2}{C_n^2} \sim 16, \tag{5.14}$$

and therefore, $0 < x \leq 1/16$. Substitution of $x^* = 16$ into Eq. (5.12) produces *finite* result:

$$\frac{1}{4}Z\left(\frac{1}{16}\right) = \left(\frac{4-\pi}{\pi}\right). \tag{5.15}$$

For $x > 1/16$ the partition function diverges, of course.

Clearly, the condition $16x^* = 1$ determines the critical temperature for the fixed value of v or determines the critical "density" v^* for a fixed temperature and the value of surface tension $\tilde{K}$ in view of Eq. (5.3).

The analysis presented above is still incomplete since one can calculate now $N(x)$ in view of Eq. (5.7). Following [35], let us multiply both sides of Eq. (5.7) by t ($t^2 = x$). We obtain,

$$\Phi(t) = tB(t^2) = tN(\Phi^2(t)). \tag{5.16}$$

Let, furthermore, $\lambda = \Phi(t)$ so that $t = \Phi^{-1}(\lambda)$. Then, we obtain,

$$N(\lambda^2) = \frac{\lambda}{\Phi^{-1}(\lambda)}. \tag{5.17}$$

Since using closed form result, Eq. (5.12), $\Phi(t) = tZ(t^2)$ can be calculated, at least in principle, then $N(\lambda^2)$ can be calculated using Eq. (5.17) as well. The radius of convergence for $N(x)$ is equal to $((4-\pi)/\pi)^2$. Obtained results are in formal qualitative accord with that known in physical literature on liquid crystals [14]. But now we know that the disordered (liquid-like) phase is actually pseudo-Anosov while the ordered (hexatic) is periodic and the reducible is solid-like. The phase transition mechanism, however, has absolutely nothing

to do with the Kosterlitz–Thouless type of transition which was discussed in Section 2. Additional details related to this subject are provided in Section 4, Part II.

Note added in proof. When this paper was completed we had found paper by Ishikawa and Lawrentovich [Europhys. Lett. 41 (1998) 171–176] in which the Whitehead moves depicted in our Fig. 4 had been detected experimentally. In addition, the paper by Penner [Adv. Math. 101 (1993) 31–49] had also came to our attention. In this paper the train-tracks are being used to describe folding of RNA. Since in [36] the folding of proteins was considered with help of the meanders, the above reference provides additional support to the results presented in Sections 3–5 relating meanders to train tracks. Some details relevant to this relationship could be also found in Section 7.5 of Part II.

Appendix A

In this appendix we would like to provide some arguments in favor of the statement made in Section 4 that the meandritic labyrinths are associated with the pseudo-Anosov homeomorphisms.

To this purpose let us consider the simplest labyrinth which is made out of two copies of the disc D^2 both of which containing just two thorns (to be in accord with P–H theorem). When glued properly, these two discs will form a sphere S^2 with foliation forming a labyrinth (e.g. see [34], p. 29). We need now to explain why such foliation could be of pseudo-Anosov type.

Consider an auxiliary problem about the foliations on the torus [17,18,42]. If we regard the torus T^2 as a quotient of $\mathbf{R}^2$ by the integer lattice Z^2, then the homeomorphisms h_α of T^2 are generated by the group $GL_2(Z)$ since any element α of $GL_2(Z)$ maps Z^2 into itself thus inducing the continuous map $h_\alpha : T^2 \to T^2$. The homeomorphism is orientation preserving if $\det(\alpha) = 1 (\forall \alpha \in GL_2(Z))$ and, in this case, $GL_2(Z)$ becomes $SL_2(Z)$ which is represented by 2×2 matrix given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(Z) \quad \text{if } ad - cb = 1. \tag{A.1}$$

The eigenvalues of the above matrix are obtained, as usual, as the roots of the characteristic polynomial

$$t^2 - (a + d)t + (ad - cb) = 0$$

or, in view of (A.1),

$$t^2 - (a + d)t + 1 = 0. \tag{A.2}$$

There are three possibilities in general:
(1) the roots of (A.2) are complex (when $a + d = 0$, 1 or -1);
(2) both roots of (A.2) are equal to ± 1 (when $a + d = 2$);
(3) the roots of (A.2) are distinct reals (when $|a + d| > 2$).

By analogy with the circle maps considered in Section 3, it can be shown [18,42], that the first possibility produces *periodic* maps of T^2, the second produces *reducible* maps of T^2 (that leave a simple Jordan curve fixed (possibly with reversed orientation))while the third possibility is responsible for the Anosov (*not pseudo-Anosov)* type of flow [17,18] on T^2 which was defined in Section 4. Consider now our S^2 made up of two copies of D^2, each containing two thorns. According to the results of Part II thorns are singularities in the complex z-plane (or on S^2) associated with punctures. As it is usually done in the complex analysis, we can make branch cuts (as depicted in Fig. 26) in order to construct the two sheeted Riemann surface since the singularities are of the type $(\sqrt{z})^{-1}$. In Fig. 26 these two sheets are depicted as onion-like sphere with two layers. If the outer sphere is

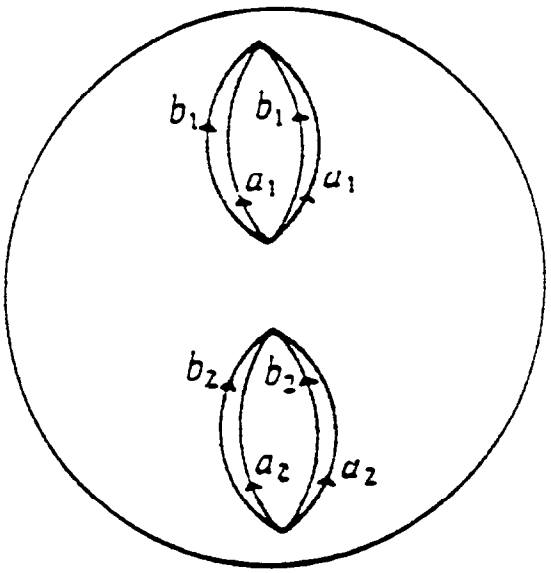


Fig. 26. Two sheeted covering of S^2 before glueing.

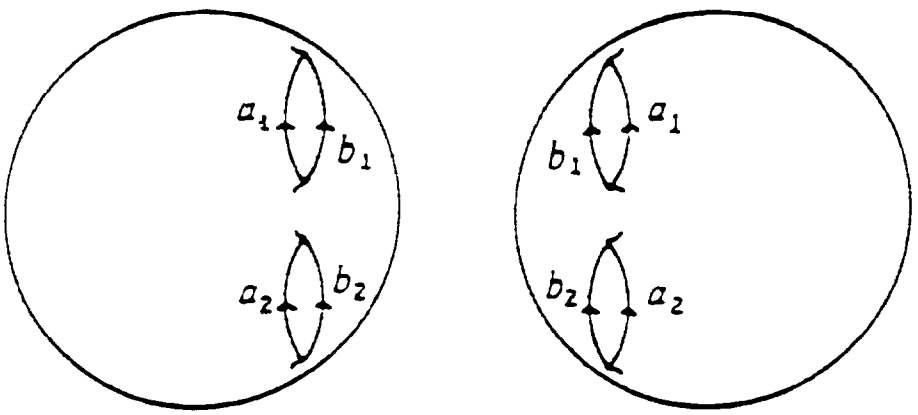


Fig. 27. Identification of cuts between the inner and the outer spheres produces torus T^2.

peeled off and placed opposite to the inner sphere, as depicted in Fig. 27, then, by obvious identification of edges, the surface which is homeomorphic to T^2 is formed. Hence, T^2 is a two-fold ramified branched covering of S^2. The points of ramification are the above thorn-like singularities. In general, one can prove the following

Theorem A.1. *Every closed orientable surface (of finite genus) is two-fold ramified branched cover of* S^2.

Proof. Please, consult [57,58]. □

If for T^2 we have the case of the Anosov flow, then it is being transferred onto S^2 in the form of pseudo-Anosov flow (because of 4 singularities placed on S^2). Details of construction of pseudo-Anosov homeomorphisms fot this case are given in [59] and are mainly based on the earlier work by Katok [60]. For more than four defects in view of Theorems A.1 and 4.4 of the main text, it is also possible, in principle, to establish connection between the labyrinths [34] and pseudo-Anosov homeomorphisms. Please, consult Expose′11 and Expose′12 of Ref. [23] and Section 4 of Ref. [45] for additional details. For more information on branched coverings, please, consult Ref. [61].

References

[1] M. Nakanara, Geometry, Topology and Physics, Adam Hilger, Bristol, 1990.
[2] J. Moore, Lectures on Seiberg-Witten Invariants, Springer, Berlin, 1996
[3] G. Hemion, The Classification of Knots and 3-Dimensional Spaces, Oxford University Press, Oxford, 1992.
[4] H. Hopf, Differential Geometry in the Large, LNM No. 1000, Springer, Berlin, 1989.
[5] V. Nemytskii, V. Stepanov, Qualitative Theory of Differential Equations, Princeton University Press, Princeton, NJ, 1960.
[6] G. Godbillon, Dynamical Systems on Surfaces, Springer, Berlin, 1983.
[7] 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.
[8] C. Oseen, The theory of liquid crystals, Faraday Transactions 29 (1993) 883–899.
[9] F. Frank, On the theory of liquid crystals, Faraday Discussions 25 (1958) 19–28.
[10] V. Poenaru, Some aspects of the theory of defects in ordered media and gauge fields related to foliations, Comm. Math. Phys. 80 (1981) 127–136.
[11] R. Langevin, Foliations, energies and liquid crystals, Asterisque 107–108 (1983) 201–213.
[12] S. Deser, R. Jackiw, G. 't Hooft, Three dimensional Einstein gravity: dynamics in flat space, Ann. Phys. 152 (1984) 220–235.
[13] P. Menotti, D. Seminara, Energy theorem for 2+1 dimensional gravity, Ann. Phys. 240 (1995) 203–221.
[14] P. Chaikin, T. Lubensky, Principles of Condensed Matter Physics, Cambridge University Press, Cambridge, 1995.
[15] M. Katanaev, I. Volovich, Theory of defects in solids and three dimensional gravity, Ann. Phys. 216 (1992) 1–28.
[16] B. Halperin, D. Nelson, Theory of two dimensional melting, Phys. Rev. Lett. 41 (1978) 121–124.
[17] V. Arnold, Geometrical Methods in the Theory of Ordinary Differential Equations, Springer, Berlin, 1983.
[18] A. Katok, B. Hasselblat, Introduction to the Modern Theory of Dynamical Systems, Cambridge University Press, Cambridge, 1997.
[19] V. Arnold, Branching covering $CP^2 \to S^4$, hyperbolicity and projective topology, Siberian Math. J. 29 (1988) 36–47.

[20] W. Thurston, The Geometry and Topology of Three Manifolds, Lecture Notes, Princeton University, 1979; http://www.msri.org/gt3m/.
[21] E. Hague, P. Hemmer, The two dimensional Coulomb gas, Physica Norvegica 5 (1971) 209–217.
[22] J. Kosterlitz, D. Thouless, Metastability and phase transitions in two dimensional systems, J. Phys. C6 (1973) 1181–1203.
[23] A. Fathi, F. Laudenbach, V. Poenaru, Asterisque 66–67 (1979).
[24] J.-P. Otal, Le theoreme d'hyperbolisation pour les varietes de dimenson 3, Asterisque 235 (1996).
[25] W. Thusrton, On the geometry and dynamics of diffeomorphisms of surfaces, Bull. Am. Math. Soc. 19 (1988) 417–432.
[26] P.-G. de Gennes, The Physics of Liquid Crystals, Clarendon Press, Oxford, 1979.
[27] L. Kinsey, Topology of Surfaces, Springer, Berlin, 1993.
[28] B. Halperin, D. Nelson, Errata, Phys. Rev. Lett. 41 (1978) 519 and Ref. [16].
[29] R. Baxter, Exactly Solved Models in Statistical Mechanics, Academic Press, New York, 1982.
[30] R. Monasson, O. Pouliquen, Entropy of particle packings: an illustration on a toy model, Physica A 236 (1997) 395–410.
[31] A. Kholodenko, Th. Vilgis, The tube diameter in polymer melts, its existence and its relation to the quantum Hall effect, J. Phys. I (France) 4 (1994) 843–862.
[32] A. Kholodenko, Th. Vilgis, Some Geometrical and Topological Problems in Polymer Physics, Physics Reports 298 (1998) 251–372.
[33] W. Thurston, Three-Dimensional Geometry and Topology, vol. 1, Princeton University Press, Princeton, 1997.
[34] H. Rosenberg, Labyrinths in the disc and surfaces, Ann. Math. 117 (1983) 1–32.
[35] S. Lando, A. Zvonkin, Meanders, Selecta Mathematica Sovetica 11 (1992) 117–144.
[36] P. Di Francesco, O. Golinelli, E. Guitter, Meanders: a direct enumeration approach, Nucl. Phys. B 482 (1996) 497–535.
[37] P. Di Francesco, O. Golinelli, E. Guttier, Meanders and the Temperley-Lieb algebra, Comm. Math. Phys. 186 (1997) 1–59.
[38] S. Lando, A. Zvonkin, Plane and projective meanders, Theor. Computer Sci. 117 (1993) 227–241.
[39] W. Thurston, Low Dimensional Topology and Kleinian Groups, Cambridge University Press, Cambridge, 1986, pp. 91–112.
[40] W. Abikoff, The Real Analytic Theory of Teichmüller Space, Springer, Berlin, 1989.
[41] P. Buser, Geometry and Spectra of Compact Riemann Surfaces, Birkhauser, Boston, 1992.
[42] A. Casson, S. Bleiler, Automorphisms of Surfaces After Nielsen and Thurston, Cambridge University Press, Cambridge, 1988.
[43] A. Marden, K. Strebel, Differential Geometry and Complex Analysis, Springer, Berlin, 1985, pp. 195–204.
[44] J. Birman, Braids, Links and Mapping Class Group, Princeton University Press, Princeton, 1975.
[45] A. Marden, K. Strebel, Pseudo–Anosov Teichmüller mappings, J. D'Analyse Mathematique 46 (1986) 194–220.
[46] R. Penner, J. Harer, Combinatorics of Train Tracks, Princeton University Press, Princeton, NJ, 1992.
[47] R. Penner, Low Dimensional Topology and Kleinian Groups, Cambidge University Press, Cambridge, 1986, pp. 77–90.
[48] L. Kauffman, On Knots, Princeton University Press, Princeton, NJ, 1987.
[49] R. Bott, M. Duffin, On the algebra of networks, Trans. Am. Math. Soc. 74 (1953) 99–109.
[50] G. Esposito, Quantum Gravity, Quantum Cosmology and Lorentzian Geometries, Springer, Berlin, 1992.
[51] J. Los, Pseudo-Anosov maps and invariant train tracks in the disc: a finite algorithm, Proc. London Math. Soc. 66 (1993) 400–430.
[52] A. Papadopoulos, R. Penner, A characterization of pseudo-Anosov foliations, Pac. J. of Math. 130 (1987) 359–377.
[53] R. Peierls, On Ising's model of ferromagnetism, Proc. Cambr. Phil. Soc. 32 (1936) 477–481.
[54] C. Itsykson, J. Drouffe, Statistical Field Theory, vol. 1, Cambridge University Press, Cambridge, 1989.
[55] Ya. Sinai, Theory of Phase Transitions, Pergamon Press, Oxford, 1982.
[56] D. Rolfsen, Knots and Links, Publish or Perish, Houston, TX, 1990.
[57] R. Fenn, Techniques of Algebraic Topology, Cambridge University Press, Cambridge, 1983.
[58] J. Stillwell, Classical Topology and Combinatorial Group Theory, Springer, Berlin, 1993.

[59] J. Llibre, R. Mackey, Pseudo-Anosov homeomorphisms on a sphere with four punctures have all periods, Math. Proc. Cambr. Phil. Soc. 112 (1992) 539–549.
[60] A. Katok, Bernoulli diffeomorphisms on surfaces, Ann. Math. 110 (1979) 529–547.
[61] I. Berstein, E. Edmonds, On the construction of branched coverings of low dimensional manifolds, Transactions Am. Math. Soc. 247 (1979) 87–124.