

Random disks of constant curvature: The lattice story

Frank Ferrari^{*}

*Service de Physique Théorique et Mathématique, Université Libre de Bruxelles (ULB)
and International Solvay Institutes, Campus de la Plaine, CP 231, B-1050 Bruxelles, Belgium*



(Received 4 July 2024; accepted 9 November 2025; published 20 January 2026)

We lay the groundwork for a UV-complete formulation of the Euclidean Jackiw-Teitelboim two-dimensional models of quantum gravity when the boundary lengths are finite, emphasizing the discretized approach. The picture that emerges is qualitatively new. For the disk topology, the problem reduces to counting so-called self-overlapping curves, that are closed loops that bound a distorted disk, with an appropriate multiplicity. We build a matrix model that does the correct counting. The theories in negative, zero, and positive curvatures have the same UV description but drastically different macroscopic properties. The Schwarzian theory emerges in the limit of very large and negative cosmological constant in the negative curvature model, as an effective theory valid on distance scales much larger than the curvature length scale. In positive curvature, we argue that large geometries are ubiquitous and that the theory exists only for a positive cosmological constant. Our discussion is pedagogical and includes a review of several relevant topics.

DOI: [10.1103/3gb4-y8r8](https://doi.org/10.1103/3gb4-y8r8)

I. INTRODUCTION

A. Two-dimensional quantum gravity: From Liouville...

The starting point of our work is to postulate that a Euclidean theory of two-dimensional quantum gravity must be a diffeomorphism invariant theory of random metrics on a two-dimensional manifold $\mathcal{M}_{h,b}$ of fixed topology, with genus h and b boundary components.¹ A rigorous formulation of such theories can be obtained by following the Wilsonian paradigm, in the spirit of statistical physics. The models are first regularized by considering discretized surfaces made of flat polygons. The problem is then purely combinatorial: One enumerates all the possible ways to build a surface by gluing polygons together. This counting of “maps” was pioneered by Tutte in the 1960s [1]. In 1978, following ideas first put forward by ’t Hooft [2], Brézin, Itzykson, Parisi, and Zuber related the combinatorics of maps to the solution of formal matrix integrals in the large- N limit [3]. This point of view greatly simplifies the counting problem. The discretized theory being solved,

one must then define a continuum limit in which the size of the polygons goes to zero. The existence of nontrivial and universal continuum limits relies on the existence of critical points [4,5].

Another approach, pioneered by Polyakov in 1981 [6], is to use a direct field theoretic formulation in the continuum, called Liouville theory. One uses conformal gauge and treats the conformal factor as an ordinary two-dimensional quantum field. Diffeomorphism invariance of the underlying quantum gravity theory implies that the resulting Liouville model must be a conformal field theory. This remains true even when gravity is coupled to an arbitrary, possibly nonconformal, matter quantum field theory. This approach has scored a number of impressive successes [7,8]. When comparable, the results obtained in the continuum Liouville approach and in the discretized approach beautifully match. This subject has become so mature that it is by now a well-established field in probability theory; see, e.g., [9] and references therein.

A “topological” version of two-dimensional quantum gravity can also be studied. Instead of considering the space of all metrics on a given surface $\mathcal{M}_{h,b}$, one restricts oneself to metrics of constant negative curvature and requires the boundaries to be geodesics of given lengths. The resulting moduli spaces of metrics are finite dimensional. The associated partition functions are, thus, given by ordinary integrals that compute the volumes of these moduli spaces with respect to the Weil-Petersson measure, as a function of the boundary lengths. These volumes can be computed by using the Mirzakhani recursion relations [10], making the link with previous famous work by Witten and

^{*}Contact author: frank.ferrari@ulb.be

¹The study of Lorentzian signature metrics and/or the consideration of the sum over all topologies are, of course, essential in some applications and will be briefly discussed later.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

Kontsevitch [11]. The Mirzakhani recursion relations are themselves equivalent to the so-called topological recursion relations in matrix models [12].

Nice reviews and relevant references on these classic results may be found in [13].

B. ...to Jackiw-Teitelboim

In recent years, a version of two-dimensional quantum gravity, called Jackiw-Teitelboim quantum gravity [14], or JT gravity, for short, has been extensively studied. JT gravity lies somewhere between topological gravity and Liouville gravity. One fixes the bulk scalar curvature and the boundary lengths to be constants, as in the case of topological gravity, but one waives the constraint that the boundaries must be geodesics.² The boundary conditions are, therefore, similar to the FZZT (Fateev–Zamolodchikov–Zamolodchikov–Teschner) branes in Liouville gravity [15].³ JT gravity comes in three versions, depending on whether the bulk curvature is chosen to be negative, zero, or positive. Most of the literature focuses on the negative curvature case because of its direct relevance for holography. The three versions are identical at distance scales much smaller than the bulk curvature length scale: They have the same UV completion. The UV-complete formulation of the models can, thus, be studied by focusing first on the simplest case of the zero curvature theory, in which all the essential ideas can be introduced in a somewhat simplified context. The cases of negative and positive curvature will also be discussed, in a second stage. At long-distance scales, we shall explain that the three models display drastically different physics.

In order to avoid a confusion in the terminology, let us stress that for us the cosmological constant Λ is defined to be the constant multiplying the bulk area term in the action.⁴ In JT gravity, it, thus, has no direct relation with the sign of the bulk curvature, which is fixed independently. In particular, the negative curvature model is defined for any value of the cosmological constant, positive or negative, the near-hyperbolic limit most relevant for holography corresponding to $\Lambda \rightarrow -\infty$. On the other hand, we shall argue that the positive curvature model is defined only for positive cosmological constant $\Lambda \geq 0$, whereas the zero

curvature model is expected to exist as long as $\Lambda \geq \Lambda_c$, for a strictly negative critical cosmological constant Λ_c .

From the physics point of view, JT gravity may seem to be too simple. Indeed, since the bulk curvature is fixed, the bulk geometry cannot fluctuate, and, thus, the local bulk physics looks nongravitational. Consequently, on closed manifolds $\mathcal{M}_{h,0}$, JT gravity is equivalent to topological gravity.

However, the situation turns out to be much more interesting on manifolds with boundaries. The core of our work will be to show, following [16], that the situation is, in fact, even more interesting and far more subtle than the existing literature suggests.

For instance, in the case of the disk topology, the space of metrics is directly related to the space of immersions of the source disk into a canonical target space of constant curvature, which is either the hyperbolic space H^2 , the Euclidean plane E^2 , or the two-sphere S^2 , depending on whether we consider the case of negative, zero, or positive curvature. The target space geometry itself is totally rigid, which translates mathematically the fact that the local bulk physics looks nongravitational. However, there are a very large number of possible immersions. By pullback, this yields an infinite-dimensional space of metrics on the source disk. Intuitively, the degrees of freedom of the theory seem to be encoded in the shape of the boundary of the immersed disk into the fixed canonical target spaces. One of the main goals of the present paper is to explain in detail the fluctuating boundary picture, from first principles. This will reveal some truly surprising and unexpected features. In particular, the standard lore according to which there is a one-to-one correspondence between the degrees of freedom of the theory and the shape of the boundary modulo the action of the isometries of the target space will turn out to be incorrect. We shall also see that typical immersed boundaries have many self-intersections and are fractal curves.

The vast majority of the literature on JT gravity has focused on the case of negative curvature in a particular limit for which the lengths of the boundaries and the absolute value of the negative bulk cosmological constant go to infinity, their ratios being kept fixed in the bulk curvature units. The intuition behind this limit is that it corresponds to taking the infinite UV cutoff limit in the dual boundary description. This is often referred to as the “near-hyperbolic limit” or the “Schwarzian limit.”⁵ The conjectured dominant configurations in the limit correspond to smoothly wiggling boundaries that fluctuate mildly on distance scales much larger than the curvature length scale. Mathematically, they are described by the so-called “reparametrization *Ansatz*” for which the JT gravity path integral reduces to an integral over the group of

²An important consequence of our work will be to show that the boundaries in JT gravity are actually fractals with a nontrivial Hausdorff dimension. The boundary lengths that are fixed are, thus, “quantum lengths,” analogous to the diffusion time for Brownian paths, as will be explained later.

³In the original FZZT construction, one fixes the boundary cosmological constant. Fixing the boundary length is obtained by a Laplace transform. The important point is that no other condition is imposed on the boundary.

⁴For the disk topology, and when the nonzero curvature models are formulated in terms of a dilaton field, as is traditional, Λ is proportional to the boundary value of the dilaton.

⁵The terminology “near AdS” limit is also frequently encountered, but we will use the name “hyperbolic space” for the space of constant negative curvature in Euclidean signature and reserve the term “anti-de Sitter” for the case of Lorentzian signature.

diffeomorphisms of the circle governed by the Schwarzian action. The resulting model is exactly solvable [17]. Despite its simplicity, it has demonstrated its relevance in shedding light on some nontrivial questions in quantum gravity. It provides a holographic description [18–21] of strongly coupled quantum mechanical systems [22,23] that are relevant to describe near-extremal black holes, including, for instance, the process of black hole formation [24] or the physics of traversable wormholes [25]. The Schwarzian description has allowed the black hole information paradox to be addressed in a controlled setup [26]. At the same time, some unexpected subtleties with the holographic framework itself have emerged, in relation with the nonfactorization of amplitudes. In particular, the Schwarzian framework is not consistent with a quantum mechanical boundary description but requires an interpretation in terms of an ensemble average over quantum theories [27].

Our results imply that the Schwarzian theory is an effective theory, valid on distance scales much greater than the curvature length scale L . It emerges as a correct IR description of the negative curvature JT gravity model in the limit $\Lambda \rightarrow -\infty$. On distances of the order of or smaller than L , the structure of the boundaries become much more complicated than the smoothly wiggling boundaries described by the Schwarzian. In particular, the boundaries are self-intersecting fractal curves, and in the microscopic description of the models the length parameter is a “quantum” length β_q that has an anomalous dimension. The usual smooth boundary length ℓ characterizing the configurations included in the Schwarzian description is itself an effective IR parameter that can be expressed in terms of the microscopic parameters β_q , Λ , and L in the $\Lambda \rightarrow -\infty$ limit.

C. Microscopic and finite cutoff Jackiw-Teitelboim

The primary motivation for the work presented in this article and the companion papers [16,28] is the observation that, despite the important applications mentioned above and the fact that the relevance of the model is widely acknowledged, the theoretical foundations of JT quantum gravity have been very little studied and are poorly understood until now. In particular, the status of the subject is far from the one achieved by Liouville gravity.

Our goal is to provide some of the keys to a rigorous study of the theory, starting from first principles, in both the discretized and the continuous approaches. Reference [16] provides a synthetic exposé of the main findings. The present article focuses predominantly, but not exclusively, on the development of the discretized approach. The continuous approach, using conformal gauge, is further studied in the upcoming paper [28]. Ideas from the continuum conformal gauge formalism were used in [29] to study the JT gravity models coupled to a conformal field theory of central charge c , in the semiclassical limit $c \rightarrow -\infty$, at the one-loop order.

An important point is that we consider Jackiw-Teitelboim gravity for finite values of the microscopic parameters: the curvature length scale L , the bulk cosmological constant Λ , and the quantum boundary lengths $\beta_{q,i}$.⁶ As we have already mentioned, the boundaries in JT gravity turn out to be fractal.⁷ Thus, they do not have smooth boundary length parameters. Such a notion emerges only on long-distance scales in the negative curvature model, in the Schwarzian limit $\Lambda \rightarrow -\infty$. Instead, the relevant notion is a renormalized, quantum length.

In the negative curvature theory, working for finite values of the parameters is interpreted as introducing a finite UV cutoff in the holographic dual. The Schwarzian limit $\Lambda \rightarrow -\infty$, in which very large, nearly hyperbolic geometries dominate,⁸ is, thus, a large UV cutoff limit from the boundary perspective. This interpretation comes from the usual UV/IR relation in holography. However, it is important to stress that, unlike in higher dimensions, the strict limit of infinite cutoff is trivial when the bulk geometry is two-dimensional. This follows from the fact that, in two dimensions, there is a unique complete metric of constant negative curvature on the disk, corresponding to hyperbolic space itself. Therefore, the theory with the cutoff strictly removed has no degree of freedom.⁹ From the boundary perspective, the impossibility of completely removing the UV cutoff is the consequence of the fact that there is no standard quantum mechanical and nontrivial conformal field theory in one dimension.¹⁰ One can find theories with conformally invariant IR fixed points, such as the Sachdev-Ye-Kitaev (SYK) model, but the IR physics can never be completely decoupled from the UV, which is not conformally invariant. In other words, conformal invariance is only approximate and is broken above some UV scale that is impossible to get rid off.

⁶Of course, in the nonzero curvature models, one may set the units so that $L = 1$. It is sometimes convenient, for the physical discussion, to keep L explicitly.

⁷The precise statement is that typical metrics will have fractal boundaries, with probability one. Metrics with smooth boundaries, of course, exist but form a subset of measure zero in the space of metrics.

⁸The precise definition of the near-hyperbolic geometries, described by the so-called reparametrization *Ansatz*, is reviewed in Sec. II E 1.

⁹This is discussed in detail in [28]. In Lorentzian signature, a similar result is valid; see, in particular, [30].

¹⁰A simple argument is as follows. A conformal primary operator $\mathcal{O}$ of conformal dimension $\Delta > 0$ must have a two-point function $\langle \mathcal{O}(t)\mathcal{O}(0) \rangle$ proportional to $1/|t|^{2\Delta}$, which, thus, diverges when $t \rightarrow 0$. But in quantum mechanics, there is no singularity when one multiplies operators at the same point and, thus, no divergence in the two-point functions when $t \rightarrow 0$. To eliminate any possible confusion, let us note that Brownian motion in two dimensions, which may be viewed as one complex scalar field $z(t)$ in one dimension, is, according to Lévy’s theorem, conformally invariant with respect to target space conformal transformations $z \mapsto z'(z)$ but not with respect to one-dimensional conformal transformations that act on the time t .

The lesson from this discussion is that it is essential to formulate the theory for finite values of the parameters. The distance scales of the order of the curvature length scale are not decoupled, and the formulation must be UV complete from the bulk perspective.

D. UV-complete Jackiw-Teitelboim quantum gravity

On which basis can we construct a UV-complete formulation of JT gravity? Is there a unique “consistent” UV completion? Why can we not simply assume that the SYK model, for example, provides a satisfactory UV completion?

The axiom on which our work is based is that, in general, Euclidean two-dimensional quantum gravity is a diffeomorphism invariant theory of random metrics. In particular, Euclidean Jackiw-Teitelboim quantum gravity is a diffeomorphism invariant theory of random constant curvature metrics. Note that, in dimensions strictly greater than two, this minimalist metric point of view may very well be too naive, but in two dimensions, there is no reason to believe this. On the contrary, the classical results on Liouville theory and our own results on JT gravity offer a rather compelling physical picture and a rigorous UV-complete formulation of quantum gravity in two dimensions.

The metric axiom rules out the SYK model, or other similar models, since these are not models of random metrics. There is another independent argument against the idea that SYK could be a satisfactory UV completion of JT gravity, based on the following observations. First, the UV completion must be the same for all three versions of JT, negative, zero, or positive curvature, since the curvature length scale is irrelevant at very short distances. Second, the Schwarzian description emerges as a long-distance effective theory only in the case of the negative curvature theory. As will be made clear in the following, the other models have a very different long-distance behavior. Third, the existence of the IR Schwarzian description is the basic reason, if not the only reason, why the SYK model has been studied in relation with quantum gravity. It seems highly improbable that SYK could provide any insight into the UV structure of the zero or positive curvature models, in which the Schwarzian plays no role. Putting these three observations together, it seems extremely unlikely and unnatural that SYK could provide a UV completion of JT gravity. We shall argue that the Schwarzian theory is a universal long-distance description of many models of closed random loops in hyperbolic space, independently of any relation with quantum gravity.

A completely different, nonminimalist but physically well-grounded route to find UV completions of JT gravity is to consider that the theory is the IR limit of a higher-dimensional quantum gravity theory. This is natural when one studies the near-horizon limit of higher-dimensional near-extremal black holes. The higher-dimensional picture also suggests to consider general dilaton gravity models in

two dimensions, for which the bulk curvature is not fixed. From this point of view, JT gravity corresponds to a linear dilaton potential, in which case the dilaton plays the role of a Lagrange multiplier field imposing the constant curvature constraint. The higher-dimensional point of view is certainly worth exploring further, but this is far beyond the scope of the present paper. In any case, building a rigorous UV completion along these lines would require a detailed understanding of the higher-dimensional quantum gravity models, at a level which does not seem to be available at the moment. We believe that an interesting guiding principle to do this could be to ensure that a matching with the purely two-dimensional UV-complete point of view developed in our work should be made.

Even if we restrict ourselves strictly to dimension two and we follow the metric axiom, is the UV completion unique? Formally, the metric axiom implies that the theory is defined by a path integral over constant curvature metrics

$$\int Dg \delta(R[g] - 2\eta/L^2) e^{-\frac{\Lambda}{16\pi} A[g]}(\dots), \quad (1.1)$$

where Dg is a formal path integral measure over metrics, the functional δ function imposes the constraint of constant curvature, with $\eta = \pm 1$ or 0 depending on the version of JT one considers, $A[g]$ is the area associated with the metric g , and the dots indicate possible insertions of diffeomorphism invariant operators. The question of uniqueness amounts to asking whether there are several inequivalent ways to make sense of the formal path integral (1.1) or, equivalently, if there are several distinct measures, in the rigorous mathematical sense, on the space of constant curvature metrics modulo diffeomorphisms. The answer to this question is yes. An obvious way to understand that there are many interesting measures is that the quantum gravity theory could be coupled to a nontrivial quantum field theory (QFT), and integrating out the QFT degrees of freedom then produces a new interesting measure. Measures that probably cannot be interpreted in this simple way will also be constructed in [28]. To achieve uniqueness, one thus needs an additional guiding principle on top of the metric axiom. There are, in fact, at least three possible angles of attack.

One point of view is to start from the real-time quantization and the requirement that the Euclidean theory computes correctly suitably defined observables in the Hilbert space framework. In spirit, this approach is similar to the logic used in constructing the Euclidean path integral in quantum mechanics, which reproduces correctly the quantum mechanical predictions if and only if one uses the Wiener measure of the set of paths, which are, thus, Brownian almost surely. However, the rigorous real-time quantization of JT gravity remains an outstanding open problem at the microscopic, finite volume, level; see [31] for an interesting approach in the near-AdS framework. We

hope that our work on the Euclidean formulation, in the present paper and in [28], will actually provides important clues to achieve it in the future.

A second point of view is to assume that the measure Dg in the path integral (1.1) is the usual ultralocal DeWitt measure on the space of metrics and proceed from there. This is nonrigorous, since ultralocal measures are only formal objects, but experience with Liouville theory indicates that this approach is valuable in providing strong clues at to what the correct, rigorously defined measure is. We shall use this point of view in the companion paper [28].

A third point of view, that we shall follow in the present paper, is to use a lattice formulation, discretizing the space of constant curvature metrics and then taking a continuum limit. This approach provides a compelling physical picture and is unique, in the sense that the continuum limit of the lattice models is universal, independent of the details of the discretization procedure. In the continuum, it yields a privileged measure on the space of constant curvature metrics. We conjecture that this measure is the same as the one obtained through the formal procedure starting from the ultralocal DeWitt measure [28] and is also the correct measure for making the link with the real-time quantization of the models.

E. Comparison with the previous literature

To the best of our knowledge, possible microscopic, UV-complete, definitions of JT gravity have been discussed in only three papers, all focusing on the disk topology.

- (i) In [32], the authors take seriously the idea that the fluctuating boundary behaves as if it were the trajectory of a “boundary particle.” The path integral is then defined in the usual quantum mechanical way, using a quadratic action which describes the continuum limit of a closed random walk in target space, which is hyperbolic space for the negative curvature theory studied in [32]. This point of view has been used to investigate various properties of the models; see, e.g., [33].
- (ii) In [34], the authors use a very different starting point. They assume that, instead of arbitrary closed random walks, the fluctuating boundaries must be self-avoiding loops, that is to say, closed curves with no self-intersection. This is a highly nontrivial model of random paths, which has been much studied in the literature because it has many applications in statistical physics and in mathematics.
- (iii) The idea to consider that the boundary must be a non-self-intersecting curve was also used in [35], albeit with different results than in [34], but making the link with an interesting “bottom-up” approach in which the finite cutoff JT theory is postulated to correspond to a $T\bar{T}$ irrelevant deformation of the Schwarzian infinite cutoff description [36].

The picture that emerges from the investigations presented below is qualitatively and quantitatively different from these earlier proposals.

As we shall explain in detail below, a description in terms of random walks, as in [32], does not satisfy the metric axiom. The model, however, is quite interesting and does share some important qualitative features with the UV-complete formulation that we propose. The approaches in [34,35], on the other hand, do describe theories of random metrics, since one can easily show that to any self-avoiding loop is associated a unique metric on the disk. But the set of metrics so obtained is a subset of measure zero with respect to the relevant measure on the space of constant curvature metrics. Models based on self-avoiding loops may, thus, be seen as some kind of minisuperspace approximation.

As already mentioned, and in sharp contrast with the existing approaches based on the fluctuating boundary picture, one of the most surprising consequences of our analysis is that *the degrees of freedom of the theory are not in one-to-one correspondence with the shape of the boundary*. This is the consequence of the fact that to a given boundary curve may be associated several distinct metrics on the disk, that is to say, metrics that are not related by a disk diffeomorphism and are, thus, truly physically inequivalent. This fact alone shows that many ideas about the model must be rethought.

F. Plan of the paper

Section II is devoted to a detailed discussion of the space of constant curvature metrics on the disk. In Secs. II A and II B, a microscopic, discretized description is given in the flat case, using surfaces built by gluing polygons. Generating functions are defined in Sec. II C. A continuum description in terms of immersions, treating in parallel all three cases of zero, positive, and negative curvatures, is given in Sec. II D. Finally, in Sec. II E, we propose an overview of earlier approaches, in order to make the paper as self-contained as possible and also to be able to better explain and highlight the novelties brought about by our own contribution. We discuss the reparametrization *Ansatz* and the associated notion of “wiggling boundary,” the Schwarzian theory, the edge mode picture, and the existing approaches to the finite cutoff models.

In Sec. III, we develop the description of the model in terms of the boundary curves. The fact that JT quantum gravity can be approached from two very different perspectives, either based on random metrics or focusing on the boundary fluctuations, is one of its most interesting properties. The analysis of existing models indeed suggests that theories of random metrics, e.g., Liouville theory, and theories of random paths, e.g., self-avoiding paths or Brownian paths, have qualitatively distinct features and describe very different physics. After a general introduction in Sec. III A, we explain in Sec. III B that the correct model for the fluctuating boundary in JT is a model of *random*

self-overlapping curves. General properties of self-overlapping curves are explained in Sec. III C. In Secs. III C 1 and III C 2, we discuss simple criteria that self-overlapping curves must satisfy, based on the Whitney index and the notion of number of overlaps. A more comprehensive analysis, in terms of Blank cuts and Blank words, is given in Sec. III C 3. In Sec. III D, we introduce the fundamental notion of *multiplicity*. We illustrate this notion on the simplest example of the Milnor curve. The multiplicity of a self-overlapping curve counts the number of distinct disks it bounds or, equivalently, the number of physically distinguishable, that is to say, diffeomorphism inequivalent, constant curvature metrics associated with it. The fact that the multiplicity can be strictly greater than one implies that the degrees of freedom of JT quantum gravity are not in one-to-one correspondence with the shape of the boundary. In Sec. III E, we define the precise model of random self-avoiding polygons that is equivalent to the model of discretized random constant curvature metrics discussed in Sec. II. We also mention some possible generalizations for which self-overlapping curves can be open, and we very briefly discuss the case of the annulus or higher topologies. Finally, we give some preliminary information on the discretized formulation in negative and positive curvatures, highlighting the subtleties and differences with the flat case.

The goal of Sec. IV is twofold. First, we provide a pedagogical review of some ideas that play a central role in the study of known models of random metrics (the Liouville theory, Sec. IV A) and of random paths (Brownian, self-avoiding, convex) in Sec. IV B. Our aim is to introduce some useful concepts and technical tools and also to highlight the similarities but also some crucial differences between random metrics and random paths in the traditional setups, having in mind that in the case of JT the random metrics and random paths formulations are strictly equivalent. We then initiate the study of the self-overlapping polygon model in Sec. IV C, both in the case of the uniform measure and in the case of the nonuniform measure taking into account the multiplicity, which is relevant for JT. We prove the existence of the connective constant for both models, which implies, in particular, an exponential bound on the multiplicity. We describe some interesting observables that are specific to these models and explain some conjectures on some critical exponents and the asymptotic properties of the area distribution law.

In Sec. V, we introduce a Hermitian matrix model that does the appropriate counting of metrics in flat JT gravity, its large- N free energy coinciding with the generating functions introduced in Secs. II and III. The model is based on the idea of dually weighted graphs, that we review in Sec. V B. As explained in Secs. V B 3 and V C, some modifications with respect to the usual formulation must be introduced in order to be able to describe JT gravity. In particular, we construct the appropriate JT gravity

macroscopic loop operators and provide a matrix model formula for the multiplicity of a self-overlapping polygon.

Section VI focuses on the cases of nonzero curvature. At short-distance scales, the positive, negative, and zero curvature models coincide, but at long distances they display drastically different physics. The large-distance behaviors can be characterized by the asymptotics of the area probability densities, for which we present conjectures. For the negative curvature model, we explain the emergence of the Schwarzian description at long distances and make the link with the ballistic behavior of random path models in hyperbolic space. For the positive curvature model, we argue that large geometries will be ubiquitous due to the absence of an isoperimetric inequality. At the quantum level, this implies that the model exists only for positive or zero cosmological constant. We also discuss the lattice formulation of the nonzero curvature models and postulate a relation with a suitable analytic continuation of the matrix model discussed in Sec. V.

Finally, in Sec. VII A, we briefly summarize some of the main results of our work. We then discuss in Sec. VII B the consequences for the study of JT gravity on manifolds of arbitrary topology. We explain that the standard approach by Saad, Shenker, and Stanford, which applies in the Schwarzian limit, cannot be generalized in any straightforward way to the case of the microscopic, finite cutoff description. We close the paper in Sec. VII C by highlighting some interesting open questions that we would like to explore in some future research project.

II. CONSTANT CURVATURE METRICS ON THE DISK

We present a construction of the constant curvature metrics on the disk, mainly focusing on the discretized version in the flat case but also discussing the general case in the continuum. The discretized formulation is central, since it provides the basis for a rigorous microscopic formulation of the JT quantum gravity theory. We put our results into perspective by comparing them with existing approaches in the literature.

A. Discretized surfaces and discretized metrics

Let us start by briefly reviewing some well-known facts about the discretization of two-dimensional surfaces and metrics.

A discretized surface is built by gluing polygons along their edges. The construction is purely combinatorial in nature: The data are the set of polygons and the pairing (gluing) of the edges. These same data define a matrix model or Feynman graph as well. To see this, one may use one of two equivalent points of view. In the first, “direct” point of view, the p -gons correspond to faces of length p in the matrix model graph, and the polygon vertices and edges correspond to the Feynman graph vertices and edges,

respectively. In the second, “dual” point of view, the p -gons correspond to vertices of degree p of the Feynman graph and a gluing is associated to an edge joining the vertices associated with the glued polygons. The vertices of the direct point of view are mapped to the faces of the dual point of view and vice-versa.

Surfaces with k boundary components can be described as closed surfaces with k marked faces that may have an arbitrary length, the edges belonging to the marked faces being interpreted as forming the surface boundary. The marked faces may be removed altogether. The Euler identity reads

$$F - E + V = 2 - 2h - b, \quad (2.1)$$

where F is the number of unmarked faces, E the number of edges, V the number of vertices, h the genus, and b the number of boundary components (or marked faces).

Assuming that the unmarked faces are made of p -gons for a fixed value of p ,¹¹ and noting $r_1, \dots, r_b$ the lengths (number of edges) of the boundaries, we have

$$2E = pF + \sum_{i=1}^b r_i. \quad (2.2)$$

Using Euler’s (2.1), we get

$$2(2 - 2h - b) + \sum_{i=1}^b r_i = (2 - p)F + 2V. \quad (2.3)$$

A metric on a discretized surface is specified by introducing a length scale ℓ_0 . One decides that all the polygons building up the surface are flat and regular, with edge length ℓ_0 . Moreover, the gluing along the edges is smooth and flat. The cutoff ℓ_0 may be interpreted as the Planck length, but we shall prefer to call it the UV lattice spacing or UV lattice edge length. In this construction, the curvature is entirely localized on the vertices. The Ricci curvature in the vicinity of a bulk vertex (defined to be a vertex that does not belong to the boundary) of degree q , at which q p -gons are glued together, reads

$$R(x) = R_{p,q} \frac{\delta(x - x_V)}{\sqrt{g(x)}}, \quad (2.4)$$

where x is a local coordinate system around the vertex situated at $x = x_V$, g the determinant of the metric, and

$$R_{p,q} = \frac{2\pi}{p} (4 - (p - 2)(q - 2)). \quad (2.5)$$

¹¹The most general case, with different types of polygons, can be treated similarly, but we do not need it for our purposes.

The extrinsic curvature in the vicinity of a boundary vertex of degree q is similarly

$$k(s) = k_{p,q} \delta(s - s_V), \quad (2.6)$$

where s is the curvilinear coordinate along the boundary and

$$k_{p,q} = \frac{\pi}{p} (2 - (p - 2)(q - 2)). \quad (2.7)$$

Let V_q and v_q be the numbers of vertices of degree q in the bulk and on the boundary, respectively. We have [also using (2.2)]

$$2E = \sum_q q(V_q + v_q) = pF + \sum_{i=1}^b r_i. \quad (2.8)$$

Moreover, the total number of boundary vertices must match the total boundary length:

$$V_b = \sum_q v_q = \sum_{i=1}^b r_i. \quad (2.9)$$

Using (2.5) and (2.7), one can then check that the Gauss-Bonnet formula

$$\int d^2x \sqrt{g} R + 2 \oint ds k = 4\pi(2 - 2h - b) \quad (2.10)$$

is equivalent to the Euler formula (2.1) or, equivalently, to (2.3).

B. Discretized flat Jackiw-Teitelboim gravity

A straightforward application of the ideas reviewed in the previous subsection allows one to define a discretized version of Euclidean Jackiw-Teitelboim quantum gravity in the zero curvature case. Working this out explicitly and explaining some simple consequences is the subject of the present subsection.

Equation (2.5) shows that the constraint of zero curvature in the bulk implies that $2p = (p - 2)q$ for each bulk vertex. The solutions are $(p, q) = (3, 6)$, $(4, 4)$, or $(6, 3)$, corresponding to the three regular tessellations of the plane, made of triangles, squares, or hexagons. These three possibilities could be used in our construction, and they are all expected to yield the same theory in the continuum limit. For illustrative purposes, we mainly focus on quadrangulations. Using hexagons may present some technical advantages that will be mentioned in Sec. V.

We choose to orient the boundaries in such a way that the outer normal is directed to the right, the surface itself being situated to the left. With this convention, a positive extrinsic curvature means that the boundary is making a left turn and

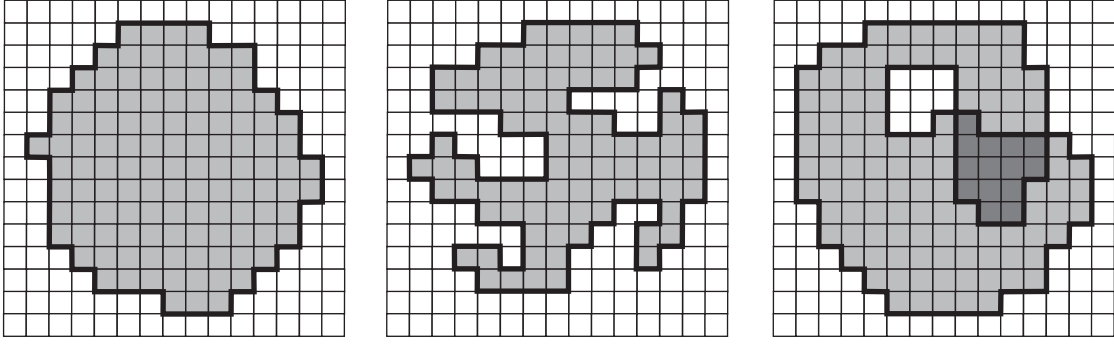


FIG. 1. Three examples of discretized disks drawn isometrically on a flat square lattice. The thick black lines represent the boundaries. The lattice tiles occupied by squares forming the disk, which are the faces of the associated fat graph in the matrix model direct representation, are shaded in gray. Dark gray tiles are occupied by two overlapping faces. Left inset: a disk with a smoothly wiggling boundary, akin to the “reparametrization *Ansatz*” configurations much studied in the literature; center inset: a disk whose boundary is a typical self-avoiding polygon; right inset: a disk with a twofold overlapping region (shaded in dark gray) and a self-intersecting boundary.

a negative extrinsic curvature that it is making a right turn. Precisely, with quadrangulations, formula (2.7) shows that the extrinsic curvature at the boundary vertices is

$$k_{4,q} = \begin{cases} \pi/2 & \text{if } q = 2(\text{left turn}), \\ 0 & \text{if } q = 3(\text{straight line}), \\ -\pi/2 & \text{if } q = 4(\text{right turn}). \end{cases} \quad (2.11)$$

We see that vertices of degree two, three, and four correspond to a left turn, going straight, or a right turn, respectively, and that the turns are made at a right angle. We restrict ourselves to vertices of degree two, three, or four on the boundary, since, intuitively, this is enough to obtain boundaries of arbitrary shapes in the continuum limit.¹²

Let us denote by V_B the number of bulk vertices and by v_L , v_S , and v_R the number of vertices of degree two, three, and four, respectively, on the boundary. Equations (2.2), (2.8), and (2.9) now take the form

$$\begin{aligned} 2E &= 4V_B + 2v_L + 3v_S + 4v_R \\ &= 4F + \sum_{i=1}^b r_i = 4F + v_L + v_S + v_R. \end{aligned} \quad (2.12)$$

In particular,

$$v_L + 2v_S + 3v_R = 4(F - V_B) \quad (2.13)$$

must be a multiple of four. Using (2.12), Euler’s identity (2.1) is found to be equivalent to

$$v_L - v_R = 4(2 - 2h - b). \quad (2.14)$$

¹²It may be sufficient to work with degrees two and four only. The case of degrees five or more will be briefly mentioned later.

Taking into account (2.11), this is, of course, also equivalent to the zero curvature Gauss-Bonnet formula $\oint k ds = 2\pi(2 - 2g - b)$. Note that, together with (2.13), we get

$$v_S = 2(F - V_B) - 2v_R - 2(2 - 2h - b). \quad (2.15)$$

The number of straight vertices must be even.

On the disk, Eq. (2.14) yields

$$v_L - v_R = 4. \quad (2.16)$$

This is a simple and useful constraint on the possible shapes of the boundary curve in JT gravity. We have derived it in the flat case, but we will see later that a version of this constraint is also valid in the positive and negative curvature cases; the difference is that in these latter cases it will not be a direct consequence of Gauss-Bonnet.

What do discretized flat disks look like? We have depicted in Fig. 1 three simple examples. All our drawings are made on an underlying square lattice. The basic building blocks of the lattice are called “tiles.” Each square face of a discretized disk is put on a lattice tile, building up a figure satisfying the constraints indicated above. The result is a graph for which all the faces, but the marked boundary, have length four, all the bulk vertices have degree four, and all the boundary vertices have degree two, three, or four.

A crucial feature is that *discretized disks may have overlaps*; see, for instance, the example depicted on the right inset in Fig. 1. As a consequence, *the boundary curves may have self-intersections*. This is an absolutely fundamental feature, which is also relevant in the positive and negative curvature cases.

Intuitively, in the continuum limit, discretized disks are arbitrary *distorted disks*, which are obtained by starting from a round disk and then allowing arbitrary smooth

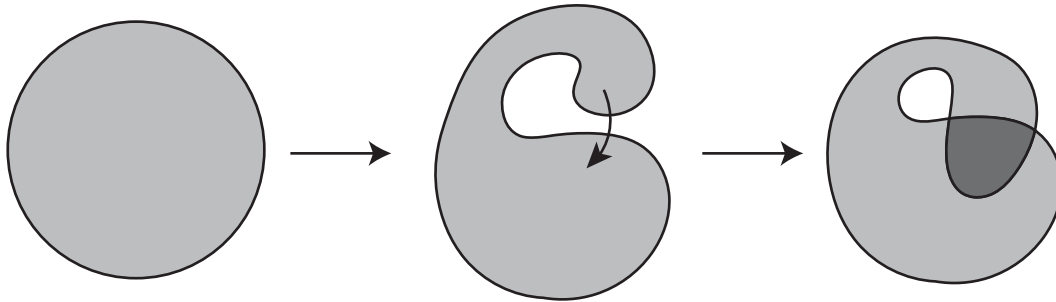


FIG. 2. A distorted disk with an overlapping region (right inset), obtained by smoothly deforming a round disk (left inset). The deformation is a regular homotopy between the trivial embedding of the circle in Euclidean space and a more general immersion. It can also be interpreted, by pullback, as a smooth deformation of the Euclidean round metric on the disk to a more general flat metric.

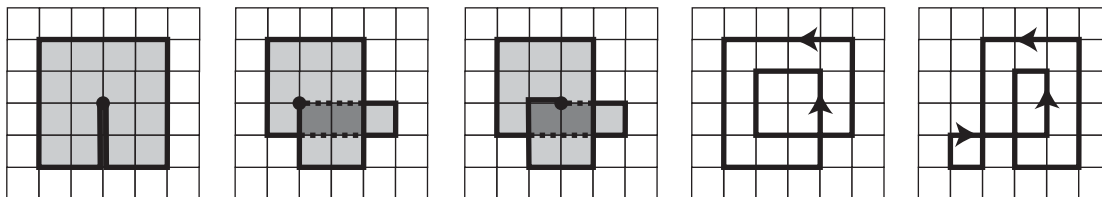


FIG. 3. Forbidden configurations corresponding to boundary vertices (black dots) of degrees 5, 6, and 7 (first, second, and third insets, respectively) and impossible boundary curves (right insets).

deformations, stretching and turning around. Under such deformations, some pieces of the original round disk may eventually cover other pieces, possibly multiple times. A deformation process, that yields a continuous configuration akin to the discretized disk depicted in the right inset in Fig. 1, is shown in Fig. 2.

Let us immediately dissipate a possible confusion. The graph associated with the disk is, of course, planar. As a consequence, it can be drawn on a plane in such a way that there is no overlap and, in particular, in such a way that the boundary has no self-intersection. This comes from a purely topological argument. The reason why we do get overlaps and self-intersections in our drawings is that we insist to depict the disks *in an isometric way*, respecting the distances defined by the metric associated with the discretization, as explained in Sec. II A. This is a very convenient and visual way to depict metrics of constant curvature, both in the discretized and in the continuum cases. This point of view will be further explored in Sec. II D.

To close this subsection, we have depicted five forbidden configurations in Fig. 3. The first three insets represent cases with boundary vertices of orders strictly greater than four: order $q = 5$, for which $k_{4,5} = -\pi$ (“U turn”); $q = 6$, for which $k_{4,6} = -3\pi/2$; and $q = 7$, for which $k_{4,6} = -2\pi$. Allowing for such configurations could be interesting, since they may allow for higher critical points in the model. In the last two insets on the right are depicted two candidate boundary curves that turn out to be forbidden, because they do not bound

distorted disks. It is indeed impossible to find discretized disks, as defined above, whose boundaries match these curves. The curve in the fourth inset has $v_L - v_R = 8$ and can, thus, be ruled out by the simple criterion (2.16). The curve in the rightmost inset satisfies the constraint $v_L - v_R = 4$ but is also ruled out because, by checking the position of the would-be distorted disk with respect to the boundary, one realizes that it would have to fill up the entire plane. These simple examples hint at the fact that the set of allowed boundary curves is constrained in a rather nontrivial and nonlocal way. This will be discussed in much more detail in Sec. III.

C. Generating functions

We denote by $W_{v,p}$ the number of distorted, flat, discretized disks with $F = p$ faces and $v = v_L + v_S + v_R$ boundary vertices. Note that (2.12) implies that v is even,¹³ and, thus, we set

$$v = 2n. \quad (2.17)$$

From (2.12) and (2.16), the total number of edges and bulk vertices, respectively, is

$$E = 2p + n, \quad V_B = p - n + 1. \quad (2.18)$$

¹³Actually, any closed path on a square lattice has even length, even paths that do not bound a distorted disks.

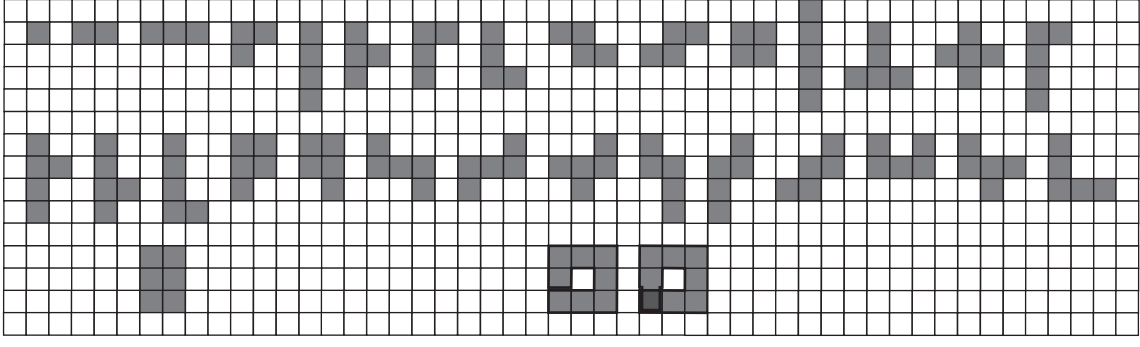


FIG. 4. All the distorted disks, modulo lattice rotations, having five faces or less or a boundary length 10 or less. In the lower part of the figure are also depicted the two simplest distorted disks whose boundaries are not self-avoiding polygons (contributing to the partition function at the order of $g^{18}t^8$) or that do have an overlap (contributing to the partition function at the order of $g^{20}t^9$).

To define $W_{2n,p}$ precisely, we identify two disks that are related by a lattice translation. From the point of view of the metric or gravitational interpretation, it would also be natural to identify configurations related by lattice rotations and reflections, since these transformations clearly do not change the metric. However, for various reasons, this is not the most convenient choice to make; see, in particular, Sec. V. We thus decide to count as distinct nonidentical configurations related by lattice rotations or reflections. For generic configurations, and, thus, also in the continuum limit, this simply amounts to multiplying the generating functions by a factor of $4 \times 2 = 8$. One may then rescale the generating functions by a global factor of $1/8$, but we will not do this, since it is more natural to work with coefficients $W_{2n,p}$ that are integers.

All the distorted disks with low values of n and p are depicted in Fig. 4. With our counting convention, the first configuration in the upper-left corner is counted one time, $W_{4,1} = 1$, the second two times, $W_{6,2} = 2$, the third two times, and the fourth four times, $W_{8,3} = 6$, etc. The numbers $W_{2n,p}$ are conveniently encoded in the generating functions

$$W(t, g) = \sum_{n,p \geq 1} W_{2n,p} t^p g^{2n}. \quad (2.19)$$

One gets, from Fig. 4,

$$W(t, g) = g^4 t + 2g^6 t^2 + 6g^8 t^3 + (18g^{10} + g^8) t^4 + (8g^{10} + 55g^{12}) t^5 + 2g^{10} t^6 + O(g^{12} t^6). \quad (2.20)$$

This low order counting is not very interesting, because for low values of n and p the boundaries are standard self-avoiding polygons (SAPs) [37]. The counting starts to deviate from the SAP case at order $g^{18}t^8$; see Fig. 4.

The parameters t and g are related to the bulk and boundary bare cosmological constants Λ_0 and λ_0 by $t = e^{-\Lambda_0 \ell_0^2 / (16\pi)}$ and $g = e^{-\lambda_0 \ell_0}$, where ℓ_0 is the UV lattice edge

length.¹⁴ The generating functions at fixed boundary length and fixed area, respectively, are

$$W_{2n}(t) = \sum_{p \geq 1} W_{2n,p} t^p, \quad \tilde{W}_p(g) = \sum_{n \geq 1} W_{2n,p} g^{2n}. \quad (2.21)$$

The full generating function (2.19) can be expressed as

$$W(t, g) = \sum_{n \geq 1} W_{2n}(t) g^{2n} = \sum_{p \geq 1} \tilde{W}_p(g) t^p. \quad (2.22)$$

For a given boundary length $2n$, the area is bounded above by $[n^2/4]$, which is $n^2/4$ or $(n^2 - 1)/4$ depending on whether n is even or odd. This is the isoperimetric inequality on the square lattice. For a given area p , the largest possible boundary length is $2(p + 1)$, which, by using (2.18), is obtained when $V_B = 0$, i.e., when the disk is treelike, or branched polymerlike, with no bulk structure. This implies that the generating functions at fixed length and fixed area are polynomials, with

$$\deg W_{2n}(t) = [n^2/4], \quad \deg \tilde{W}_p(g) = 2(p + 1). \quad (2.23)$$

In particular,

$$w_{2n} = W_{2n}(1) = \sum_{p \geq 1} W_{2n,p}, \quad \tilde{w}_p = \tilde{W}_p(1) = \sum_{n \geq 1} W_{2n,p} \quad (2.24)$$

are finite and count the total number of distorted disks with fixed boundary length $2n$ and fixed area p , respectively. From (2.20), one finds

¹⁴We use the traditional convention in JT gravity of introducing a factor of $\frac{1}{16\pi}$ in the definition of the bulk cosmological constant.

$$\begin{aligned}
w_2 = 0, \quad w_4 = 1, \quad w_6 = 2, \quad w_8 = 7, \quad w_{10} = 28, \\
\tilde{w}_1 = 1, \quad \tilde{w}_2 = 2, \quad \tilde{w}_3 = 6, \quad \tilde{w}_4 = 19, \quad \tilde{w}_5 = 63.
\end{aligned}
\tag{2.25}$$

The functions W , W_{2n} , or $\tilde{W}_p$ are *a priori* formal generating functions encoding the numbers $W_{2n,p}$. Their radii of convergence and the possibility to define a continuum limit will depend on the asymptotic behavior at large n and p . These aspects will be discussed in Sec. IV.

D. Constant curvature metrics on the disk in the continuum

The above discussion has a natural extension in the continuum that we present in this subsection. The cases of zero, positive, or negative curvatures are treated in parallel.

1. Immersions and pullback metrics

Let us first recall the definition of an immersion and a few simple facts about the concept.

Definition 2.1. An immersion $F: \mathcal{M} \rightarrow \mathcal{N}$ between two differentiable manifolds $\mathcal{M}$ and $\mathcal{N}$, $\dim \mathcal{N} \geq \dim \mathcal{M}$, is a differentiable function such that dF has maximum rank everywhere, that is to say, $\text{rk} dF|_P = \dim \mathcal{M}$ for any point P of $\mathcal{M}$.

In local coordinate systems (x^μ) on $\mathcal{M}$ and (y^a) on $\mathcal{N}$, an immersion F is described by functions $y^a(x)$ such that the matrix $\partial y^a / \partial x^\mu$ has rank $\dim \mathcal{M}$ everywhere.

We shall use this in the following way. If $F: \mathcal{M} \rightarrow \mathcal{N}$ is an immersion and $\mathcal{N}$ is endowed with a metric δ , then there is a naturally associated metric $g = F^* \delta$ on $\mathcal{M}$, called the pullback metric, given in local coordinates by

$$g_{\mu\nu}(P) = \frac{\partial y^a}{\partial x^\mu} \frac{\partial y^b}{\partial x^\nu} \delta_{ab}(F(P)). \tag{2.26}$$

The fact that the metric tensor g is positive definite follows from the facts that δ is positive definite and that $\partial y^a / \partial x^\mu$ does not have a zero eigenvalue. If we waive the condition that F is an immersion, then the symmetric tensor g defined by (2.26) will still be positive, but it will be degenerate (noninvertible) at the points where dF is not invertible. Thus, it will not be a metric in the usual sense.

An immersion $F: \mathcal{M} \rightarrow \mathcal{N}$ is a local diffeomorphism between $\mathcal{M}$ and its image $F(\mathcal{M})$. When an immersion is injective, one says that the immersion is an embedding. There is then a global identification between the spaces $\mathcal{M}$ and $F(\mathcal{M})$. Of course, a generic immersion is not an embedding.

The simplest case to consider is that of immersions $F: S^1 \rightarrow \mathbb{R}^2$ of the circle into the plane. The image $\text{Im } F = \gamma$ is a closed curve. For a general immersion F , γ has self-intersections. The immersion is an embedding if and only if γ has no self-intersection. Another case, which is very

important for our purposes, is that of immersions $F: \mathcal{D} \rightarrow \mathbb{R}^2$ from the disk to the plane. The discretized disks in Sec. II B are the discretized versions of the images $F(\mathcal{D})$ of immersions F . The associated metrics are the pullback metrics (2.26), where δ is the Euclidean metric. This idea yields an identification between the space of constant curvature metrics and appropriate quotients of the space of immersions, that we are going to work out below.

Consistently with the convention we have already used, we always denote by $\mathcal{D}$ the disk manifold. This is the “source” space, on which the quantum gravity theory is defined. It can be described by the usual polar coordinates (ρ, θ) , $0 \leq \rho < 1$, or complex coordinates $z = \rho e^{i\theta}$,

$$\mathcal{D} = \{z \in \mathbb{C} | |z| < 1\}. \tag{2.27}$$

The boundary is $S^1 = \partial \mathcal{D} = \{z \in \mathbb{C} | |z| = 1\}$, and the closed disk is denoted by $\bar{\mathcal{D}} = \mathcal{D} \cup \partial \mathcal{D}$.

2. The space of zero curvature metrics on the disk, I

Let $E^2 = (\mathbb{R}^2, \delta)$ be the two-dimensional Euclidean plane, where δ is the usual flat metric. In Cartesian coordinates (x, y) and complex coordinates $w = x + iy$,

$$\delta = dx^2 + dy^2 = |dw|^2. \tag{2.28}$$

A flat two-dimensional disk $(\mathcal{D}, g)$, where g is a flat metric on $\mathcal{D}$, is locally isometric to E^2 . This is equivalent to saying that there exists an immersion

$$F: \mathcal{D} \rightarrow \mathbb{R}^2 \tag{2.29}$$

such that

$$g = F^* \delta. \tag{2.30}$$

Let us denote by $\text{Imm}(\mathcal{D}, \mathbb{R}^2)$ the space of immersions $F: \mathcal{D} \rightarrow \mathbb{R}^2$. This is a bit imprecise, because we do not specify at this stage the boundary behavior of the allowed immersions; the discussion of this important point is postponed to the next subsection.

We can now make precise the notion of “distorted disk” that we have already encountered above.

Definition 2.2. A distorted disk is an immersion belonging to $\text{Imm}(\mathcal{D}, \mathbb{R}^2)$.

On $\text{Imm}(\mathcal{D}, \mathbb{R}^2)$ act the groups $\text{Iso}(E^2) = \text{ISO}(2) = \text{T}(2) \rtimes \text{O}(2)$ of isometries of the Euclidean plane, on the left, and the group $\text{Diff}_+(\bar{\mathcal{D}})$ of orientation-preserving diffeomorphisms of the closed disk, on the right. We consider only orientation-preserving diffeomorphisms for convenience, but including complex conjugation $z \mapsto \bar{z}$ is, of course, straightforward. By definition, the group $\text{Diff}_+(\bar{\mathcal{D}})$ includes boundary reparametrizations. At the infinitesimal level, it is generated by vector fields on $\mathcal{D}$ that

are tangent to the boundary at the boundary, as is standard for a manifold with boundaries.

If we note $g_F = F^* \delta$ the metric associated with the immersion F as in Eq. (2.30), then, by definition of the isometry group, $g_F = g_{\psi \circ F}$ if $\psi \in \text{Iso}(\mathbb{E}^2)$. Moreover, if $\varphi \in \text{Diff}_+(\bar{\mathcal{D}})$, then $g_F = \varphi^* g_{F \circ \varphi^{-1}}$; i.e., the metrics g_F and $g_{F \circ \varphi^{-1}}$ are related by the usual action of the group of diffeomorphisms of $\bar{\mathcal{D}}$. They are, thus, identified. Overall, we have shown that the space of flat metrics on the disk is

$$\text{Met}^0(\mathcal{D}) = \text{Iso}(\mathbb{E}^2) \backslash \text{Imm}(\mathcal{D}, \mathbb{R}^2) / \text{Diff}_+(\bar{\mathcal{D}}). \quad (2.31)$$

This can be further simplified by noting that, by acting with $\text{Diff}_+(\bar{\mathcal{D}})$ together with the reflection $F \mapsto \bar{F}$, which is an element of $\text{Iso}(\mathbb{E}^2)$, we can always find a holomorphic representative F , with F' never vanishing, for each equivalence class of immersions in the quotient space on the right-hand side of (2.31). For a holomorphic immersion, the pullback metric (2.30) is

$$g_F = |F'(z)|^2 |dz|^2. \quad (2.32)$$

This result is a special case of a more general theorem that ensures that any metric on $\mathcal{D}$, not necessarily flat, can always be put in the so-called conformal gauge,

$$g = e^{2\Sigma} |dz|^2, \quad (2.33)$$

for a conformal factor (“Liouville field”) Σ . The flatness condition is equivalent to the fact that Σ is harmonic on the disk and is, thus, the real part of a holomorphic function H , $\Sigma = \text{Re}H$. If $F' = e^H$,

$$e^\Sigma = |F'(z)|, \quad (2.34)$$

and, thus, Eq. (2.33) is equivalent to (2.32). Let us denote by $\text{hImm}(\mathcal{D})$ the set of holomorphic immersions $F: \mathcal{D} \rightarrow \mathbb{C}$, i.e., the set of holomorphic functions on the disk such that $F'(z) \neq 0$ for all $z \in \mathcal{D}$.¹⁵ On $\text{hImm}(\mathcal{D})$ act the group $\text{Iso}_+(\mathbb{E}^2) = \text{ISO}_+(2) = \text{T}(2) \rtimes \text{SO}(2)$ of orientation-preserving isometries of the Euclidean plane on the left and the group $\text{Aut}(\mathcal{D})$ of holomorphic diffeomorphisms of the disk on the right. This group is the remnant of the full group of diffeomorphisms in conformal gauge (2.33). It is generated by the conformal Killing vectors on the disk and is isomorphic to $\text{PSL}(2, \mathbb{R})$. We, thus, have

$$\begin{aligned} \text{Met}^0(\mathcal{D}) &= \text{Iso}_+(\mathbb{E}^2) \backslash \text{hImm}(\mathcal{D}) / \text{Aut}(\mathcal{D}) \\ &= \text{ISO}_+(2) \backslash \text{hImm}(\mathcal{D}) / \text{PSL}(2, \mathbb{R}). \end{aligned} \quad (2.35)$$

¹⁵As for $\text{Imm}(\mathcal{D}, \mathbb{R}^2)$, we let unspecified the behavior at the boundary. This is discussed in the next subsection.

The explicit actions of the groups on the disk and on the Euclidean plane are given by

$$\text{PSL}(2, \mathbb{R}): z' = e^{i\alpha} \frac{z - z_0}{1 - \bar{z}_0 z}, \quad |z_0| < 1, \quad (z, z') \in \bar{\mathcal{D}}^2, \quad (2.36)$$

$$\begin{aligned} \text{ISO}_+(2): w' &= e^{i\alpha}(w - w_0), \quad w_0 \in \mathbb{C}, \\ (w, w') &\in \mathbb{C}^2 = \mathbb{E}^2 \times \mathbb{E}^2. \end{aligned} \quad (2.37)$$

Conformal gauge will be thoroughly studied in a companion paper [28].

Remarks.

- (i) Both the group $\text{Diff}_+(\bar{\mathcal{D}})$ of diffeomorphisms of the closed disk and the group $\text{Iso}(\mathbb{E}^2)$ of isometries of the Euclidean plane must be considered as gauge symmetries. This statement is equivalent to the presentation (2.31) of the space of metrics as a quotient space. It is useful to understand that the origin of these two gauge symmetries is very different. On the one hand, the gauge symmetry $\text{Diff}_+(\bar{\mathcal{D}})$ is consubstantial with our problem: It is the fundamental diffeomorphism gauge symmetry present in any gravity theory.¹⁶ On the other hand, the gauge symmetry $\text{Iso}(\mathbb{E}^2)$ appears only because we use immersions to describe the metrics. It is a redundancy that enters via the formula (2.30). The same can be said after the partial gauge fixing that yields (2.35): The groups $\text{PSL}(2, \mathbb{R})$ and $\text{ISO}_+(2)$ are both gauge symmetries, but $\text{PSL}(2, \mathbb{R})$ is always present in conformal gauge for quantum gravity on the disk, whereas $\text{ISO}_+(2)$ appears because we parametrize the metrics in terms of holomorphic immersions.
- (ii) In the spirit of what we have done in Sec. II B, where configurations obtained by performing lattice rotations were counted as inequivalent, we could quotient only by the group of translations $\text{T}(2)$ instead of the full isometry group of $\mathbb{E}^2$. This kind of modification is largely irrelevant. It simply introduces trivial factors of 2π in the definition of the partition function. The important point is to take into account the noncompact pieces of the gauge symmetries in order to work with well-defined path integrals.

3. The space of zero curvature metrics on the disk, II

The discussion above does not address the problem of the boundary properties of the disk metrics. There are many mathematically natural possibilities and, thus, a large

¹⁶Note that so-called “large” diffeomorphisms, that act non-trivially on the boundary, must be gauged because we are considering the theory on finite geometries (i.e., at finite cutoff). This point is particularly important in the continuum approach; see [28].

number of *a priori* interesting spaces of flat metrics on $\mathcal{D}$. A precise definition of the Euclidean quantum JT gravity theory requires one to single out one of these spaces and to define a probability measure on it.

A seemingly natural choice (which has been made implicitly in virtually all studies on JT gravity so far; see, in particular, the review in Sec. II E) is to impose, in conformal gauge, Eq. (2.33), that the boundary Liouville field $\sigma = \Sigma|_{\partial\mathcal{D}}$ is continuous.¹⁷ With this choice, the boundary length

$$\ell = \int_0^{2\pi} e^\sigma d\theta \quad (2.38)$$

is well defined and one can consider the space $\text{Met}_\ell^0(\mathcal{D})$ of flat metrics on the disk with fixed boundary length ℓ .

An important consequence of the analysis presented in our work, consistently with the discussion in [16], is that this smoothness condition on the disk boundary is not well founded from the point of view of the microscopic, UV-complete definition of the models.

To clearly appreciate the nature of the problem, it is useful to use a simple analogy with the standard functional integral formulation of quantum mechanics. In this case, one wishes to define a sum over “paths.” But what is exactly the space of paths one must consider? There are two physically motivated approaches to answer this question. The first is to construct a path integral that reproduces the results obtained from canonical quantization in the traditional operator formalism of quantum mechanics. The second is to start from a discretized model, in which the sum over path is a well-defined finite problem, and look for a critical point that allows one to define a continuum limit. Both procedures yield the Wiener measure on the set of *continuous* paths, describing Brownian motion. The typical Brownian paths turn out to be almost surely nowhere differentiable and have Hausdorff dimension two. If we had started from the seemingly natural but *ad hoc* idea that the paths had to be differentiable, or rectifiable, we would not have been able to obtain the correct theory.

We propose that exactly the same logic should be used in the case of JT gravity. One could try to set up a rigorous canonical quantization formalism in real time and then explore its consequences for the Euclidean path integral formulation. This is very interesting but not in the spirit of the present paper. We favor the discretized formulation, as defined in Secs. II B and II C. The continuum limit should then yield precise predictions for the boundary properties of the typical flat disk metrics. Alternatively, one could start from a formal continuum approach, based on the formal ultralocal measure on the space of metrics, and try to make sense of it; see [16,28].

¹⁷Note that the bulk Liouville field is always infinitely differentiable, because it must be a harmonic function.

We conjecture that all these methods (canonical quantization, discrete formulation, and continuous formulation) yield the same quantum gravity theory. The results announced in [16], and confirmed in [29] by studying a particular limit of JT gravity coupled to conformal matter, strongly suggest that the resulting disk metrics are highly singular on the disk boundary. In conformal gauge, the boundary Liouville field is predicted to be a random distribution, not a random function.¹⁸ The resulting typical immersed boundary curves are fractal curves of Hausdorff dimension $d_H > 1$.¹⁹ In particular, the usual geometrical boundary length defined by (2.38) is infinite for such metrics, but it is possible to define a notion of renormalized length, or “quantum” length β_q .²⁰ The explicit definition of β_q is given in Sec. IV in the discretized formulation and in [16,28] in the continuum formulation. It is similar to the diffusion time in a Brownian process, which provides a notion of “quantum length” for a Brownian path. An important qualitative feature is that the quantum length does not have the dimension of a geometrical length. Instead, β_q^{1/d_H} has the dimension of length. The fact that the renormalized length β_q exists is a desirable feature for physics, since it allows one to define canonical and microcanonical ensembles; see below.

A similar discussion applies to the spaces of constant positive and negative curvature metrics that we are going to discuss below. The relevant metrics and immersions are highly singular on the boundary, with the same UV properties as in the flat case.

4. The spaces of constant positive and negative curvature metrics on the disk

The discussion in Sec. II D 2 can be repeated, with straightforward modifications, to describe the spaces of constant positive and negative curvature metrics. The difference is that the Euclidean space, which plays the role of the canonical target space in the case of the flat metrics, is replaced by the round two-sphere S^2 for positive curvature and by the hyperbolic space H^2 for negative curvature.

Let S^2 be the round two-sphere, with canonical constant positive curvature metric $R = +2$ given by

¹⁸This is in sharp contrast with the traditional one-dimensional quantum mechanical variables, which are random continuous functions. This behavior is closer to that of a quantum field in dimension two.

¹⁹For the pure JT gravity model, which corresponds to the discretized model discussed in Sec. II B, the Hausdorff dimension of the disk boundary is predicted to be two [16]. This is the same value as for Brownian paths, even though the disk boundary in JT is not Brownian; see the discussion in Secs. III and IV.

²⁰We use the notation β_q for the quantum length of the boundary, in order to avoid confusion with another parameter β , that we shall denote by β_S , which is traditionally used in the Schwarzian limit; see Eq. (2.50).

$$\delta^+ = \frac{4|dw|^2}{(1 + |w|^2)^2} \quad (2.39)$$

in the usual projective coordinates $w \in \bar{\mathbb{C}}$. Let H^2 be the hyperbolic space, with canonical constant negative curvature metric $R = -2$ given by

$$\delta^- = \frac{4|dw|^2}{(1 - |w|^2)^2} \quad (2.40)$$

in the usual Poincaré disk coordinates $|w| < 1$. Let us note that H^2 is topologically a disk, endowed with the metric δ^- , so we can write $H^2 = (\mathcal{D}, \delta^-)$.

Let $\text{Met}^\pm(\mathcal{D})$ be the spaces of constant positive or negative curvature metrics, $R = \pm 2$, on $\mathcal{D}$, with suitable boundary properties, as described in Sec. II D 3. All such metrics can be written as

$$g = F^* \delta^\pm \quad (2.41)$$

for some immersion $F: \mathcal{D} \rightarrow S^2$ or $F: \mathcal{D} \rightarrow H^2$. Let $\text{Imm}(\mathcal{D}, S^2)$ and $\text{Imm}(\mathcal{D}, H^2)$ be the associated spaces of immersions of $\mathcal{D}$ into S^2 or H^2 . By following the same logic that yields Eq. (2.31) in the flat case, we obtain

$$\text{Met}^+(\mathcal{D}) = \text{Iso}(S^2) \backslash \text{Imm}(\mathcal{D}, S^2) / \text{Diff}_+(\bar{\mathcal{D}}), \quad (2.42)$$

$$\text{Met}^-(\mathcal{D}) = \text{Iso}(H^2) \backslash \text{Imm}(\mathcal{D}, H^2) / \text{Diff}_+(\bar{\mathcal{D}}). \quad (2.43)$$

Going to conformal gauge, Eq. (2.33), we can write the metrics in terms of holomorphic functions F :

$$e^\sigma = \frac{2|F'(z)|}{1 \pm |F(z)|^2} \quad \text{for } R = \pm 2. \quad (2.44)$$

In the case of positive or negative curvature, the holomorphic functions are from $\mathcal{D}$ to $\bar{\mathbb{C}}$ or from $\mathcal{D}$ to $\mathcal{D}$, respectively. This yields

$$\begin{aligned} \text{Met}^+(\mathcal{D}) &= \text{Iso}_+(S^2) \backslash \text{hImm}(\mathcal{D}, \bar{\mathbb{C}}) / \text{Aut}(\mathcal{D}) \\ &= \text{SO}(3) \backslash \text{hImm}(\mathcal{D}, \bar{\mathbb{C}}) / \text{PSL}(2, \mathbb{R}), \end{aligned} \quad (2.45)$$

$$\begin{aligned} \text{Met}^-(\mathcal{D}) &= \text{Iso}_+(H^2) \backslash \text{hImm}(\mathcal{D}, \mathcal{D}) / \text{Aut}(\mathcal{D}) \\ &= \text{PSL}(2, \mathbb{R}) \backslash \text{hImm}(\mathcal{D}, \mathcal{D}) / \text{PSL}(2, \mathbb{R}). \end{aligned} \quad (2.46)$$

The automorphism group $\text{PSL}(2, \mathbb{R})$ acting on the right is as in (2.36), whereas the isometry groups act on the left as

$$\begin{aligned} \text{SO}(3): w' &= e^{i\alpha} \frac{w - w_0}{1 + \bar{w}_0 w}, \quad w_0 \in \bar{\mathbb{C}}, \\ (w, w') &\in \bar{\mathbb{C}}^2 = S^2 \times S^2, \end{aligned} \quad (2.47)$$

$$\text{PSL}(2, \mathbb{R}): w' = e^{i\alpha} \frac{w - w_0}{1 - \bar{w}_0 w}, \quad |w_0| < 1,$$

$$(w, w') \in \mathcal{D}^2 = H^2 \times H^2. \quad (2.48)$$

Note that, following remark (i) at the end of Sec. II D 2, the two $\text{PSL}(2, \mathbb{R})$ groups, acting on the right and on the left in (2.46), are of totally different origins, even though they coincide as abstract groups; this is an important point to keep in mind—see [28] for further discussion.

5. Quantum JT gravity on the disk

The disk partition function of quantum JT gravity in zero, positive, or negative curvature, in the canonical ensemble, is expressed as a path integral

$$W^{0,\pm}(\beta_q, \Lambda) = \int_{\text{Met}_{\beta_q}^{0,\pm}(\mathcal{D})} d\mu(g) e^{-\frac{\Lambda}{16\pi} A[g]} \quad (2.49)$$

over the spaces $\text{Met}_{\beta_q}^{0,\pm}(\mathcal{D})$ of constant curvature metrics on the disk with a fixed quantum length β_q .²¹ These spaces are subspaces of the spaces $\text{Met}^{0,\pm}(\mathcal{D})$ defined above, for which we impose the constraint of fixed quantum boundary length; see [16,28]. In (2.49), $A[g]$ denotes the area of the disk for the metric g , and Λ is the cosmological constant. The expression (2.49) is, of course, formal. The challenge is to define measures $d\mu(g)$ over the spaces $\text{Met}_{\beta_q}^{0,\pm}(\mathcal{D})$. As we have already mentioned, an approach for constructing these measures is to start from the discretized formulation and to take an appropriate continuum limit of the generating functions defined in Sec. II C; see Sec. IV. A direct continuum approach is also proposed in [16,28].

Remark on stability. The partition functions (2.49) will always be well defined for positive or zero bulk cosmological constants, $\Lambda \geq 0$. The situation for $\Lambda < 0$ is much more subtle.

At the classical level, we usually consider metrics that extend smoothly to the boundary. In zero and negative curvature, the isoperimetric inequalities then imply that the area is bounded above for fixed boundary length. In this sense, the models are, thus, classically stable. However, in positive curvature, there is no isoperimetric inequality and the model is classically unstable for $\Lambda < 0$; see Sec. VI C.

At the quantum level, the situation is much more subtle. Even in zero or negative curvature, the isoperimetric inequalities do not apply, since the boundary is a fractal curve of infinite length. There are, thus, configurations of infinite area that have a finite quantum length. As a consequence, the classical action $\frac{\Lambda}{16\pi} A$ is not bounded

²¹Throughout the paper, we use the letter W to denote quantum gravity partition functions in fixed topology, instead of Z , to emphasize the fact that they are continuum limits of generating functions which are themselves interpreted as the sum over connected Feynman graphs in suitable matrix models.

from below when $\Lambda < 0$. But this does not imply that the models are quantum mechanically ill defined, since the entropy of the configurations must be taken into account.

We conjecture the following behavior: (i) The negative curvature theory is well defined for all values, positive or negative, of the cosmological constant Λ . (ii) The zero curvature theory is well defined only for a range $\Lambda > \Lambda_c$ of cosmological constants, where the critical value Λ_c is strictly negative. (iii) The positive curvature theory is well-defined only for $\Lambda \geq 0$. This conjectured behavior will be further discussed in Sec. VI.

E. Discussion of the existing literature

Now that the stage is set, at least at the conceptual level, it is useful to review some of the basic ideas that underlie the vast existing literature on JT gravity. This subsection is written mainly for nonexperts in the field of SYK and JT gravity. Emphasis is put on important differences between standard approaches and our own work, in order to put into perspective the new point of view that we are introducing.

By far the most popular framework to study the negative curvature model is based on the so-called reparametrization *Ansatz* and the Schwarzian action. This description is usually assumed to become correct in a particular near-hyperbolic limit²²

$$\ell \rightarrow \infty, \quad \Lambda \rightarrow -\infty, \quad \frac{2\ell}{|\Lambda|} = \beta_S \text{ fixed.} \quad (2.50)$$

As we shall argue, the Schwarzian theory is an effective description of JT gravity, that emerges in the limit $\Lambda \rightarrow -\infty$ and is valid on distance scales much larger than the curvature length scale. The very definition of the limit in Eq. (2.50) is “effective,” since it refers to a smooth boundary length parameter ℓ that does not exist at the microscopic level. The relation between the macroscopic length ℓ and the microscopic parameters will be discussed in Sec. VI.

An interesting attempt at a microscopic definition of JT gravity was proposed by Kitaev and Suh in [32]. The idea is to assume that the disk boundaries are random Brownian loops. The resulting path integral is then related to the quantum mechanics of a charged particle in hyperbolic space. As is clear from the discussion in Sec. II B, the vast majority of the configurations introduced in this way are not associated with disk metrics, because a typical random loop does not bound a distorted disk. Therefore, the Kitaev-Suh model is not a model of random metrics.

²²The factor of 2 in the definition of β_S is inserted to match with the standard definition in the literature. Note that the cosmological constant Λ is related to the boundary value Φ_b of the dilaton field in the usual formulation of JT gravity by $|\Lambda| = -\Lambda = 2\Phi_b$.

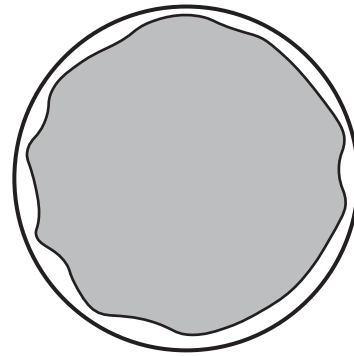


FIG. 5. A reparametrization *Ansatz*, or near-hyperbolic geometry, as defined in [20]. The open disk with the thick boundary is hyperbolic space $H^2 = (\mathcal{D}, \delta^-)$ in the Poincaré disk representation, Eq. (2.40). The geometry is defined by the cut-out region, in gray, which is itself determined by the shape of a gently wiggling boundary curve. In our language, the cut-out regions associated with reparametrization *Ansätze* are distorted disks associated with a very special class of embeddings of the source disk, on which the quantum gravity theory is defined, into hyperbolic space.

Other proposed microscopic descriptions are based on using self-avoiding loops [34,35]. This captures only a very small subset of the allowed configurations, actually a subset of measure zero. These approaches are, thus, in the same spirit as the minisuperspace approximations to quantum gravity. They are unlikely to provide a good approximation to the microscopics of JT gravity, even qualitatively. For instance, the Hausdorff dimension of self-avoiding loops counted with the uniform measure, as in [34], is $4/3$, to be compared to the correct value 2 for pure JT. In the approach of [35], the boundaries are smooth. A picture based on “edge modes” or broken diffeomorphisms has also been studied; it is another minisuperspace approximation, closely related to the other approaches.

1. Reparametrization Ansatz

The reparametrization *Ansatz* is relevant only in the case of the negative curvature theory.

Brief review. The authors of [20] considered the possibility of defining negative curvature geometries by “cutting out pieces of hyperbolic spaces” with particular shapes, as they phrase it. A typical example is depicted in Fig. 5. The piece that is cut out is defined by the shape of its boundary. The boundary curve is assumed to be smooth, of length ℓ , and is described as follows. We introduce the arc-length coordinate s along the curve and polar coordinates (r, ϕ) , $w = re^{i\phi}$, such that the hyperbolic metric (2.40) reads

$$\delta^- = \frac{4(dr^2 + r^2 d\phi^2)}{(1 - r^2)^2}. \quad (2.51)$$

We parametrize

$$r(s) = 1 - \frac{1}{\ell} h(\vartheta), \quad \phi(s) = f(\vartheta), \quad \text{where } \vartheta = \frac{2\pi s}{\ell} \quad (2.52)$$

is a 2π -periodic variable. The length ℓ is considered to be large.²³ The function f is assumed to be smooth, strictly monotonic, and such that $f(\vartheta + 2\pi) = f(\vartheta) + 2\pi$. These conditions define the group

$$\begin{aligned} \text{Diff}_+(S^1) &= \{f \in C^1(\mathbb{R}) | f(\vartheta + 2\pi) \\ &= f(\vartheta) + 2\pi \quad \text{and} \quad f'(\vartheta) > 0\} / 2\pi\mathbb{Z} \end{aligned} \quad (2.53)$$

of orientation-preserving diffeomorphisms of the circle. Denoting with a dot the derivatives with respect to s , the function h is related to f by the constraint

$$\dot{r}^2 + r^2 \dot{\phi}^2 = \frac{1}{4} (1 - r^2)^2. \quad (2.54)$$

One also assumes that the functions f and h are infinitely differentiable and their derivatives $f^{(k)}$ and $h^{(k)}$, $k \geq 1$, with respect to ϑ , are of the order of one when ℓ is large. This implies that $\dot{r} = O(\ell^{-2})$, $\ddot{r} = O(\ell^{-3})$, $\dot{\phi} = O(\ell^{-1})$, $\ddot{\phi} = O(\ell^{-2})$, etc. Under these assumptions, Eq. (2.54) implies that h can be expressed in terms of f in a large- ℓ expansion of the form

$$h = 2\pi \sum_{n \geq 0} \left(\frac{2\pi}{\ell} \right)^n h_n, \quad (2.55)$$

where the coefficients h_n are local expressions involving higher and higher derivatives of f . For instance, the first terms are computed to be

$$\begin{aligned} h_0 &= f', \quad h_1 = -\frac{1}{2} f'^2, \quad h_2 = \frac{f''^2}{2f'}, \\ h_3 &= \frac{1}{8} f'^4 - \frac{1}{2} f''^2, \quad h_4 = -\frac{1}{2} f' f''^2 - \frac{5f''^4}{8f'^3} + \frac{f''^2 f'''}{f'^2}. \end{aligned} \quad (2.56)$$

The resulting curve is a gently “wiggling” boundary. The associated large- ℓ geometries are called “nearly hyperbolic.” Everything is fixed in terms of the diffeomorphism f , hence the name reparametrization *Ansatz*.

Reparametrization Ansätze and distorted disks. In the language of Sec. II D, the nearly hyperbolic geometries correspond to particular immersions $F: \mathcal{D} \rightarrow \mathbb{H}^2$ and, thus, to particular disk metrics $g = F^* \delta^-$. The cut-out piece of hyperbolic space is the image $F(\mathcal{D})$ and, thus, a special instance of a distorted disk. The immersions, or,

²³Note that, with the choice of metric (2.51), the units are chosen in such a way that the curvature length scale is one. “Large” thus means large compared to the curvature length scale.

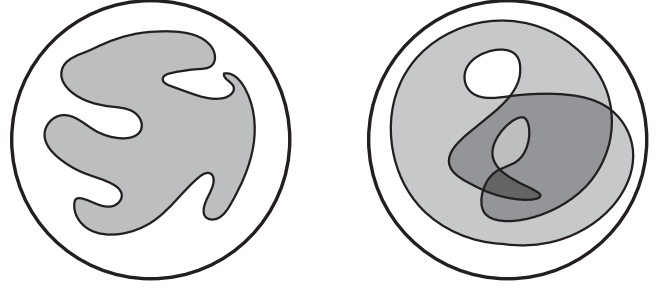


FIG. 6. Left inset: an embedding that is not a reparametrization *Ansatz*, because the function $\phi(s)$ is not monotonic. Right inset: an immersion that is not an embedding, because the associated distorted disk has overlaps. Note that both configurations are associated with constant curvature metrics on the disk and, thus, contribute to the JT quantum gravity path integral.

equivalently, the distorted disks, that are taken into account by the reparametrization *Ansatz*, are *very* atypical:

- (i) The reparametrization *Ansatz* immersions are actually embeddings, which implies, in particular, that the distorted disks have no overlaps.
- (ii) Since f is assumed to be strictly monotonic, the boundary curve cannot make zigzags.

An embedding that is not a reparametrization *Ansatz*, as well as an immersion that is not an embedding, is depicted in Fig. 6. Note that immersions that are not embeddings cannot be seen as “cut-out” pieces of hyperbolic space.

- (iii) The condition that the derivatives $f^{(k)}$ of the reparametrization are of the order of one means that we cut off frequencies of the order of ℓ or higher. In other words, the distances under which the wiggling of the boundary curve occurs must be much larger than the curvature scale.
- (iv) The reparametrization *Ansatz* describes geometries that have smooth, rectifiable boundaries, whereas typical geometries will have fractal boundaries almost surely.

Replacing the full path integral (2.49) by a path integral over the smooth wiggling boundaries described by the reparametrization *Ansatz* is, thus, an approximation in which only a measure zero subset of the possible metrics is taken into account. We shall argue in Sec. VI that this approximation is justified in the limit $\Lambda \rightarrow -\infty$ as an effective, long-wavelength description, for which the curvature length scale plays the rôle of an ultraviolet cutoff.

Remarks.

- (i) For any given cut-out region, F exists and is unique modulo the right action of $\text{Diff}_+(\bar{\mathcal{D}})$. In conformal gauge, this result is equivalent to the Riemann mapping theorem.
- (ii) The condition $f(\vartheta + 2\pi) = f(\vartheta) + 2\pi$ on the reparametrization f is usually imposed in a rather *ad hoc* way. One might wonder why other sectors with higher winding number, $f(\vartheta + 2\pi) = f(\vartheta) + 2\pi n$,

$n \in \mathbb{N}$, are not included (changing the sign of n would simply amount to changing the orientation). The answer is that it is only in the case $n = 1$ that the immersion F exists, and, therefore, it is only in this case that the gently wiggling curve is associated with a metric on the disk. This follows from the continuous version of the constraint (2.16); see Sec. III C 1.

The space $\text{Wig}(\mathcal{D})$ of wiggling boundaries. Exactly as in our discussion in Sec. II D 4, configurations that are related by the action of the $\text{PSL}(2, \mathbb{R})$ isometry group of $\mathbb{H}^2$ [Eq. (2.48)] must be identified. This action yields an action on $\text{Diff}_+(S^1)$, via the relation $\phi(s) = f(\vartheta)$ in (2.52), whose explicit form may be derived straightforwardly in the large- ℓ expansion by using (2.48) and $w = re^{i\phi}$. At leading order, the action matches with the left action of the standard $\text{PSL}(2, \mathbb{R})$ subgroup of $\text{Diff}_+(S^1)$, which amounts to identifying circle diffeomorphisms f and $\tilde{f}$ related by a relation of the form

$$\tan \frac{\tilde{f}}{2} = \frac{a \tan \frac{f}{2} + b}{c \tan \frac{f}{2} + d}, \quad ad - bc = 1. \quad (2.57)$$

The parameters (z_0, α) in (2.48) and (a, b, c, d) , constrained by $ad - bc = 1$, are related by

$$e^{i\alpha} = \frac{b - c - i(a + d)}{c - b - i(a + d)}, \quad z_0 = -\frac{b + c + i(a - d)}{b - c - i(a + d)}. \quad (2.58)$$

Moreover, diffeomorphisms that differ by a constant shift of the variable ϑ should also be identified, since there is no privileged point on the boundary.²⁴ The set of metrics associated with the reparametrization *Ansatz*, thus, form the space

$$\text{Wig}(\mathcal{D}) = \text{PSL}(2, \mathbb{R}) \backslash \text{Diff}_+(S^1) / S^1. \quad (2.59)$$

The notation Wig refers to the “wiggling” boundary.

Note that, unlike in (2.43), the group $\text{Diff}_+(\tilde{\mathcal{D}})$ does not appear in the quotient defining $\text{Wig}(\mathcal{D})$. This is consistent: The description in terms of the circle diffeomorphism f does not refer directly to the source disk $\mathcal{D}$, and, thus, the group $\text{Diff}_+(\tilde{\mathcal{D}})$ does not act on f .

In the above discussion, to derive (2.59) with an action of the isometry group given by (2.57), we have implicitly restricted ourselves to isometry transformations that are independent of ℓ . But, actually, the reparametrization *Ansatz* and the large- ℓ expansion are not consistent with the action of the full isometry group. Indeed, the hyperbolic space $\mathbb{H}^2$ is homogeneous and $\text{PSL}(2, \mathbb{R})$ acts transitively.

²⁴In most of the previous literature, it is actually implicitly assumed that one has a marked point on the boundary. This is not natural from the quantum gravity point of view. This point is discussed further below.

This implies, for example, that we can always find an isometry that maps any point on the wiggling boundary parametrized as in (2.52) to the center $r = 0$ of the Poincaré disk. This new configuration is clearly not of the form (2.52) for $h = O(1)$. One can also find isometries that create turning points and are, thus, inconsistent with the condition $f' > 0$. Since the full $\text{PSL}(2, \mathbb{R})$ is gauged, this is *a priori* problematic. However, the situation is not so different in spirit from the usual perturbative treatment of gauge theories, for which the usual gauge fixing conditions have Gribov ambiguities. The perturbative consistency is ensured by the fact that the Gribov copies are too far away in field space to be seen in perturbation theory.

The Schwarzian theory. The area $A[g]$ for a reparametrization *Ansatz* can be computed straightforwardly. One starts from the exact formula giving the extrinsic curvature of the boundary

$$k = \frac{2(1 + r^2)}{1 - r^2} \dot{\phi} + \frac{r \ddot{\phi}}{\dot{r}}. \quad (2.60)$$

As for h in (2.55) and (2.56), the extrinsic curvature has a local large- ℓ expansion involving higher and higher derivatives of f :

$$k = 1 + \sum_{n \geq 1} \left(\frac{2\pi}{\ell} \right)^n k_n. \quad (2.61)$$

The first few terms can be easily computed:

$$\begin{aligned} k_1 &= 0, & k_2 &= \text{Sch}[f] + \frac{1}{2} f'^2, & k_3 &= 0, \\ k_4 &= \frac{f'' f^{(4)}}{f'^2} + \left(\frac{f'''}{f'} \right)^2 - \frac{11 f''^2 f'''}{2 f'^3} - \frac{1}{2} f' f''' \\ &\quad + \frac{27}{8} \left(\frac{f''}{f'} \right)^4 - \frac{1}{8} f'^4 + \frac{1}{4} f''^2, \quad \text{etc.} \end{aligned} \quad (2.62)$$

The symbol $\text{Sch}[f]$ denotes the Schwarzian derivative of f :

$$\text{Sch}[f](\vartheta) = \{f, \vartheta\} = \frac{f'''}{f'} - \frac{3}{2} \left(\frac{f''}{f'} \right)^2 = \left(\frac{f''}{f'} \right)' - \frac{1}{2} \left(\frac{f''}{f'} \right)^2. \quad (2.63)$$

Using the Gauss-Bonnet formula (2.10) for $(h, b) = (0, 1)$ and $R = -2$, one then gets

$$A[g] = \ell - 2\pi + \frac{2\pi}{\ell} \int_0^{2\pi} \text{Sch} \left[\tan \frac{f}{2} \right] d\vartheta + O(\ell^{-3}). \quad (2.64)$$

The partition function within the reparametrization *Ansatz* approximation is then

$$W_{\text{Sch}}(\beta_S) = \int_{\text{Wig}(\mathcal{D})} d\mu(f) e^{-S_{\text{Sch}}[f]}, \quad (2.65)$$

for a “Schwarzian” action

$$\begin{aligned} S_{\text{Sch}}[f] &= -\frac{1}{4\beta_S} \int_0^{2\pi} \text{Sch} \left[\tan \frac{f}{2} \right] d\vartheta \\ &= \frac{1}{8\beta_S} \int_0^{2\pi} \left[\left(\frac{f''}{f'} \right)^2 - f'^2 \right] d\vartheta, \end{aligned} \quad (2.66)$$

where the parameter β_S was defined in (2.50). The formal measure $d\mu(f)$ is postulated to be the Fadeev-Popov measure obtained after gauge fixing the $\text{PSL}(2, \mathbb{R})$ symmetry from the natural formal left-invariant measure on $\text{Diff}_+(S^1)$. Modulo a multiplicative factor of the form $e^{\zeta_0 - \beta_S \varepsilon_0}$, where ζ_0 and ε_0 are arbitrary constants, the Schwarzian partition function can be computed exactly and turns out to be one-loop exact [17]:

$$W_{\text{Sch}}(\beta_S) = e^{\zeta_0 - \beta_S \varepsilon_0} \beta_S^{-3/2} e^{\frac{\pi}{4\beta_S}}. \quad (2.67)$$

The constants ε_0 and ζ_0 can be interpreted as renormalized ground state energy and zero temperature entropy, respectively; see below. They cannot be predicted in the Schwarzian theory, because they correspond to *a priori* arbitrary finite local counterterms one may add to the action, associated with a boundary cosmological constant and the topological Einstein-Hilbert terms, respectively.²⁵ This formalism, called the “Schwarzian theory,” is at the basis of essentially all the analytic calculations made in JT gravity; see, e.g., [38] and references therein.

The Schwarzian partition function (2.67) can be expressed in terms of a density

$$\rho_{\text{Sch}}(E) = \frac{2}{\pi} e^{\zeta_0} \sinh \sqrt{\pi(E - \varepsilon_0)} \quad (2.68)$$

as

$$W_{\text{Sch}}(\beta_S) = \int_{\varepsilon_0}^{\infty} \rho_{\text{Sch}}(E) e^{-\beta_S E} dE. \quad (2.69)$$

This may suggest a boundary quantum mechanical interpretation for which β_S is an inverse renormalized temperature and ρ_{Sch} the density of state for a boundary Hamiltonian H , such that

$$W_{\text{Sch}}(\beta_S) \stackrel{?}{=} \text{tr} e^{-\beta_S H}. \quad (2.70)$$

We put a question mark on top of the equal sign because this formula cannot be strictly correct as such. Indeed, the partition function for a Hamiltonian with a continuous

spectrum is always infinite²⁶ and, thus, (2.70) does not make sense. The state-of-the-art interpretation of the Schwarzian partition function, together with its generalizations, when multiple boundaries are present, is that it gives an ensemble average for a suitable random matrix model, the random matrix being the Hamiltonian itself [27,39]. Formula (2.70) is, thus, replaced by

$$W_{\text{Sch}}(\beta_S) = \langle \text{tr} e^{-\beta_S H} \rangle_{\text{disk}}, \quad (2.71)$$

where only the disk contribution in the genus expansion of the matrix model is kept on the right-hand side.²⁷

Discussion.

- (1) The Schwarzian partition function W_{Sch} should be the limit of the “microscopic” partition function $W^-(\beta_q, \Lambda)$ defined in (2.49), when $\Lambda \rightarrow -\infty$, with a particular scaling given by (2.50). Recall that ℓ is a macroscopic length parameter, which, in principle, can be expressed in terms of the microscopic parameters β_q and Λ ; see Sec. VI.

However, it turns out that the partition function (2.67), with the coefficient of proportionality $\beta_S^{-3/2}$, corresponds to a counting for which one marks a point on the disk boundary. The marking yields an additional factor of β_S with respect to the standard quantum gravity conventions, that we are using in the present paper.²⁸ Another way to state this is that the Schwarzian partition function corresponds to an integral over $\text{PSL}(2, \mathbb{R}) \backslash \text{Diff}_+(S^1)$ instead of $\text{Wig}(\mathcal{D}) = \text{PSL}(2, \mathbb{R}) \backslash \text{Diff}_+(S^1)/S^1$. The fact that the proportionality factor $\beta_S^{-3/2}$ corresponds to a counting with a marked point on the boundary is also supported by the calculations presented in [29] and is consistent with what one finds in simplified models [32,34,40].

- (2) The difficulties in interpreting the Schwarzian theory in a full-fledged quantum mechanical framework do not prevent the Schwarzian theory from capturing correctly some important features of genuine quantum mechanical systems in some limit. For example, the free energy predicted in the Schwarzian framework is of the form

$$\beta_S F = -\ln W_{\text{Sch}} = \beta_S \varepsilon_0 - \zeta_0 - \frac{\pi}{4\beta_S} + \frac{3}{2} \ln \beta_S. \quad (2.72)$$

²⁵In principle, counterterms should be discussed in the UV-complete microscopic description, not in the effective Schwarzian description. Consistency between the two descriptions imply an interesting constraint; see Sec. VI.

²⁶This is a consequence of the fact that the stationary states are not normalizable in the case of a continuous spectrum. A similar remark was also made in [17].

²⁷Higher genus contributions correspond to higher genus topologies on the JT gravity side.

²⁸Marking a point on the boundary is not natural in quantum gravity, because all the markings are associated with the same metric.

In the SYK model (or in similar quantum mechanical tensor or vector-matrix models), the free energy has the general form [22,32]

$$\beta F_{\text{SYK}} = \beta E_0 - N \left(s_0 + \frac{a_1}{\beta J} + \frac{a_2}{(\beta J)^2} + o(1/(\beta J)^2) \right) + \frac{3}{2} \ln(c\beta J) + o(N^0). \quad (2.73)$$

The parameters are the inverse temperature β , a dimensionful coupling constant J that plays the role of a UV cutoff for the Schwarzian description, and the number of degrees of freedom N ; a_1 , a_2 , and c are dimensionless numerical constants.²⁹ E_0 and Ns_0 represent the ground state energy and the zero temperature entropy, respectively. The expansion (2.73) is valid at large N and large dimensionless coupling βJ . Interestingly, the ratio $\beta J/N$ may be of the order of one. The leading terms at large N and βJ in (2.73), even when $\beta J/N$ is of the order of one, can be identified with (2.72) if we set

$$\beta_s = \frac{2\ell}{|\Lambda|} = \frac{\pi\beta J}{4Na_1}. \quad (2.74)$$

The Schwarzian description is also consistent with the thermodynamic properties of near-extremal black holes; in particular, it reproduces the linear growth of the heat capacity at low temperature. The emergence of a continuous spectrum is also in line with the idea that a large- N limit is taken. The near-conformal properties are associated with the large cutoff $\beta J \gg 1$.

- (3) In spite of its great interest, the Schwarzian description is far from capturing the full complexity of the SYK model or of similar quantum mechanical models. It is an infrared effective description that does not correctly account for the UV properties that are crucial to ensure the quantum consistency of the models. Simple conjectured extensions of the Schwarzian result, as reviewed below in Sec. II E 4, are also not able to reproduce the details of the large- βJ expansion of the quantum mechanical models, that involves fractional powers of $1/(\beta J)$. This is, of course, well known. What may not have been fully realized or emphasized until now is that the Schwarzian description has the same status with respect to JT gravity as it has with respect to SYK. It is an effective description of JT gravity that falls short of capturing essential features of the microscopic description, like, for instance, the fractal nature of the boundary. As a consequence, the one-loop

exactness of the partition function (2.67) does not have any fundamental meaning.

- (4) It is interesting to pinpoint the various levels at which we make an approximation when we use the Schwarzian description.

- (i) The description takes into account only a subset of measure zero in the space of constant negative curvature disk metrics. For instance, configurations like the ones depicted in Fig. 6 are not taken into account. Moreover, the boundary is assumed to be smooth, whereas it is a fractal at the microscopic level.
- (ii) The Schwarzian description is based on the use of an *approximation* of the area term $\frac{\Lambda}{16\pi} A$ in the action, valid on distance scales much larger than the curvature scale. It is easy to construct reparametrization *Ansatz* configurations for which the area is finite and yet not given by the Schwarzian action. For instance, take

$$f(\vartheta) = \vartheta + \frac{\alpha}{n^2} \cos(n\vartheta), \quad (2.75)$$

where the frequency n is of the order of ℓ . This *Ansatz* is such that f' and f'' are of the order of one at large ℓ , because the amplitude is chosen to go as $1/n^2$, but $f''' = O(\ell)$, $f^{(4)} = O(\ell^2)$, etc. In this case, the local expansions (2.55) and (2.61), and, thus, the formula (2.64) for the area, are not valid. However, the small amplitude in $1/n^2$ ensures that the condition

$$A[g] - \ell + 2\pi = O(1/\ell), \quad (2.76)$$

which yields the finiteness of the action after the usual counterterms have been subtracted, is satisfied. Note that it is also easy to construct finite action configurations that have many microscopic zigzags or overlaps.

- (iii) Last, but not least, the correct UV structure of the theory is not governed by the area term but by a completely different induced boundary gravitational action *à la* Liouville, as explained in [16,28]. The induced gravitational action is essential to ensure the consistency of the theory on all scales.

2. Edge modes

Basic idea. The so-called *edge modes* appear when gauge theories are formulated on manifolds with boundaries in such a way that the boundary conditions break part of the gauge symmetry on the boundary. The broken generators then become genuine dynamical degrees of freedom. A typical example is BF gauge theory on a disk with gauge group G . Naively, because the BF theory gauge

²⁹In the SYK model, one often writes $a_1 = 2\pi^2\alpha_s$.

connection is flat and the disk is contractible, there is no degree of freedom. But the standard boundary conditions break the gauge symmetry on the boundary. Edge modes are then described by maps $h: \partial\mathcal{D} \rightarrow G$, and the model is equivalent to the dynamics of a particle moving on G . A similar phenomenon occurs in the gauge theoretic description of JT gravity [41].

Gravitational edge modes are similarly associated with the partial breaking of the group of diffeomorphisms. They play a central role in three-dimensional asymptotically hyperbolic or AdS gravity, where the asymptotic boundary conditions break part of the full group of diffeomorphisms [42]. It is natural to study whether a similar description is possible in two dimensions as well [43]. Since imposing the constraint of constant curvature eliminates the local metric degrees of freedom, it is legitimate to seek a formulation in terms of edge modes. The wiggling boundary picture seems consistent with this idea.

However, there are major differences between the two- and the three-dimensional cases, and these differences seem to be at the origin of some confusion.³⁰ An important difference is that, in two dimensions, there is a unique complete asymptotically hyperbolic geometry in two dimensions, which is hyperbolic space itself, $H^2 = (\mathcal{D}, \delta^-)$. We, thus, work with finite cutoff geometries (finite quantum boundary length β_q or finite smooth length ℓ if one deals with smooth boundaries). But, in this case, the group of diffeomorphisms $\text{Diff}_+(\bar{\mathcal{D}})$ is *not* broken: It remains an *exact* gauge symmetry of the model, for any finite value of the microscopic parameters, and this unbroken gauge symmetry includes the boundary reparametrizations. *There is, thus, no standard edge mode in the fundamental gravitational description of the model on finite geometries.*

A construction in JT gravity and the space of embeddings. Let us consider the group $\text{Diff}_+(H^2)$ of diffeomorphisms of hyperbolic space. This is, of course, a group of diffeomorphisms of the disk $\mathcal{D}$, but, unlike for $\text{Diff}_+(\bar{\mathcal{D}})$, the diffeomorphisms of $\text{Diff}_+(H^2)$ do not need to extend smoothly to $\partial\mathcal{D}$, which is at infinite distance in H^2 . The group $\text{Diff}_+(H^2)$ acts on the space of immersions $\text{Imm}(\mathcal{D}, H^2)$ by left multiplication, $F \mapsto \psi \circ F$. Using $g = F^* \delta^-$, we get an action on the space of metrics $\text{Met}^-(\mathcal{D})$. The space $\text{Met}^-(\mathcal{D})$, thus, decomposes into a disjoint union of orbits under $\text{Diff}_+(H^2)$.

One of the orbits corresponds exactly to the space of embeddings, $\text{Emb}(\mathcal{D}, H^2) \subsetneq \text{Imm}(\mathcal{D}, H^2)$, which is associated to the subspace of metrics

$$\text{SAL}^-(\mathcal{D}) = \text{Iso}(H^2) \backslash \text{Emb}(\mathcal{D}, H^2) / \text{Diff}_+(\bar{\mathcal{D}}) \subsetneq \text{Met}^-(\mathcal{D}). \quad (2.77)$$

The notation $\text{SAL}^-(\mathcal{D})$ is motivated by the fact that embeddings are in one-to-one correspondence with the boundary curves that are self-avoiding loops. Note that all the embeddings are in the same orbit, because there is always a diffeomorphism that maps the interior of two distinct self-avoiding loops.

To obtain an explicit description of this orbit, we can start picking a particular element of the orbit, i.e., a given embedding F_0 , associated with a background metric $g_0 = F_0^* \delta^-$. For instance, we can use a round disk, $F_0(z) = Rz$ for some R , $0 < R < 1$. We denote by $\mathcal{D}_0 = \text{Im } F_0 \subset H^2$ the associated distorted disk and by $\text{Diff}_+(\bar{\mathcal{D}}_0)$ the subgroup of $\text{Diff}_+(H^2)$ that maps $\mathcal{D}_0$ onto $\mathcal{D}_0$ and $\partial\mathcal{D}_0$ onto $\partial\mathcal{D}_0$. Any $\psi \in \text{Diff}_+(\bar{\mathcal{D}}_0)$ is such that $\psi|_{\mathcal{D}_0} = F_0 \circ \varphi \circ F_0^{-1}$ for some $\varphi \in \text{Diff}_+(\bar{\mathcal{D}})$. For $\psi \in \text{Diff}_+(\bar{\mathcal{D}}_0)$, the embeddings F_0 and $\psi \circ F_0$, thus, correspond to the same metric, since $\psi \circ F_0 = F_0 \circ \varphi$ for $\varphi \in \text{Diff}_+(\bar{\mathcal{D}})$. The group $\text{Diff}_+(\bar{\mathcal{D}}_0)$ is, thus, a genuine gauge symmetry that realizes the group of diffeomorphisms of the source disk embedded in H^2 via F_0 . In other words, $\text{Diff}_+(\bar{\mathcal{D}}_0)$ is the isotropy group of the embedding F_0 under the action of $\text{Diff}_+(H^2)$ in the quotient (2.77). Overall, this implies that

$$\text{SAL}^-(\mathcal{D}) = \text{Iso}(H^2) \backslash \text{Diff}_+(H^2) / \text{Diff}_+(\bar{\mathcal{D}}_0). \quad (2.78)$$

This is an “edge-mode-like” description: One might say that the group $\text{Diff}_+(H^2)$ is broken and that the “edge modes,” that are the nontrivial diffeomorphisms with respect to the quotient (2.78), are the degrees of freedom of the theory. These edge modes are the diffeomorphisms that do not let the boundary $\partial\mathcal{D}_0$ invariant. They are sometimes called “radial diffeomorphisms.”

Let us emphasize that the edge mode terminology in the present context is dangerous, because it hides some crucial differences with the gauge theoretic models for which a fundamental edge mode description is possible. The differences are twofold.

- (i) First, the group $\text{Diff}_+(H^2)$ is not the group $\text{Diff}_+(\bar{\mathcal{D}})$ of diffeomorphisms of the source manifold $\mathcal{D}$. The group $\text{Diff}_+(\bar{\mathcal{D}})$ is the fundamental gauge symmetry of the gravitational theory. The group $\text{Diff}_+(H^2)$ appears when one describes the metrics via immersions in H^2 as the group of diffeomorphisms of the target space H^2 .
- (ii) Second, the space of metrics (2.78) described in this formalism is only a small subset of the full set of metrics, associated with embeddings, whereas the general case involves immersions.

³⁰We thank Junggi Yoon for a stimulating discussion on these topics. This discussion has motivated us to write the present subsection, the aim of which is to clarify certain subtle and potentially confusing aspects in a pedagogical way.

Let us note that the reparametrization *Ansatz*, which is itself a special case of embeddings, can be obtained from (2.78) by considering infinitesimal elements of $\text{Diff}_+(\mathbb{H}^2)$.

Some clarifications on the nature of the degrees of freedom in JT gravity. The formulation in terms of the wiggling boundary and/or in terms of edge modes may lead to considerable confusion as to the nature of the degrees of freedom of the model and in the way these degrees of freedom have to be treated. One might think that the metric δ^- on the target space $\mathbb{H}^2$, the wiggling boundary, and the metric g on the source disk $\mathcal{D}$ are three independent sets of variables that must be varied independently in an action principle or that must be integrated over in a path integral. This is incorrect. *The only degrees of freedom are the constant curvature metrics g on the source disk $\mathcal{D}$, modulo the action of $\text{Diff}_+(\bar{\mathcal{D}})$, as described in Secs. II A–II D. The wiggling boundary is nothing more than a convenient trick to describe a subset of the metrics on $\mathcal{D}$, using $g = F^* \delta^-$ and a set of special embeddings F . Varying the wiggling boundary amounts to varying F and, thus, g . Note that δ^- is fixed. Thus, the wiggling boundary does not correspond to new degrees of freedom in the model. As for the metric δ^- on $\mathbb{H}^2$, it is unique. Varying it or integrating over it does not make any sense.*

3. Approaches to the finite cutoff models

The first proposal for a microscopic definition of JT quantum gravity, for the case of negative curvature, was made by Kitaev and Suh in [32]. The proposal is based on two assumptions: (i) The degrees of freedom of the theory are in one-to-one correspondence with the shape of a boundary curve immersed in $\mathbb{H}^2$, and (ii) the boundary curve can be an arbitrary closed random walk.

These assumptions lead to an equivalence between quantum JT gravity in negative curvature and the quantum theory of a particle in $\mathbb{H}^2$. As argued in [32], the properties of the model so obtained are consistent with the results expected in the Schwarzian limit (2.50).

However, if we assume, as we do, that JT gravity is a theory of random constant curvature metrics, the assumptions made in [32] are inadequate. Indeed, on the basis of what we have already explained, it is clear that the vast majority of closed random walks do not bound a distorted disk and are, therefore, not associated with a metric. Simple examples of forbidden closed walks were depicted in Fig. 3. The rather subtle conditions under which a closed curve is associated with a metric will be explored in detail in Sec. III. So assumption (ii) is ruled out by the metric axiom. Moreover, and this is one of the biggest surprises to emerge from the analysis of Sec. III, the very idea that the degrees of freedom are in one-to-one correspondence with the shape of the boundary is actually erroneous. Assumption (i) must, therefore, also be called into question.

The conclusion is that the model proposed by Kitaev and Suh, even though perfectly well defined, is not a theory of

quantum gravity, if one is willing to insist on the fact that a theory of quantum gravity in two dimensions must be a diffeomorphism invariant theory of random metrics.

This being said, the Kitaev-Suh model is a very valuable tool to explore, at the qualitative level, some important physical features of JT gravity, like, for instance, the emergence of the Schwarzian description on long-distance scales for a large negative cosmological constant. A notable fact, that suggests that the Kitaev-Suh model is a sensible approximation to JT, is that the Hausdorff dimension of the Brownian paths is 2, matching the expected value for pure JT [16]. It can also be generalized to the flat and positive curvature cases, yielding interesting qualitative clues on the possible behavior of JT gravity in these cases as well. Further discussion of the model will be given in Sec. IV B and in [40].

Stanford and Yang have proposed in [34] a very different approach. They do keep the assumption (i) by Kitaev and Suh, that the degrees of freedom are in one-to-one correspondence with the shape of the boundary curve, but, instead of using arbitrary closed random walks, they propose that the boundary curves should be self-avoiding loops. As explained above, each self-avoiding loop corresponds to an embedding of the disk and is, thus, associated with a unique metric on the disk, belonging to the space $\text{SAL}^-(\mathcal{D})$ defined in (2.77). However, this takes into account only a measure zero fraction of the full set of constant negative curvature metrics. In particular, the Hausdorff dimension of a typical SAL boundary is $4/3$ and not 2 as in JT. It is unlikely that the SAL model could provide a satisfactory approximation to JT gravity, but we found some qualitative aspects of the model and of the discussion in [34] very instructive. For these reasons, we shall provide more information on the self-avoiding loop model in Sec. IV B.

Iliesiu *et al.*, in [35], use the same basic assumptions as in [34], but instead of starting from the usual microscopic definition of random self-avoiding loops, which is based on the natural uniform measure in the discretized lattice formulation and which yields, in particular, fractal boundaries, they develop the model directly in the continuum, using various assumptions on the path integral integration measure, etc., which are in line with the usual assumptions made in the Schwarzian description. In particular, they deal with smooth rectifiable boundaries. Of course, the results obtained in this way are very different from the results presented in [34]. Instead of an attempt at a microscopic definition of the model, the work in [35], and its interesting generalization in [44], may be better interpreted as being a proposal for a possible extension of the effective Schwarzian framework, trying to push its range of validity beyond the leading order in the near-hyperbolic limit (2.50) but by remaining within the framework of an effective long-wavelength description.

In the next subsection, we are going to show that the results in [35] yield a nontrivial conjecture on the form of

the semiclassical expansion of JT gravity that does not seem to have been noted previously. As we shall argue, this prediction appears to be hard to reconcile with the expected microscopic properties of the theory.

4. A conjecture to all orders in the large- $|\Lambda|$ expansion

According to [35], the finite cutoff disk partition function in negative curvature is given by³¹

$$W_{\text{IKTV}}^-(\beta_S) \equiv \int_{\varepsilon_0}^{\infty} \rho_{\text{IKTV}}(E) e^{-\beta_S E} dE \quad (2.79)$$

for a density

$$\rho_{\text{IKTV}}(E) = \frac{2}{\pi} e^{\zeta_0} \left(1 - \frac{32\pi}{\Lambda^2} (E - \varepsilon_0) \right) \times \sinh \sqrt{\pi(E - \varepsilon_0) \left(1 - \frac{16\pi}{\Lambda^2} (E - \varepsilon_0) \right)} \quad (2.80)$$

generalizing the Schwarzian density (2.68). The symbol $\equiv$ in Eq. (2.79) means that the formula must be understood perturbatively, as explained below. Indeed, if taken seriously as a strict equality, Eq. (2.79) would predict an imaginary part in the partition function. Instead, one must imagine that there is a cutoff which is $o(|\Lambda|^2)$ in the integral over E . One can also give an expression in terms of a modified Bessel function:

$$W_{\text{IKTV}}^-(\beta_S) \equiv e^{\zeta_0 - \beta_S \varepsilon_0} \frac{\Lambda^2}{8\ell} \frac{e^{-\frac{|\Lambda|\ell}{16\pi}}}{1 + \left(\frac{2\pi}{\ell}\right)^2} I_2 \left[\frac{|\Lambda|\ell}{16\pi} \sqrt{1 + \left(\frac{2\pi}{\ell}\right)^2} \right] \quad (2.81)$$

or an alternative integral representation using the Schwarzian density (2.68):

$$W_{\text{IKTV}}^-(\beta_S) \equiv \int_{\varepsilon_0}^{\infty} \rho_{\text{Sch}}(E) e^{\frac{\ell|\Lambda|}{16\pi} [1 - \sqrt{1 - 64\pi(E - \varepsilon_0)/\Lambda^2}]} dE, \quad (2.82)$$

which is suggested by an approach via the one-dimensional $T\bar{T}$ deformation [36].

The formulas (2.79)–(2.82) all yield the same near-hyperbolic expansion, corresponding to the limit (2.50), to all orders in $1/\ell$ at fixed β_S ; see below. However, they contain imaginary parts and differ by terms that are exponentially small, of the order of $\exp(-\frac{|\Lambda|\ell}{8\pi})$. These nonperturbative pieces in the formulas are unphysical.

³¹To compare with the formulas in [35], note that, in our conventions, the cosmological constant is related to the boundary value ϕ_b of the dilaton in the usual formulation of JT gravity by $|\Lambda| = -\Lambda = 2\phi_b$.

Reference [35] discusses possible ways to constrain the nonperturbative contributions, and the nonperturbative structure from the point of view of resurgence is also studied in [44].

The near-hyperbolic expansion to all orders. The near-hyperbolic expansion may be obtained straightforwardly by expanding the density (2.80). One can also use the known large- z asymptotic expansion of the Bessel function:

$$I_2(z) = \frac{e^z}{\sqrt{2\pi z}} \sum_{k \geq 0} \frac{\Gamma(k - 3/2) \Gamma(k + 5/2)}{2^k \pi k! z^k}. \quad (2.83)$$

Note that the (unphysical) nonperturbative corrections to (2.83) are proportional to ie^{-z} . One gets

$$\ln W_{\text{IKTV}}^- = \zeta_0 - \beta_S \varepsilon_0 + \frac{\pi}{4\beta_S} - \frac{3}{2} \ln \beta_S + \frac{\pi}{\beta_S} \sum_{k \geq 1} \left(\frac{\pi}{\ell} \right)^{2k} P_k(\beta_S/\pi), \quad (2.84)$$

where P_k is a polynomial of degree $k + 1$,

$$\begin{aligned} P_1(x) &= -\frac{1}{4} - 5x - 15x^2, & P_2(x) &= \frac{1}{2} + 10x + 30x^2 - 60x^3, \\ P_3(x) &= -\frac{5}{12} (3 + 64x + 216x^2 - 576x^3 + 432x^4), \\ P_4(x) &= \frac{7}{2} + 80x + 300x^2 - 960x^3 + 1080x^4 + 1440x^5, \\ P_5(x) &= -\frac{21}{2} - 256x - 1050x^2 + 3840x^3 - 5400x^4 \\ &\quad - 11520x^5 + 47520x^6, \text{ etc.} \end{aligned} \quad (2.85)$$

The first terms in (2.84) reproduce the Schwarzian partition function (2.67), which is the expected leading piece in the limit (2.50). These leading terms get contributions only at tree level and one loop, as is well known [17]. Interestingly, the above formulas also predict that the subleading corrections, at any given order $1/\ell^{2k}$, get contributions only up to $k + 1$ loops, the loop counting parameter being here β_S . To summarize, the IKTV framework suggests two non-trivial properties: (i) The effective Schwarzian description can be extended to all orders in $1/\ell^2$, and (ii) within this effective description, a consistent loopwise expansion exists and there is a nonrenormalization theorem, that generalizes the famous one-loop exactness at leading order, stating that the contributions at the order of $1/\ell^{2k}$ are $(k + 1)$ -loop exact, for all $k \geq 0$.

These properties are atypical for an effective, long-wavelength description. Usually, descriptions of this kind do not yield consistent loopwise expansions, since loop corrections probe arbitrarily short-distance scales. Indeed, our discussion below will cast some doubts on the validity

of these results, at least within our purely metric and two-dimensional point of view on JT gravity.

A *semiclassical large- $|\Lambda|$, fixed- ℓ expansion*. The IKTV framework actually has even more powerful consequences than those we discussed above and which do not seem to have been noted in the literature. By reordering the terms of the original large- ℓ , fixed- $\beta_S = 2\ell/|\Lambda|$ expansion, it is indeed possible to get a new

$$\Lambda \rightarrow -\infty, \quad \ell \text{ fixed} \quad (2.86)$$

expansion of the form

$$\ln W_{\text{IKTV}}^- = \zeta_0 - \beta_S \varepsilon_0 + |\Lambda| f_0(\ell) + f_1(\ell, \Lambda) + \sum_{L \geq 2} |\Lambda|^{1-L} f_L(\ell), \quad (2.87)$$

with

$$|\Lambda| f_0(\ell) = \frac{|\Lambda| \ell}{16\pi} \left[\sqrt{1 + \left(\frac{2\pi}{\ell} \right)^2} - 1 \right], \quad (2.88)$$

$$f_1(\ell, \Lambda) = -\frac{3}{2} \ln \frac{2\ell}{|\Lambda|} - \frac{5}{4} \ln \left[1 + \left(\frac{2\pi}{\ell} \right)^2 \right], \quad (2.89)$$

$$f_L(\ell) = p_L \left[1 + \left(\frac{\ell}{2\pi} \right)^2 \right]^{\frac{1-L}{2}}, \quad L \geq 2, \quad (2.90)$$

with numerical constants p_L ,

$$\begin{aligned} p_2 &= -15, & p_3 &= -60, & p_4 &= -180, \\ p_5 &= 1440, & p_6 &= 47520, & p_7 &= 921600, \\ p_8 &= \frac{121953600}{7}, & p_9 &= 373248000, \\ p_{10} &= 9628761600, \quad \text{etc.} \end{aligned} \quad (2.91)$$

This expansion may be derived either from (2.82) by using a saddle-point approximation or directly from (2.83). Note that the usual expansion (2.84) can also be obtained from (2.87), replacing $|\Lambda|$ by $2\ell/\beta_S$, β_S being fixed, using the fact that $|\Lambda|^{1-L} f_L(\ell) \propto 1/\ell^{2L-2}$ for $L \geq 2$ and reordering the terms. The two expansions, Eqs. (2.84) and (2.87), thus contain exactly the same information but presented in different ways.

The notable fact about the formulas (2.88)–(2.90) is that they show that the infinite series in $1/\ell^2$ contributing at a given order $|\Lambda|^{1-L}$ according to IKTV have a finite radius of convergence and can, thus, be summed. In principle, we then obtain an information which is valid for finite values of ℓ .

The term $\zeta_0 - \beta_S \varepsilon_0$ in (2.87) corresponds to the arbitrary bulk Einstein-Hilbert and boundary cosmological constant counterterms, as already discussed in previous subsections. The leading-order contribution (2.88) matches with the

renormalized classical action $\frac{|\Lambda|}{16\pi} (A_{\text{cl}} - \ell + 2\pi)$. The contribution $-\ell + 2\pi$ can be absorbed in the counterterms $\zeta_0 - \beta_S \varepsilon_0$, and A_{cl} is the area of a round disk of boundary length ℓ embedded into hyperbolic space. This configuration corresponds to the maximal area metric and is, thus, the classical solution of JT gravity. This is all perfectly consistent with the idea that the expansion (2.87) could represent the standard quantum gravity semiclassical expansion, at fixed boundary length ℓ , the semiclassical expansion parameter being $1/|\Lambda|$, consistently with the fact that the classical action is proportional to Λ ; see also Appendix A.

However, the extremely simple form of the result [Eqs. (2.89) and (2.90)] appears to be hard to reconcile with some basic predictions of our microscopic definition of JT quantum gravity. Actually, the very existence of the semiclassical limit (2.86) seems to be very difficult to make consistent with the microscopic formulation.

Indeed, as we have emphasized above, in particular, in Sec. II E 1, the smooth boundary length parameter ℓ is an effective, macroscopic parameter that makes sense only on length scales that are much larger than the curvature scale. Therefore, this parameter cannot enter, *a priori*, in formulas valid at a strictly finite cutoff. Instead, one expects that loop corrections should generate an anomalous dimension for the length, revealing the fractal nature of the boundary. This is exactly what is found in the context of the semiclassical $c \rightarrow -\infty$ limit *à la* Zamolodchikov studied in [29]. Moreover, the UV structure of the theory, which is essential in making any loopwise expansion sensible, is not governed by the area term but by an induced gravitational action that does not become semiclassical when $\Lambda \rightarrow -\infty$ [16,28]. Explicit evidence that this precludes the existence of a semiclassical expansion at $\Lambda \rightarrow -\infty$ is given in [29], and general arguments in this direction will also be presented in Sec. IV.

Actually, a difficulty with the expansion (2.87) can be directly detected, by looking at the flat space limit of the theory. To discuss the flat space limit, it is convenient to introduce explicitly the curvature length scale, denoted by L , by writing the Ricci scalar $R = -2/L^2$. The dimensionless parameters of the theory then become ℓ/L and ΛL^2 , and the flat space limit is $L \rightarrow \infty$ at ℓ and Λ fixed. In this limit, the tree-level term (2.88) reduces to $\frac{|\Lambda| \ell^2}{64\pi^2}$, which is the expected result: The classical solution of flat JT gravity corresponds to a Euclidean disk of circumference ℓ , which has area $\frac{\ell^2}{4\pi}$, and, thus, the classical action is $-\frac{|\Lambda|}{16\pi} \times \frac{\ell^2}{4\pi} = -\frac{|\Lambda| \ell^2}{64\pi^2}$. However, at the one-loop level, (2.89) yields a diverging result:

$$\begin{aligned} f_1 &= -\frac{3}{2} \ln \frac{2\ell}{|\Lambda| L^3} - \frac{5}{4} \ln \left[1 + \left(\frac{2\pi L}{\ell} \right)^2 \right] \\ &= \frac{1}{2} \ln \frac{|\Lambda|^3 \ell^2 L^4}{256\pi^2} + O(L^{-2}). \end{aligned} \quad (2.92)$$

This seems to be a serious problem but should not come as a surprise. The macroscopic parameter ℓ does not make sense at all in flat space JT gravity. The flat space theory does not have a regime that can be described by a Schwarzian effective theory or by smooth boundary curves. It is impossible to obtain information about the flat space theory within the near-hyperbolic framework. The reliability of the would-be one-loop result (2.89) and, by extension, that of the would-be higher loop results (2.90) or of the all-order near-hyperbolic expansion (2.84), which indicate that the large- ℓ series can be resummed to obtain results at finite ℓ , must therefore be called into question.

III. CURVES THAT BOUND A DISTORTED DISK

A. General introduction

Models of (discretized) random paths are pervasive in theoretical physics and have also been studied extensively by mathematicians. The simplest and most important example is the random walk. The lattice version has a universal continuum limit called Brownian motion. The associated probability measure is the Wiener measure, which is used, in particular, to define the path integral in Euclidean quantum mechanics. Other well-known examples of random paths are the Schramm-Loewner evolution processes, which are conjectured to describe the continuum limit of various models of physical interest: self-avoiding walks and polygons, percolation exploration paths, Ising model interfaces, etc. For some entry points in the literature on these subjects, see, e.g., [45] and the excellent textbook [46].

JT gravity is fundamentally a theory of random metrics. Yet, a description in terms of fluctuating boundaries seems very natural as well, at least when the base manifold is a disk. Actually, the fact that two equivalent dual formulations may exist, one in terms of random metrics and the other in terms of random paths, is conceptually nontrivial, since models of random geometries and models of random paths are usually associated with qualitatively different behaviors and different types of observables. In a sense, JT gravity seems to stand at a delicate equilibrium point between the two classes of models. This may be its distinctive and most interesting feature.

Our goal in this section is to show that a formulation in terms of random closed paths is indeed possible and that this formulation is actually much more subtle than what has been envisioned so far. We are going to show that the relevant random path model is of a new kind, with unusual properties.

Focusing on the disk topology, for which there is only one boundary, the main features are as follows.

The set of allowed closed curves must obey a subtle constraint: *Only closed curves that bound a distorted disk are allowed.* The notion of distorted disk, both at the

discretized level and in the continuum, has been introduced in Sec. II. Following a terminology found in the mathematical literature, we call the allowed curves *self-overlapping polygons* (SOPs) in the discretized case and simply self-overlapping curves in the continuum.

The set of SOPs is strictly between the set of SAPs and the set of arbitrary closed loops. This means that SAPs are SOPs, but a generic SOP is not a SAP, and that SOPs are closed loops (obviously), but a generic closed loop is not a SOP. As for SAPs, the constraint defining a SOPs is non-Markovian. However, whereas it is immediate to decide if a closed polygon is self-avoiding or not, the case of SOPs is much more subtle.

Fortunately, the problem of deciding whether a given closed curve is self-overlapping or not has been studied extensively in the mathematical literature, from many different points of view (analytic, geometric, algorithmic), starting with the works of Titus and Blank [47], as reviewed by Poénaru in [48]. The work of Blank completely solves the problem, using a nice method of “cuts” that will be illustrated below on examples. Motivated by possible applications in robotics and the design of integrated circuits, Shor and Van Wyk proposed an efficient algorithm which allows to determine if a curve is self-overlapping [49]. For a discretized curve of length $2n$, their algorithm runs in $O(n^3 \ln n)$ time. A more recent algorithm was proposed in [50], with average running time $O(n^2)$. There are also applications in the study of benzenoid hydrocarbons; see, e.g., [51] and references therein. Other references that we found very instructive include [52].

One of the most important messages of [16] and the present paper is that the correct formulation of JT quantum gravity in terms of a fluctuating boundary is a *theory of random self-overlapping polygons*. This provides an entirely new application of these objects and also emphasizes the importance of studying *random* SOPs, a subject on which virtually nothing is known.

The most surprising aspect of the model is that, in general, a given SOP does not determine a unique distorted disk; equivalently, it does not determine a unique metric. Since this is so important and understandably very confusing, let us say it again: There exist cases where *distinct* distorted disks, for which the associated disk metrics are physically different (which means that they are not equivalent modulo the action of disk diffeomorphisms), have exactly the *same* boundary curve. This implies that the correct measure on the set of SOP is not the uniform measure: A nontrivial multiplicity index must be attached to each SOP, which counts the number of distinct metrics associated with it. This multiplicity can be computed, in principle, by using the Shor–Van Wyk or the Mukherjee algorithms [49,50]. At a conceptual level, this phenomenon implies that *the degrees of freedom of JT quantum gravity are not in one-to-one correspondence with the shape of the boundary.*

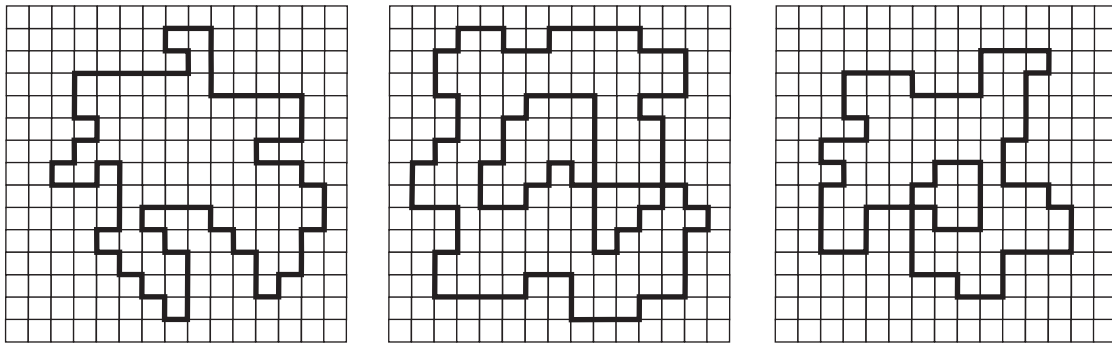


FIG. 7. Three instances of closed paths, or polygons, on the square lattice. The paths in the left and center insets are SOPs, but the path in the right inset is not; see the main text.

B. SOPs, boundary code, and symmetry factors

Consistently with Sec. II, we use a square lattice in our discussion; a similar treatment applies to triangular or hexagonal lattices.

1. Basic definitions

A closed path, or polygon, on the square lattice, of length ν , is a cyclic sequence of lattice points $(P_0, P_1, \dots, P_\nu)$, with $P_k = P_{k+\nu}$, such that P_{k+1} is obtained from P_k by walking along a single edge of the lattice (i.e., P_k and P_{k+1} are nearest neighbors), with the constraint that $P_{k+1} \neq P_{k-1}$ (we do not allow U turns). Note that ν is necessarily even, and, thus, we note $\nu = 2n$. We have depicted in Fig. 7 three typical polygons.

A path may be entirely coded by telling, at each step, in which direction we are going: straight, right, or left.³² For instance, the “path code” for the simplest possible closed path, encircling a unique lattice tile, is LLLL. Of course, only the cyclic ordering of the letters in the code is important, since the starting point along the path may be chosen arbitrarily.³³ The two possible orientations of a given path are considered to be equivalent, so two codes obtained by reversing the cyclic ordering and at the same time exchanging the L and R moves are identified.

A SOP is a closed path bounding a discretized flat disk (distorted disk) as defined in Sec. II. SOPs have a canonical “anticlockwise” orientation, which is such that the disk it bounds is on the left. The SOP code will always be chosen to be canonical. Note that the relation (2.16) is then automatically satisfied: The number of left turns minus the number of right turns is always equal to four. This rules out, for example, the polygon in the right inset in Fig. 7.

2. Counting and symmetry factors

To define a theory of random SOPs, we need to define a measure on the set of SOPs. For now, let us limit ourselves to a few basic remarks. This discussion will be refined, in a subtle and surprising way, in Sec. III D.

If we describe a SOP by its code, with the canonical orientation, two SOPs that coincide modulo lattice translations and rotations are automatically identified. Moreover, a reflection with respect to a vertical or horizontal axis acts by reversing the order of the letters in the path code. Because of this, we might consider the possibility of identifying a code and its reverse. Strictly speaking, this is the natural choice in the metric interpretation of the theory, but it turns out that it is not very convenient and, in particular, it does not match with the conventions used in Sec. II C. Since we want to keep using the same conventions, we make the choice of not identifying a configuration with its image under a reflection. Moreover, we also count with an appropriate multiplicity SOPs that do not match under the action of lattice rotations. This amounts to counting any “generic” SOP path code with a factor of 4. This choice will, thus, have a trivial effect in any well-defined continuum limit of the model.

Let us be a bit more precise. Let us pick a SOP of length $2n$. To each vertex v , we associate the “rooted” code B_v at v , which by definition is the representative of the code for which the first letter corresponds to the vertex v . A SOP is *generic* if the $2n$ rooted codes B_v , for any v , are all distinct. Generic SOPs are conventionally counted with a factor of 4.

More generally, pick a vertex v and let v' be the closest vertex following v along the SOP such that $B_v = B_{v'}$. Assume that the distance between v' and v is d , which we note $v' = v + d$. Thus, $B_v = B_{v+d} = B_{v+pd}$ for any integer p . The sequences of d letters from $v + pd$ to $v + (p+1)d - 1$ are the same for any p ; let us note this sequence σ . Moreover, since d is the smallest possible distance such that $B_v = B_{v+d}$, d necessarily divides $2n$. We can, thus, write $2n = dr$ for some positive integer r . The code of the SOP is, thus, a repetition of r times the

³²For the hexagonal lattice, only right and left are available.

³³This is consistent with the fact, mentioned in Sec. II, that we do not have a marked point on the boundary.

sequence σ . If v_L^σ and v_R^σ represent the numbers of left and right vertices, respectively, in the sequence σ , then, using (2.14), we get $v_L - v_R = 4 = r(v_L^\sigma - v_R^\sigma)$. Thus, r divides 4. There are only three possibilities: Either $r = 4$, in which case the SOP has a $\mathbb{Z}_4$ symmetry corresponding to a 90° lattice rotation; or $r = 2$ and the SOP has a $\mathbb{Z}_2$ symmetry corresponding to a 180° lattice rotation; or $r = 1$ in which case the SOP is generic.

The precise counting rule is then that $\mathbb{Z}_4$ -symmetric SOPs are counted once, $\mathbb{Z}_2$ -symmetric SOPs are counted twice, and generic SOPs are counted four times. This counting is equivalent to counting distinct rooted codes with a weight $2/n$.

An additional multiplicity factor, of a very different nature, will be introduced in Sec. III D.

C. General properties of self-overlapping curves

We are now going to present criteria that allow to decide whether a closed curve is self-overlapping or not, by using information directly contained in the shape of the curve. This discussion allows us to develop a better understanding of our problem and suggests some natural generalizations. Since the ideas apply equally well to discrete polygons and to continuous piecewise smooth curves, we deal with this latter case, which is more general. Note that in the continuum limit, which will be discussed in Sec. IV, these curves may become fractal. Strictly speaking, the discussion below applies before the continuum limit is taken.

Let us set up some notations and recall some of the results explained in Sec. IID. The source disk is as in Eq. (2.27), with coordinates

$$z = x^1 + ix^2 = \rho e^{i\theta}. \quad (3.1)$$

It is oriented by the volume form $dx^1 \wedge dx^2$ (anticlockwise orientation). The disk boundary $\partial\mathcal{D} = S^1$ is parametrized by $\theta \equiv \theta + 2\pi$. The target space $\mathbb{R}^2 = \mathbb{C}$ is parametrized by Cartesian coordinates (w^1, w^2) or by the complex coordinate $w = w^1 + iw^2$.

We consider closed, continuous, piecewise smooth curves $\gamma: S^1 \rightarrow \mathbb{R}^2 = \mathbb{C}$ such that $\gamma' \neq 0$, which is equivalent to saying that γ is an immersion of S^1 into $\mathbb{C}$.

The curve γ is self-overlapping if and only if it can be extended into an immersion of $\mathcal{D}$ into $\mathbb{C}$, that is to say, if and only if there exists an immersion $F: \mathcal{D} \rightarrow \mathbb{R}^2$ such that $F|_{\partial\mathcal{D}} = \gamma$. Without loss of generality, we always use orientation-preserving immersions, for consistency with the orientation conventions made up to now. We also know from Sec. IID that the immersion F may be chosen to be holomorphic, in which case

$$\gamma(\theta) = F(e^{i\theta}). \quad (3.2)$$

We want to discuss necessary (and even, ideally, sufficient)

conditions on γ that ensure that the extension by an immersion F exists. For the sake of simplicity, we assume that γ is generic, in the sense that it has a finite number of self-intersection points and that the self-intersections are transverse and correspond to double points (the discussion can be straightforwardly adapted to the more general cases).

Note that the most important points of the discussion in the present and the next Sec. IIID are purely topological in nature and do not depend on the metric on the target space. They, thus, apply to the cases of zero, positive, or negative curvature equally well.

1. Whitney index

Since the tangent vector γ' never vanishes, it has a well-defined winding number

$$W(\gamma) = \frac{1}{2i\pi} \int_0^{2\pi} \frac{\gamma''}{\gamma'} d\theta. \quad (3.3)$$

This is an integer, often called the Whitney index of the curve, which may also be seen as the winding number of the map from S^1 to S^1 defined by the normalized tangent vector $\gamma'/|\gamma'|$.³⁴ The Whitney index is an invariant of the curve modulo regular homotopies (homotopies γ_λ that are immersions, $\gamma'_\lambda \neq 0$, for any homotopy parameter λ). Whitney theorem states that closed curves in the plane are classified by their Whitney index modulo regular homotopies.

Proposition 3.1. Any self-overlapping curve with the anticlockwise orientation has a Whitney index $W(\gamma) = +1$.

Proof. A simple proof uses the representation (3.2) in terms of a holomorphic function F . One gets in this way

$$W(\gamma) = 1 + \frac{1}{2i\pi} \oint_{\partial\mathcal{D}} \frac{F''}{F'} dz. \quad (3.4)$$

Since $F' \neq 0$ everywhere in the disk by the immersion property, F''/F' is holomorphic in $\mathcal{D}$ and the contour integral over $\partial\mathcal{D}$ vanishes.

Another interesting proof uses the discretization. An immersion $F: \mathcal{D} \rightarrow \mathbb{C}$ being given, we can always assume, even if we are dealing with the positive or negative curvature models, that the target space is endowed with the flat Euclidean metric. We can then use the discretization in terms of which the curve γ becomes a polygon on the square lattice. It is clear that the total winding angle of the tangent vector is $\frac{\pi}{2}(v_L - v_R)$; the Whitney index is this divided by 2π and is, thus, equal to one by (2.16). ■

³⁴The normalized tangent vector $\gamma'/|\gamma'|$ is the unit tangent vector in the zero curvature case, when the target space is the Euclidean plane, but this is irrelevant for computing the winding number.

Corollary 3.1. Any self-overlapping curve has an even number of self-intersections.

Proof. This comes from the fact that the Whitney index and the number of self-intersections have opposite parities. Indeed, the number of self-intersections is always preserved modulo 2 by a regular homotopy. Since Whitney theorem ensures that any self-overlapping curve, which must have $W(\gamma) = 1$, is regular homotopic to a circle, and since a circle has no self-intersection, we conclude. ■

We immediately get an interesting generalization to arbitrary topologies. Imagine we have an immersion $F: \mathcal{M}_{g,b} \rightarrow \mathbb{R}^2$ from a surface of genus g with b circle boundaries into the plane. Let $\gamma_1, \dots, \gamma_b$ be the images of the boundary components of $\mathcal{M}_{h,b}$ by F . Then, (2.14) yields

$$\sum_{i=1}^b W(\gamma_i) = 2 - 2h - b. \quad (3.5)$$

2. Number of overlaps

Consider an arbitrary point P in $\mathbb{C} \setminus \gamma$, with complex coordinate w_P . The winding number $\nu_\gamma(P)$ of γ with respect to P is defined in the usual way to be the winding number of the ray vector from P to γ when the curve is traveled all over once in the anticlockwise direction:

$$\nu_\gamma(P) = \frac{1}{2i\pi} \int_0^{2\pi} \frac{\gamma'}{\gamma - w_P} d\theta. \quad (3.6)$$

The set

$$\mathbb{C} \setminus \text{Im } \gamma = (\cup_i \Omega_i) \cup \Omega_\infty \quad (3.7)$$

is the disjoint union of simply connected open sets, which are all bounded except Ω_∞ . The winding $\nu_\gamma(P)$ is integer valued and constant in each components Ω_i . Clearly, it is zero in the unbounded component.

If γ is a self-overlapping curve, it bounds a distorted disk which may overlap with itself. The number of layers is a constant in each open set Ω_i . We denote it by ν_i . Clearly, this number is always positive and is zero in the unbounded component.

Proposition 3.2. Let γ be self-overlapping. Then, the number of layers ν_i in a bounded simply connected component Ω_i of $\mathbb{C} \setminus \text{Im } \gamma$ is equal to the winding number $\nu_\gamma(P)$ of γ with respect to any $P \in \Omega_i$. In particular, $\nu_\gamma(P) \geq 0$ for any P .

Proof. Using (3.2), the integral (3.6) defining the winding number $\nu_\gamma(P)$ can be rewritten as

$$\nu_\gamma(P) = \frac{1}{2i\pi} \oint_{\partial\mathcal{D}} \frac{F'(z)}{F(z) - w_P} dz. \quad (3.8)$$

The contour integral picks a contribution 1 from each pole

of the integrand, i.e., from each solution of $F(z) = w_P$ for $z \in \mathcal{D}$. The number of solutions is the number of preimages of P under F , which is precisely the number of layers at P . ■

This result is particularly interesting, because the criterion $\nu_\gamma(P) \geq 0$ that any self-overlapping curve must satisfy is easy to check. It also yields a very simple way of counting the number of layers at any point without having to understand the details of the structure of the distorted disk bounded by the curve. This is actually quite nontrivial, as will become clear in the next Sec. III D.

Corollary 3.2. Let g be a constant curvature metric on $\mathcal{D}$. Then, the area $A[g]$ of $\mathcal{D}$ with respect to g depends only on the shape of the boundary of the associated distorted disk.

Proof. Writing as usual $\mathbb{C} \setminus \text{Im } \gamma = (\cup_i \Omega_i) \cup \Omega_\infty$, we have

$$A[g] = \sum_i \nu_i A[\Omega_i], \quad (3.9)$$

where ν_i is the number of layers of the distorted disk covering Ω_i and $A[\Omega_i]$ is the area of the component Ω_i computed with the relevant target space metric, which is given by (2.28), (2.39), or (2.40) depending on whether we are considering the case of zero, positive, or negative curvature. ■

Three simple examples, with the winding numbers in each components explicitly indicated, are depicted in Fig. 8.

Remark 1. A simple and convenient way to determine the winding around a point P is to draw a generic oriented half line from P to infinity and to count the number of crossings of this half line with the curve, with a plus sign if the curve crosses from right to left and a minus sign if the curve crosses from left to right.

Remark 2. If γ is an arbitrary closed curve, there is no obvious notion of “area enclosed by γ .” But we can always write $\mathbb{C} \setminus \text{Im } \gamma = (\cup_i \Omega_i) \cup \Omega_\infty$ and define the winding numbers $\nu_i \in \mathbb{Z}$ with respect to the bounded simply connected components Ω_i . This yields three natural notions of area:

$$\begin{aligned} A_{\text{Lévy}}[\gamma] &= \sum_i \nu_i A[\Omega_i], \\ A_{\text{wind}}[\gamma] &= \sum_i |\nu_i| A[\Omega_i], \\ A_{\text{arith}}[\gamma] &= \sum_i A[\Omega_i] \end{aligned} \quad (3.10)$$

called the Lévy (or algebraic, or stochastic), winding, and arithmetic areas, respectively. The algebraic area was studied long ago by Lévy in the context of Brownian motion [53]. It plays an essential rôle in the definition of the Kitaev-Suh model [32] and its generalizations; see Sec. IV B and [40].

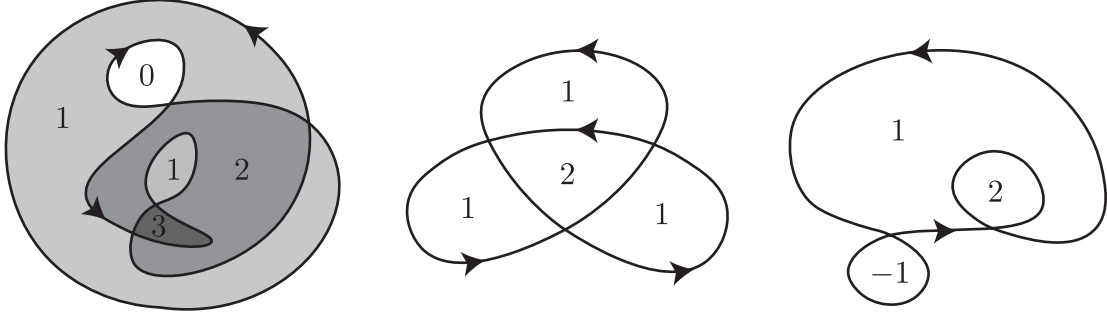


FIG. 8. Three examples of closed curves γ for which we have indicated the winding numbers ν_γ in each bounded simply connected components of $\mathbb{C} \setminus \text{Im } \gamma$. The curve on the left is self-overlapping. It is a slightly more complicated version of the example in Fig. 2. The curve in the center satisfies the constraints $\nu_\gamma(P) \geq 0$ everywhere but has an odd number of self-intersections (and Whitney index 2) and, thus, cannot be self-overlapping. It is actually an interior boundary; see Sec. III C 3. The curve on the right has Whitney index one, but there is a region for which $\nu_\gamma < 0$ and, thus, it is not self-overlapping (nor an interior boundary).

3. Blank cuts, Blank words, interior boundaries

A cut of an oriented closed curve γ is a choice of two distinct points P and Q on γ together with a cut path from P to Q that crosses γ transversally at P and Q . To a cut of γ is associated a decomposition of γ into two closed oriented curves: one obtained by going from P to Q along γ and then from Q to P along the cut path and the other obtained by going from Q to P along γ and then from P to Q along the cut path.

There are two types of cut. A *Blank cut* is such that the curve crosses the cut path from left to right at one of its end points and from right to left at the other end point. The two possibilities and the associated decomposition are shown in Fig. 9. If γ decomposes in γ_1 and γ_2 , the Whitney indices are such that $W(\gamma_1) + W(\gamma_2) = W(\gamma) + 1$ for a Blank cut and $W(\gamma_1) + W(\gamma_2) = W(\gamma)$ otherwise.

It can be shown [47,48] that a curve is self-overlapping if and only if a decomposition in terms of Blank cuts having the following properties exists. First, using a Blank cut, $\gamma = \gamma_0$ is decomposed into a simple (i.e., non-self-intersecting) curve $\tilde{\gamma}_0$ and a new closed curve γ_1 . If γ_1 is simple, the procedure stops. Otherwise, we use a new Blank cut to decompose γ_1 into a new simple curve $\tilde{\gamma}_1$ and γ_2 . And we keep going in this way until, in the last step, the two curves produced by the decomposition are simple.

The intuition behind this approach is that the interiors of the simple curves $\tilde{\gamma}_i$, called *Blank domains*, are pieces of the distorted disk that do not overlap, from which the full distorted disk can be reconstructed by gluing along the cuts.

This procedure can be made systematic in the following way. In each bounded connected component Ω_i of $\mathbb{C} \setminus \text{Im } \gamma$, $1 \leq i \leq m$, pick a point P_i and draw an oriented half line L_i from P_i to infinity in a generic way. Then build a word out of m letters a_i as follows. Go all over γ starting from an arbitrary base point. Each time you cross a half line L_i from right to left, add the “positive” letter a_i to the word. Each time you cross a half line L_i from left to right, add the “negative” letter a_i^{-1} to the word. The result is called a Blank word. Since the base point on γ is arbitrary, the Blank word is defined modulo cyclic permutations of the letters.

Note that the number of layers ν_i of the distorted disk on Ω_i is equal to the number of times the letter a_i occurs minus the number of times a_i^{-1} occurs in the Blank word.

A Blank cut corresponds to a pair (a_i, a_i^{-1}) in the word such that only positive letters appear between a_i and a_i^{-1} (or between a_i^{-1} and a_i). A step in the Blank cut decomposition corresponds to “reducing” the word by removing a_i, a_i^{-1} and the sequence of positive letters in between. The curve γ is self-overlapping if and only if $W[\gamma] = 1$ and the word can be reduced in such a way that at the end we are left with positive letters only. One can show that the conclusion does

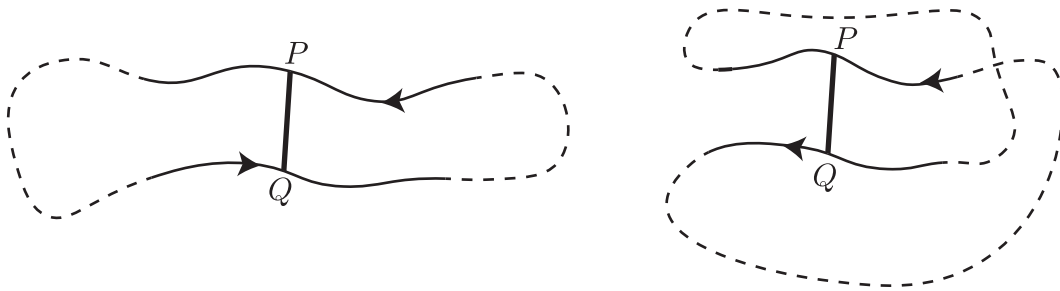


FIG. 9. A Blank cut (left inset) and a non-Blank cut (right inset). The cut path is the thick line.

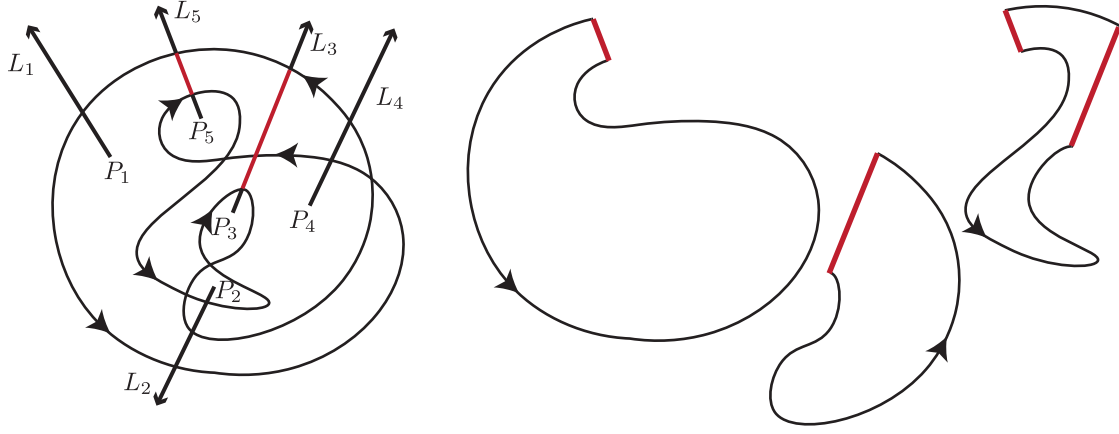


FIG. 10. Blank cut decomposition of a self-overlapping curve, with three Blank domains. The Blank word is $a_5 a_1 a_2 a_4 a_3 a_5^{-1} a_2 a_3^{-1} a_2 a_4 a_3$. It is reduced to a_2 by removing the sequences $a_5 a_1 a_2 a_4 a_3 a_5^{-1}$ and then $a_3^{-1} a_2 a_4 a_3$. These sequences, together with the final word, represent the three simple curves depicted on the right. These simple curves bound pieces from which the distorted disk can be reconstructed, by gluing along the Blank cuts (outlined in red).

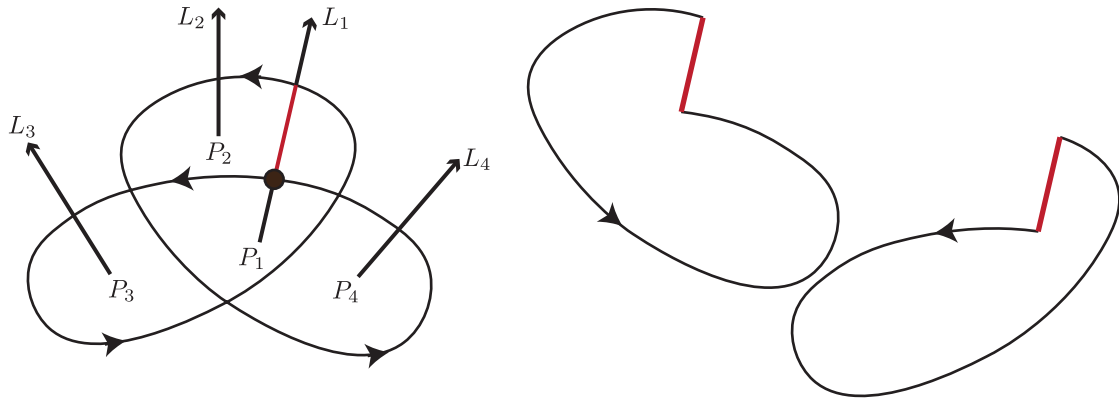


FIG. 11. An interior boundary with Blank word $a_1 a_2 a_4 a_1 a_3$. It can be reconstructed by gluing the interior of two simple curves, shown on the right, corresponding to the non-Blank cut $a_1 a_2 a_4 a_1$. The associated map $F: \mathcal{D} \rightarrow \mathbb{C}$ has a branch point, indicated by a black bullet.

not depend on the set of generic points P_i and half lines L_i we pick. The problem of recognizing a self-overlapping curve is, thus, reduced to a purely combinatorial analysis of the properties of the Blank word.

To illustrate the Blank word method, we have depicted in Fig. 10 a Blank cut decomposition for the self-overlapping curve on the left in Fig. 8.

An interesting generalization corresponds to the case when a decomposition in terms of simple curves is possible, but only if one allows both Blank and non-Blank cuts. This amounts to allow pairings of a_i with itself (or of a_i^{-1} with itself) in the Blank word. An example is depicted in Fig. 11. Curves with this property are called *interior boundaries* in the mathematical literature. Note that the configurations depicted in the second and third insets in Fig. 3 are interior boundaries, their boundary curve being of the type of the example depicted in the center in Fig. 8. Interior boundaries clearly satisfy the

positivity constraint $\nu_\gamma(P) \geq 0$ on the winding numbers, but their Whitney index may be an arbitrary positive integer. One can show that interior boundaries of Whitney index k may be split by removing recursively k self-overlapping pieces corresponding to closed arcs of the curve starting and ending at a self-intersection point. In particular, interior boundaries that have $W = 1$ are self-overlapping.

The defining property of a self-overlapping curve γ is that there exists an immersion $F: \mathcal{D} \rightarrow \mathbb{C}$ such that $F|_{\partial \mathcal{D}} = \gamma$. In the case of an interior boundary, one still has a map $F: \mathcal{D} \rightarrow \mathbb{C}$ such that $F|_{\partial \mathcal{D}} = \gamma$, but F is constrained only to be open (the image of an open set is an open set), light (the preimage of a point is a collection of distinct points), and orientation preserving. Modulo a disk diffeomorphism, one can always assume that F is analytic. For a self-overlapping curve, $F' \neq 0$ everywhere, but for an interior boundary F may have a finite number of branch

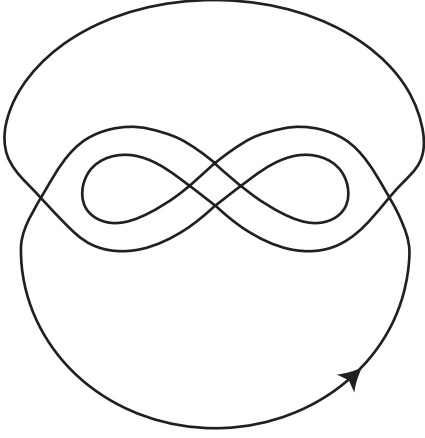


FIG. 12. The Milnor curve.

points around which it behaves as $F \sim (z - z_0)^r$ for $r \geq 2$.³⁵ Interior boundaries, thus, do not correspond to smooth metrics on the disk.

D. The Milnor curve and multiplicity

When the boundary curve is a self-avoiding loop, it separates the plane into two regions, the interior and the exterior. The associated distorted disk is then easily and uniquely identified with the interior region. When the boundary curve has self-intersections, it is no longer obvious, in general, to decide whether it bounds a distorted disk or not and, if it does, to reconstruct the distorted disk explicitly. A systematic procedure to solve this problem, using Blank cuts and words, has been outlined in the preceding subsection.

Let us now consider the closed curve depicted in Fig. 12. We call this curve the Milnor curve.³⁶ This is a typical example where a simple visual inspection is not enough to immediately decide whether it bounds a distorted disk or not. But we can straightforwardly compute a Blank word and look at its possible reductions. We let the reader do this very instructive exercise carefully.

The result is extremely interesting: One finds that the curve is self-overlapping, but *it actually bounds two distinct distorted disks*.

The two disks correspond to two distinct reductions of a Blank word, using distinct set of cuts. This is depicted in Fig. 13. It is important to realize that the pieces making up the two distorted disks are genuinely different and could not be made identical by cutting in different ways. Mathematically, this means that the two immersions $F_i: \mathcal{D} \rightarrow \mathbb{C}$, $i = 1$ or 2 , corresponding to the two distorted disks, are inequivalent in the sense that there is no diffeomorphism ϕ of the disk such that $F_1 = F_2 \circ \phi$. In

particular, and this is the most relevant point for us, the two disk metrics $g_1 = F_1^* \delta$ and $g_2 = F_2^* \delta$ are physically distinguishable.

This fundamental example shows that the correspondence between self-overlapping curves and disk metrics is not one to one. In JT gravity language, this is equivalent to saying that *the degrees of freedom of the theory (the space-time metrics) are not in one-to-one correspondence with the shape of the allowed (self-overlapping) boundaries*.

Definition 3.1. The number of distorted disks bounded by a closed curve γ is called the multiplicity of γ and is denoted by μ_γ .

The integer μ_γ can always be determined systematically by counting the number of distinct reductions of a Blank word of the curve or, equivalently, by running the algorithm of Shor and Van Wyk.

Let us mention two general and useful qualitative features.

- (i) If $\mu_\gamma > 1$, then one can show that there must exist a simply connected component of $\mathbb{C} \setminus \text{Im } \gamma$ for which the number of layers is $\nu \geq 3$. This result was conjectured in [49] and proven by Graver and Cargo [52]; see also [51]. In the case of the Milnor curve, Figs. 12 or 13, the small central connected component of $\mathbb{C} \setminus \text{Im } \gamma$, roughly square in shape, is covered three times by any of the two associated distorted disks.
- (ii) As we explained in Sec. III C 2, the number of layers ν_i in each simply connected component of $\mathbb{C} \setminus \text{Im } \gamma$ is entirely determined by the shape of the boundary curve only. The sets of integers ν_i , thus, match for all the distinct distorted disks bounded by a given curve. Suppose that $\mu_\gamma \geq 2$ and denote by $g_1, \dots, g_{\mu_\gamma}$ the corresponding metrics. It then follows from Corollary 3.2, or, equivalently, from Eq. (3.9), that the areas of the disk calculated with these different metrics must all be equal:

$$A[g_1] = A[g_2] = \dots = A[g_{\mu_\gamma}]. \quad (3.11)$$

This is nontrivial, because the metrics g_i are not equivalent under the action of a diffeomorphism.

Remarks. (i) The Milnor curve can be generalized in a rather straightforward way to construct examples that have an arbitrary integer multiplicity; see Fig. 3 in [48] and the discussion in Sec. IV. (ii) In the case of interior boundaries, there will typically be a moduli space associated with a given curve, since the positions of the branching points depend on the particular choice of Blank cuts. An example of this is given by the configurations in the second and third insets in Fig. 3.

E. Generating functions and generalizations

1. Generating functions

We now have all the ingredients we need to define the JT gravity models from the point of view of the boundary

³⁵See, e.g., [54] for details.

³⁶It seems to be acknowledged that this curve was first found by Milnor, even though his discovery was never published.

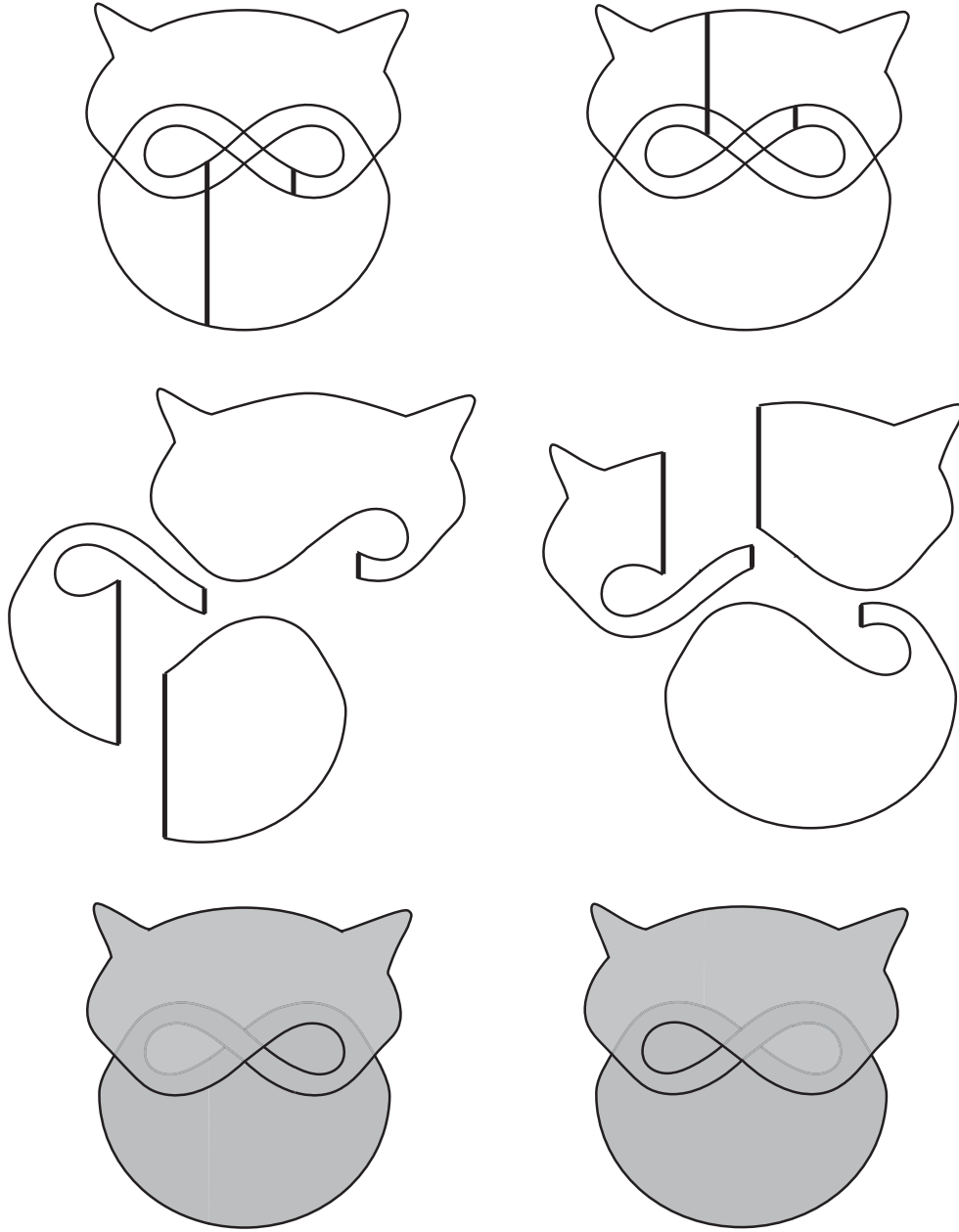


FIG. 13. The Milnor curve bounds two distinct distorted disks, depicted on the left and on the right. Upper insets: the Milnor curves with two distinct sets of Blank cuts. Central insets: the resulting simple curves. Lower insets: reconstruction of the disks by gluing the simple pieces along the cuts.

curve. We focus here on the zero curvature case. The cases of negative and positive curvatures will be briefly discussed in Secs. III E 4 and VI.

The generating (or partition) function is

$$W(t, g) = \sum_{\gamma} \mu_{\gamma} t^{F(\gamma)} g^{v(\gamma)} = \sum_{n, p \geq 1} W_{2n, p} t^p g^{2n}. \quad (3.12)$$

The sum is over all self-overlapping polygons, counted as explained in Sec. III B, with the additional multiplicity

factor μ_{γ} defined in Sec. III D. $F(\gamma)$ is the number of faces of the associated distorted disks, which depends only on γ , according to (3.11), and $v(\gamma)$ is the boundary length or, equivalently, the number of letters in the boundary code.

The definition (3.12) of the generating function has been carefully designed to be equivalent with the definition in Sec. II C.

As already mentioned in Sec. II C, at low orders, the generating function W matches with the standard generating function for self-avoiding polygons. The counting starts

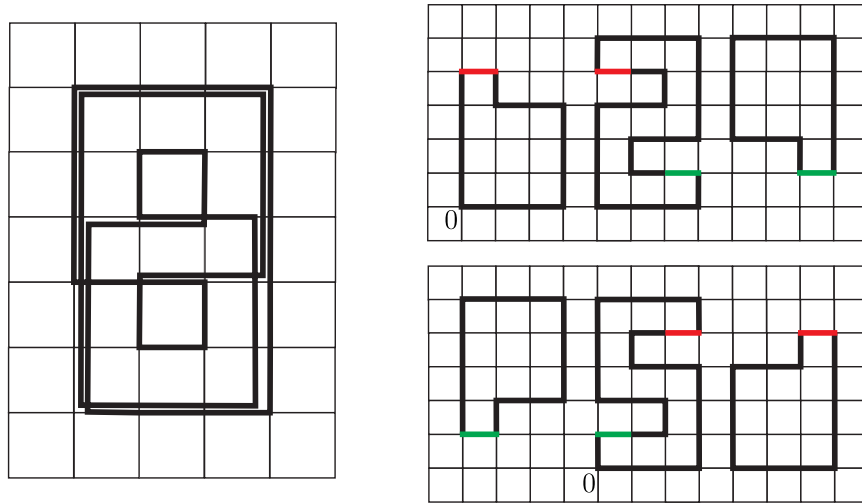


FIG. 14. The smallest SOP with multiplicity two on a square lattice, with boundary code LSSLSSLSRRRRSLSS LSSLSSSLSSLSSLSRRRRSLSSLSSSS, $2n = 48$, and $p = 31$ (the code is rooted on a vertex denoted by 0 in the figure; see also Fig. 21). This configuration is $\mathbb{Z}_2$ symmetric (invariant under a 180° rotation) and has multiplicity 2. It is, thus, counted $2 \times 2 = 4$ times in the generating function. Similar Milnor configurations with no symmetry can be easily obtained, for instance, by adding a face at the bottom, with new boundary code LSSLSSLSRRRRSLSSLSSSLSSLSSLSRRRRSLSS**RLRL**SSSS (the added sequence is in boldface). Such a configuration must be counted 8 times in the generating function.

to deviate at the order of $g^{18}t^8$, for which there exists a SOP that is not a SAP (see Fig. 4). Moreover, the lowest order at which a SOP with higher multiplicity appears is $t^{31}g^{48}$.³⁷ This configuration is a Milnor-like polygon, depicted in Fig. 14.

The fact that higher multiplicities first appear at relatively high orders in the expansion of the generating function may convey the impression that higher-multiplicity curves are very rare, but this is misleading. A simple argument given in Sec. IV shows that the number of SOPs of length $2n$ that have a higher multiplicity is exponentially large in n . We shall also see that, for a given SOP, the multiplicity index can also be exponentially large in n .³⁸

The generating function (3.12) can be generalized to a more refined version in a natural way. Let $\mathscr{W}_{v_L, v_S, v_R, p, \mu}$ be the number of SOPs such that the boundary code contains v_L , v_S , and v_R letters L, S, and R, respectively, of total area p and of multiplicity μ . The generalized generating function is

$$\mathcal{W}(t, g_L, g_S, g_R, u) = \sum_{\substack{p \geq 1, v_L \geq 4 \\ v_S, v_R \neq 0}} \mathcal{W}_{v_L, v_S, v_R, p, \mu} t^p g_L^{v_L} g_S^{v_S} g_R^{v_R} u^\mu. \quad (3.13)$$

Note that, due to the constraint (2.14), there is a trivial redundancy in the parameters.

³⁷We have not tried to prove this fact rigorously. For an analysis on the hexagonal lattice, see [55].

³⁸I thank Peter Shor for pointing this out to me.

2. Open self-overlapping curves

It is possible to define a natural notion of open self-overlapping curves, joining two distinct points w_1 and w_2 in $\mathbb{C}$. We consider a fixed oriented curve γ_{21} from w_2 to w_1 . An obvious possibility is to choose a straight line, which would correspond to “chordal self-overlapping curves,” but more generally γ_{21} may be an arbitrary continuous curve from w_2 to w_1 . We then say that an oriented open curve γ from w_1 to w_2 is self-overlapping if and only if the closed curve $\gamma_{21} \cup \gamma$ is self-overlapping.

This notion is not directly relevant for the quantum JT gravity application that we study in this paper, but investigating random open self-overlapping curves may present some mathematical interest.

3. Remarks on other topologies

The concept of distorted disks explained in Sec. II can be straightforwardly generalized to the concept of distorted surfaces of arbitrary topologies, with genus h and b boundary components. In the continuum, distorted surfaces are described by immersions $F: \mathcal{M}_{h,b} \rightarrow \mathbb{C}$ that, by pull-back, yield constant curvature metrics on $\mathcal{M}_{h,b}$.

For higher topologies, however, the point of view in terms of the boundary curves becomes extremely subtle, even more so than the disk case discussed above. The picture that emerges seems staggering.

Even for the annulus, which seems to be the simplest case beyond the disk, the ideas developed for the disk are no longer applicable. For instance, one can readily see that boundary curves of arbitrary shapes are allowed, by

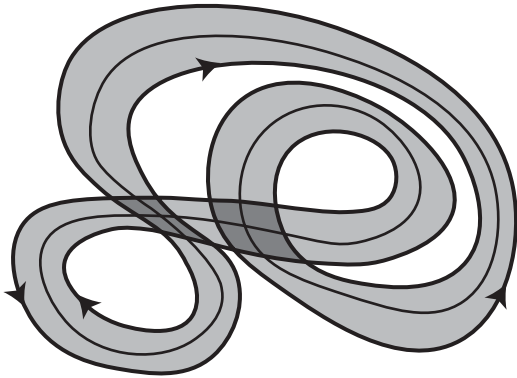


FIG. 15. A distorted annulus obtained by the thickening of a non-self-overlapping closed curve, which is depicted as a thin black line.

“thickening”—see Fig. 15. At a deeper level, it was shown in [56] that the problem of determining whether a curve bounds an arbitrary immersed surface or not is NP-complete, in sharp contrast with the case of the disk which is polynomial [49].

Moreover, and unlike for the disk, the set of constant curvature metrics obtained by pullback from immersions is only a subset of the full set of constant curvature metrics on surfaces of more complicated topologies. The associated flat connections are indeed constrained to have trivial holonomies along the noncontractible cycles of the surface.³⁹

The case of higher topologies will be briefly discussed again in Secs. V and VII.

4. The cases of nonzero curvature, light version

The discretized formulation of flat JT gravity presented in Sec. II or above in Sec. III is based on the use of a regular tessellation of the flat Euclidean space, for instance, the square lattice. This can be straightforwardly generalized to constant negative or positive curvatures, by using tessellations of hyperbolic space or the two-sphere, but these generalizations come with interesting subtleties and limitations.

In negative curvature, there is an infinite number of regular tessellations $\{p, q\}$ made of regular p -gons and degree q vertices. The edges and vertices associated with these tessellations form the so-called regular hyperbolic lattices. The only constraint that the integers p and q must satisfy is

$$(p - 2)(q - 2) > 4, \quad (3.14)$$

which follows from the formula for the curvature given in (2.5). The hyperbolic lattices are commonly used to discretize or approximate hyperbolic space; see, e.g., [57]

for recent interesting applications. Random self-avoiding walks have also been studied on these lattices [58,59]. A natural approach to JT quantum gravity in negative curvature is, thus, to consider models of random self-overlapping polygons, counted with multiplicity, on the regular hyperbolic lattices. As will be explained in Sec. VI, these models may have a very interesting physical interpretation, but they are not suitable to define JT gravity in a UV-complete way.

In positive curvature, regular tessellations correspond to the five Platonic solids. In particular, they have a finite number of tiles. Such discretizations are clearly inadequate to approximate JT gravity in any useful way. More general irregular tessellations of the two-sphere must be used to obtain an interesting model.

We refer to Sec. VI for a more detailed discussion of the models in nonzero curvature.

IV. ON RANDOM SELF-OVERLAPPING CURVES

The aim of this section is to provide a basic introduction to the study of random self-overlapping curves, a subject on which nothing seems to have ever been published. Our goal will be to outline some interesting expected properties of the model, emphasizing, in particular, the continuum limit. We focus on the flat theory, but we also discuss, mainly at the qualitative level, the cases of negative and curvature. The microscopic properties are the same, whether the curvature is positive, zero, or negative, but the macroscopic features will differ, even qualitatively, between the three cases.

We begin the section with two review subsections on some aspects of standard Liouville gravity and random curves models. The problem we are studying is subtly lying somewhere in between these two classes of models and having a synthetic presentation of some relevant features at hand is very useful (and very hard to find in the existing literature). Our goal is to put our analysis into an appropriate context, introduce some important concepts, refine our intuition, and motivate some of the ideas and conjectures we will propose. References include [8,13,15,60] on the Liouville side and [53,61] and [37,45,46] for Brownian and self-avoiding curves, respectively, as well as many additional references therein.

A. Review of Liouville gravity

1. Pure Liouville gravity on the sphere

Discretized formulation. The standard two-dimensional gravity is a theory of random metrics on a two-dimensional surface of fixed topology.⁴⁰ It can be defined rigorously by using the ideas reviewed in Sec. II A, starting from the set

³⁹I thank the referee for making this useful observation.

⁴⁰One may also consider the possibility to sum over all the possible topologies, but this aspect is not discussed in the present paper.

of discretized metrics obtained by quadrangulating the surface and then taking an appropriate continuum limit.⁴¹ The counting of the possible configurations was performed long ago by Tutte [1] by using combinatorial arguments. The counting can also be made elegantly by solving a quartic matrix model in the large- N limit [3].

The generating function

$$W^{(s)}(t) = \sum_{p \geq 0} W_p^{(s)} t^p \quad (4.1)$$

encodes the numbers $W_p^{(s)}$ of quadrangulations of the sphere with p squares.⁴² In terms of a formal matrix integral over an $N \times N$ Hermitian matrix X , we have

$$e^{N^2 W^{(s)}(t)} = \int dX e^{-N(\frac{1}{2} \text{tr} X^2 - \frac{t}{4} \text{tr} X^4)}, \quad (4.2)$$

where it is understood that only planar diagrams are kept.⁴³ The series (4.1) must clearly diverge when $t = 1$, since there is an infinite number of possible quadrangulations of the sphere. One finds

$$W_p^{(s)} = \frac{3^p (2p)!}{2p(p+1)(p+2)p!^2}. \quad (4.3)$$

When $p \rightarrow \infty$, which corresponds to large surfaces,

$$W_p^{(s)} \underset{p \rightarrow \infty}{\sim} \frac{1}{2\sqrt{\pi}} \frac{12^p}{p^{7/2}}. \quad (4.4)$$

This implies, in particular, that the series (4.1) has a strictly positive radius of convergence t_* such that

$$\lim_{p \rightarrow \infty} (W_p^{(s)})^{1/p} = \frac{1}{t_*} = 12. \quad (4.5)$$

Continuum limit. This $p \rightarrow \infty$ behavior allows one to define a continuum limit as follows. Consider the generating function at fixed area $W_p^{(s)} t^p$. Introducing the UV lattice spacing ℓ_0 , the area is

$$A = p \ell_0^2. \quad (4.6)$$

The factor t^p can be written as

$$t^p = e^{p \ln t} = e^{-\frac{\Lambda_0}{16\pi} A}, \quad (4.7)$$

where $\Lambda_0 = \frac{16\pi}{\ell_0^2} \ln \frac{1}{t}$ plays the rôle of a bare cosmological constant.⁴⁴ A nontrivial continuum limit

$$p \rightarrow \infty, \quad \ell_0 \rightarrow 0, \quad A = p \ell_0^2 \text{ fixed} \quad (4.8)$$

is then obtained by adjusting the bare cosmological constant in such a way that

$$t \rightarrow t_*^- \quad \text{and} \quad \Lambda = \frac{16\pi}{\ell_0^2} \left(1 - \frac{t}{t_*}\right) \geq 0 \text{ is fixed.} \quad (4.9)$$

The intuition behind this scaling is that, when $t \rightarrow t_*$ from below, the quadrangulations with a very large number of squares dominate and the scaling ensures a finite macroscopic area. Strictly speaking, this intuition is valid only if we are considering quadrangulations of surfaces of genus $h \geq 2$; see Eq. (4.12) below. In the case of the sphere, the $1/p^{7/2}$ asymptotic behavior in (4.4) shows that only the higher moments $\langle p^k \rangle$ for $k \geq 3$ actually diverge when $t \rightarrow t_*$. The continuum limit is then defined after subtracting the piece that is regular when $t \rightarrow t_*$ in the partition function $W^{(s)}(t)$, in such a way that the leading term after subtraction is

$$W_{\text{subt}}^{(s)}(t) \underset{t \rightarrow t_*^-}{\sim} -\frac{4}{15} \left(1 - \frac{t}{t_*}\right)^{5/2}. \quad (4.10)$$

A similar remark applies to the torus topology.

Note that the renormalized cosmological constant obtained in this way is always positive, because the limit has to be taken from below, $t < t_*$. There is no getting around this fact: The generating function (4.1) is ill defined for $t > t_*$, because the number of quadrangulations of the sphere is infinite. As a consequence, one cannot define a continuum limit for which $t \rightarrow t_*$ from above. This is not surprising: For strictly negative Λ , the classical action is not bounded from below, and one expects the theory to be inconsistent.

Note also that the precise value of t_* is irrelevant in the continuum limit. This is consistent with the fact that this value is not universal and depends, for example, on the type of polygons we use to discretize the surface. The only universal property of t_* is that $0 < t_* < 1$. On the other hand, the exponent $7/2$ in (4.4) is universal. The generating function at fixed area (4.4) scales as⁴⁵

⁴¹The continuum limit is universal and does not depend on the type of polygons we use. We consider squares for simplicity.

⁴²The precise definition of $W_p^{(s)}$ involves symmetry factors attached to each discretized surface, according to their automorphism group. This is similar to the discussion in Sec. III B 2; see also Sec. V. In the continuum limit, these symmetry factors play a trivial rôle.

⁴³Note that we use here the “dual” point of view, as defined at the beginning of Sec. II A, the matrix model vertex $\text{tr} X^4$ being associated with a square of the sphere quadrangulation.

⁴⁴The factors of 16π are included to be consistent with the usual conventions for the cosmological constant used in the context of JT gravity; see, e.g., Eq. (2.49).

⁴⁵Here and below, the proportionality sign $\propto$ indicates that the generating functions in the continuum limit are fixed up to a trivial, parameter-independent, multiplicative constant, that may involve powers of the cutoff ℓ_0 .

$$W^{(s)}(A) \propto A^{-7/2} e^{-\frac{\Lambda}{16\pi} A}. \quad (4.11)$$

This is a famous result, which can be generalized to random closed surfaces of arbitrary genus h :

$$W^{(h)}(A) \propto A^{-1+(\gamma_{\text{str}}-2)(1-h)} e^{-\frac{\Lambda}{16\pi} A}. \quad (4.12)$$

The exponent γ_{str} , called the string susceptibility, is $\gamma_{\text{str}} = -1/2$ in the case of pure gravity and takes the more general KPZ form

$$\gamma_{\text{str}} = \frac{1}{12} (c - 1 - \sqrt{(c-1)(c-25)}) \quad (4.13)$$

when two-dimensional gravity is coupled to an arbitrary matter conformal field theory (CFT) of central charge $c < 1$. There is a so-called barrier at $c = 1$, beyond which Eq. (4.13) predicts a nonsensical imaginary string susceptibility. At $c = 1$, logarithmic corrections to the scaling law (4.12) are believed to be present [62], with

$$W^{(s)}(A) \propto A^{-3} (\ln A)^{-2} e^{-\frac{\Lambda}{16\pi} A}. \quad (4.14)$$

2. Pure Liouville gravity on the disk

To come closer in spirit to the problem we are studying in the present paper, let us make one more step in the framework of Liouville gravity and review the case of the disk topology. On the disk, we can define in Liouville gravity the exact analogs of the JT gravity generating functions (2.19) and (2.21). To avoid confusion and for the sake of consistency, we continue to use the notation W for JT gravity and use instead $\mathbf{W}$ for Liouville gravity. The difference between $\mathbf{W}$ and W is that, in the first case, we sum over all possible discretized disk metrics, whereas in the second case we restrict ourselves to the metrics of zero curvature.⁴⁶

Let us mention right away that thinking that JT has to be simpler than Liouville because we have to take into account less metrics is naïve and probably misleading: Problems with constraints are typically more difficult.⁴⁷

Discretized formulation. From direct combinatorial counting *à la* Tutte or from the solution of the matrix model, the Liouville disk generating function can be computed exactly, with only slightly more effort than in the case of the sphere. The number of quadrangulations of the disk of area p and boundary length $2n$ is found to be

$$W_{2n,p} = 3^p \frac{(2n-1)!(2p+n-1)!}{n!(n-1)!(p+n+1)!p!}. \quad (4.15)$$

As a consistency check, note that $W_{4,p-1}$ counts the number of quadrangulations of the sphere with $p-1$ unmarked and one marked squares, the marked square corresponding to the boundary of length four. This implies that $W_{4,p-1} = pW_p^{(s)}$, where $W_p^{(s)}$ is given in Eq. (4.3). This is indeed satisfied.

The generating function at fixed boundary length is found to be

$$\begin{aligned} W_{2n}(t) &= \sum_{p \geq 1} W_{2n,p} t^p \\ &= \frac{(2n-1)!}{2^{2n-1} n! (n+2)!} a^{2n}(t) \left(n+1 - \frac{1}{8} n a^2(t) \right) \\ &\quad - \frac{(2n-1)!}{n! (n+1)!}, \end{aligned} \quad (4.16)$$

where

$$a^2(t) = \frac{2}{3t} (1 - \sqrt{1-12t}) \quad (4.17)$$

is a solution of $3ta^4 - 4a^2 + 16 = 0$.⁴⁸

From the point of view of the matrix model (4.2), it is interesting to note that the generating function $W_{2n}(t)$ is a simple expectation value:

$$W_{2n}(t) = \frac{1}{2n} \left\langle \frac{1}{N} \text{tr} X^{2n} \right\rangle. \quad (4.18)$$

Indeed, in the dual Feynman graph point of view, the insertion of $\text{tr} X^{2n}$ corresponds to a marked face, which is removed to get the disk topology. The factor $1/(2n)$ takes into account the fact that all the edges of the boundary are equivalent (there is no marked edge on the boundary).⁴⁹ One can also show straightforwardly that the full generating function

$$W(t, g) = \sum_{n \geq 1} W_{2n}(t) g^{2n} \quad (4.19)$$

satisfies

$$\begin{aligned} g \frac{\partial W}{\partial g} &= \sum_{n \geq 1} 2n W_{2n}(t) g^{2n} \\ &= \frac{1}{2g^2} \left[1 - \frac{t}{g^2} + t \left(\frac{1}{g^2} + \frac{a^2(t)}{2} - \frac{1}{t} \right) \sqrt{1 - a^2(t)g^2} \right]. \end{aligned} \quad (4.20)$$

⁴⁶Or more generally to constant curvature; see Secs. III E 4 and VI.

⁴⁷A famous illustration of this “principle” is the Ising model. The exact Onsager solution of the model on a fixed flat lattice is much more difficult than the solution on a random lattice [63].

⁴⁸The interval $[-a, a]$ represents the support of the planar matrix model eigenvalue distribution.

⁴⁹For the same reason, there is a factor $1/4$ in the interaction term in the matrix integral (4.2).

These results show that $W_{2n}(t)$ has a finite radius of convergence $t_* = 1/12$, matching with the radius of convergence of the sphere generating function.⁵⁰ Similarly, one may check that $\tilde{W}_p(g) = \sum_{n \geq 1} W_{2n,p} g^{2n}$ has a radius of convergence $g_* = 1/2$ and that $\sum_{n \geq 1} W_{2n}(t) g^{2n}$ has a radius of convergence $g_*(t) = 1/a(t)$, etc.

Continuum limit. As in the case of the sphere, we can define a universal continuum limit. Let us work with the generating function at fixed boundary length $W_{2n}(t)g^{2n}$. In terms of the lattice spacing ℓ_0 , the boundary length is

$$\ell = 2n\ell_0. \quad (4.21)$$

The factor g^{2n} plays the rôle of a bare boundary cosmological constant:

$$g^{2n} = e^{2n \ln g} = e^{-\lambda_0 \ell}, \quad (4.22)$$

with $\lambda_0 = (\ln \frac{1}{g})/\ell_0$. The continuum limit is

$$n \rightarrow \infty, \quad \ell_0 \rightarrow 0, \quad \ell = 2n\ell_0 \text{ fixed}, \quad (4.23)$$

adjusting at the same time the bare bulk cosmological constant t as in (4.9), which is equivalent to setting

$$\frac{t}{t_*} = 1 - \frac{\ell^2 \Lambda}{64\pi n^2}. \quad (4.24)$$

In this limit, $W_{2n}(t)$ behaves as

$$W_{2n}(t) \underset{n \rightarrow \infty}{\sim} \frac{1}{\sqrt{\pi}} \frac{8^n}{n^{7/2}} e^{-\frac{\ell\sqrt{\Lambda}}{8\sqrt{\pi}}} \left(1 + \frac{\ell\sqrt{\Lambda}}{8\sqrt{\pi}}\right). \quad (4.25)$$

The exponential growth 8^n can be canceled, as usual, by adjusting the bare boundary cosmological constant g to its critical value $g_*(t_*) = g_* = 1/a(t_*) = 1/(2\sqrt{2})$:

$$g \rightarrow g_*, \quad \lambda = \frac{1}{\ell_0} \left(1 - \frac{g}{g_*}\right) \geq 0 \text{ fixed}. \quad (4.26)$$

We then get from (4.25) that $W_{2n}(t)g^{2n}$ scales as

$$W(\ell, \Lambda) \propto \ell^{-7/2} e^{-\lambda\ell - \frac{\sqrt{\Lambda}}{8\sqrt{\pi}}\ell} \left(1 + \frac{\ell\sqrt{\Lambda}}{8\sqrt{\pi}}\right). \quad (4.27)$$

We could absorb the Λ -dependent contribution in the exponential by redefining λ , but we prefer to keep the formula in this way, with the parameter Λ directly coupling to the area; see below.

Note that, as we have already noticed in the case of the sphere topology, the theory makes sense only if $\Lambda \geq 0$. For

$\Lambda < 0$, even on a disk of fixed boundary length, the theory is unstable. As explained in Sec. II D 5 and further discussed in Sec. VI and [40], this is similar to the expected behavior of JT gravity in the positive curvature case but in sharp contrast to the case of negative curvature, for which JT gravity is believed to exist for all positive and negative values of the cosmological constant. The behavior in zero curvature is probably intermediate, the theory existing for all positive and at least for some range of negative values of the cosmological constant.

Bulk and boundary scalings. An important and interesting feature of Liouville gravity is that the relative scaling of the bulk area and the boundary length is “classical” in the continuum limit, in the sense that the area scales as the square of the boundary length. For instance, working at fixed boundary length as above, the existence of the moments of the bulk area

$$\langle A^k \rangle = \frac{\ell_0^2}{W_{2n}(t)} \frac{\partial^k W_{2n}(t)}{\partial \ln t^p} = \ell_0^{2k} \langle p^k \rangle \quad (4.28)$$

in the continuum limit (4.23) and (4.24) implies that the expectation values $\langle p^k \rangle$ scale as n^{2k} in this limit, for any integer $k \geq 1$. This can be checked explicitly from Eq. (4.16). The expectation values $\langle A^k \rangle$ can actually be computed straightforwardly in the continuum from (4.27), for instance,

$$\langle A \rangle = -\frac{16\pi}{W} \frac{\partial W}{\partial \Lambda} = \frac{\ell^2}{8(1 + \frac{\ell\sqrt{\Lambda}}{8\sqrt{\pi}})}. \quad (4.29)$$

One could also take the point of view of the generating function for which both the area and the boundary length are fixed [Eq. (4.15)]. The above discussion implies that there exists a nontrivial continuum limit $\ell_0 \rightarrow 0$ for which both the area and the boundary length are finite. In other words, we take $n \rightarrow \infty$ and $p \rightarrow \infty$, with

$$\frac{p}{(2n)^2} = \frac{A}{\ell^2} \text{ fixed}. \quad (4.30)$$

Dealing with the usual exponential divergence by renormalizing the bare bulk and boundary cosmological constant using (4.9) and (4.26), we get the continuum partition function at fixed ℓ and A :

$$t^p g^{2n} W_{2n,p} \underset{\substack{n,p \rightarrow \infty \\ p=4n^2 A/\ell^2}}{\propto} W(A, \ell) = \frac{e^{-\frac{\Lambda}{16\pi} A - \lambda\ell - \frac{\ell^2}{16A}}}{\ell^{1/2} A^{5/2}}. \quad (4.31)$$

By integration,

⁵⁰This fact and its generalizations are easily understood in the matrix model framework.

$$W(\ell, \Lambda) = \int_0^\infty W(A, \ell) dA, \quad (4.32)$$

we find again (4.27).

In the continuum formulation of Liouville gravity, the bulk and boundary cosmological constant operators are represented by renormalized bulk $:e^{\gamma_B \varphi}:$ and boundary $:e^{\gamma_b \varphi}:$ exponentials of the Liouville field φ for which the boundary constant γ_b is equal to half the bulk constant γ_B , $\gamma_b = \frac{1}{2}\gamma_B$.⁵¹ This implies the classical scaling $A \propto \ell^2$.

As we have already emphasized in Sec. II D 3 and in part of the discussion in Sec. II E, the situation in JT gravity is completely different. The boundary cosmological operator is still represented by a renormalized boundary exponential $:e^{\gamma_b \varphi}:$ of the boundary Liouville field, with an associated renormalized quantum boundary length $\beta_q = \int :e^{\gamma_b \varphi}: d\theta$, but the bulk operator is no longer renormalized, due to the constant curvature constraint. This yields a quantum-corrected scaling $A \propto \beta_q^{2\nu}$, for a nontrivial critical exponent $\nu = 1/d_H$ which is the inverse of the Hausdorff dimension of the boundary [16,28].

Generalization. Based on the fact $\ell\sqrt{\Lambda}$ is the only nontrivial dimensionless parameter in the model at fixed boundary length, one expects that the scaling limit of the generating function is of the general form

$$W(\ell, \Lambda) \propto \ell^{\alpha-3} e^{-\lambda\ell} f\left(\frac{\ell\sqrt{\Lambda}}{8\sqrt{\pi}}\right), \quad (4.33)$$

for an exponent α and a function f which are *a priori* unknown. For pure gravity, Eq. (4.27) yields $\alpha = -1/2$ and $f(z) = e^{-z}(1+z)$. In spite of its very simple form, it is actually quite challenging to derive this result from a direct continuum approach within Liouville theory [8,13,15,60]. Strong arguments [15] show that, in the more general case for which gravity is coupled to an arbitrary CFT of central charge $c < 1$, α matches with the string susceptibility exponent γ_{str} given by (4.13) and

$$W(\ell, \Lambda) \propto \ell^{\gamma_{\text{str}}-3} e^{-\lambda\ell} (\ell\sqrt{\Lambda})^{1-\gamma_{\text{str}}} K_{1-\gamma_{\text{str}}}\left[\frac{\ell\sqrt{\Lambda/(16\pi)}}{\sqrt{\sin\frac{\pi}{1-\gamma_{\text{str}}}}}\right], \quad (4.34)$$

where K is the modified Bessel function of the second kind.

For pure gravity, (4.34) is consistent with (4.27), modulo a multiplicative constant in the definition of $\ell\sqrt{\Lambda}$:

$$(\ell\sqrt{\Lambda})_{\text{sq}} = \frac{2\sqrt{2}}{3^{1/4}} (\ell\sqrt{\Lambda})_{\text{cont}}. \quad (4.35)$$

⁵¹This condition comes from the fact that the bulk operator $:e^{\gamma_B \varphi}:$ must be of conformal dimension 2, whereas the boundary operator $:e^{\gamma_b \varphi}:$ must be of conformal dimension 1.

The subscript “sq” on the left-hand side refers to the quantity defined within the discretized approach using quadrangulations [Eqs. (4.6) and (4.21)], whereas the subscript “cont” on the right-hand side refers to the quantity defined directly within the continuum Liouville approach used in [15]. A finite multiplicative renormalization as in (4.35) is harmless. It simply shows that the choice of lattice model in Liouville affects renormalized lengths only by a multiplicative factor in the continuum limit.

Semiclassical limit. The existence of a semiclassical regime in which the fluctuations of the geometry are tamed and the physics is dominated by a smooth classical geometry is a desirable feature of any interesting quantum gravity model.

Pure Liouville quantum gravity does not have a semiclassical limit. On the disk, one may want to consider $\Lambda \rightarrow +\infty$ at fixed ℓ , but this limit is dominated by singular geometries for which the disk has a branched polymer structure. As for the limit $\Lambda \rightarrow -\infty$, it does not make sense since the model is unstable for $\Lambda < 0$. Nevertheless, as noticed by Zamolodchikov in [64], when Liouville gravity is coupled to conformal matter of central charge c , the limit $c \rightarrow -\infty$ does correspond to a semiclassical limit. Indeed, at large $|c|$, the Liouville action, that governs the gravitational dynamics, is multiplied by a constant proportional to $-c$. To obtain an interesting limit, one must also rescale the cosmological constant in such a way that the cosmological constant term in the action is proportional to $|c|$ as well. The semiclassical limit of Liouville gravity on the disk, thus, corresponds to

$$c \rightarrow -\infty, \quad \Lambda \rightarrow +\infty, \quad \Lambda/|c| = \frac{2}{3}\mu \text{ and } \ell \text{ fixed.} \quad (4.36)$$

The factor $2/3$ in the definition of μ is added because it allows us to simplify some formulas below. The semiclassical loopwise expansion of $W(\ell, \Lambda)$ takes the form

$$\ln W(\ell, \Lambda) = -\lambda\ell + (\gamma_{\text{str}} - 3) \ln \frac{\ell}{2\pi} + \sum_{L \geq 0} |c|^{1-L} f_L\left(\frac{\ell\sqrt{\mu}}{2\pi}\right), \quad (4.37)$$

with, according to (4.13),

$$\gamma_{\text{str}} - 3 = -\frac{|c|}{6} - \frac{25}{6} + \frac{6}{|c|} + O(1/|c|^2). \quad (4.38)$$

The contributions f_L at any loop order can be straightforwardly obtained from the exact formula (4.34) and the asymptotic expansion of the Bessel function for large argument

$$K_\nu(\nu z) \underset{\nu \rightarrow \infty}{=} \sqrt{\frac{\pi}{2\nu}} \frac{e^{-\nu h(z)}}{(1+z^2)^{1/4}} \sum_{k \geq 0} (-1)^k U_k \left(\frac{1}{\sqrt{1+z^2}} \right) \nu^{-k}, \quad (4.39)$$

where

$$h(z) = \sqrt{1+z^2} + \ln \frac{z}{1+\sqrt{1+z^2}} \quad (4.40)$$

and the U_k are polynomials of degree $3k$ that can be computed recursively, for example,

$$U_0(x) = 1, \quad U_1(x) = \frac{1}{24}(3x - 5x^3), \quad \text{etc.} \quad (4.41)$$

Normalizing W in such a way that

$$W(\ell = 2\pi, \mu = 0, c) = e^{-\lambda \ell}, \quad (4.42)$$

we find, up to two loops,

$$f_0(z) = \frac{1}{6} \ln \frac{1+\sqrt{1+z^2}}{2} - \frac{1}{6} \left(\sqrt{1+z^2} - 1 \right), \quad (4.43)$$

$$f_1(z) = \frac{13}{6} \ln \frac{1+\sqrt{1+z^2}}{2} - \frac{13}{12} \left(\sqrt{1+z^2} - 1 \right) - \frac{1}{4} \ln(1+z^2), \quad (4.44)$$

$$f_2(z) = -6 \ln \frac{1+\sqrt{1+z^2}}{2} + \frac{313z^4 + 759z^2 + 506}{48(1+z^2)^{3/2}} - \frac{253}{24} + \frac{13}{4} \frac{z^2}{1+z^2} - \frac{\pi^2}{2} \left(\sqrt{1+z^2} - 1 \right). \quad (4.45)$$

One can check explicitly that the leading terms in (4.37), corresponding to the contribution $-|c|/6$ in (4.38) and to f_0 in (4.43), precisely match with the expected result obtained from a classical disk geometry; see Appendix A. In principle, the higher loop contributions can be obtained from the path integral following the same logic as in [64]. This is checked explicitly in [29], where the analog of the limit (4.36) is also studied for JT gravity.

B. Review of random paths models

Our goal in this subsection is to summarize the standard framework in which models of random polygons, which yield theories of random closed paths in the continuum limit, are usually analyzed. We, thus, consider a general model of random polygons on the square lattice. As we shall see, there are similarities, but also crucial differences, with the Liouville models of random metrics reviewed in the previous subsection. Understanding this physics in detail is important in order to elucidate which aspects, some more metriclike, others more pathlike, are likely to be

relevant for the new model of SOPs we are primarily interested in; see Sec. IV C. Indeed, as already emphasized many times, the most characteristic property of this model is to have dual interpretations, either in terms of random polygons or in terms of random metrics.

Our discussion is illustrated on the examples of arbitrary random polygons (RPs), SAPs, and convex polygons (CPs), which may all be seen as toy approximate models for flat JT gravity. The RP model is the flat version of the model proposed by Kitaev and Suh in negative curvature [32]. The SAP model is the one studied in Stanford and Yang in [34]. The CP model shares some properties with the one-loop exact reparametrization *Ansatz* used in the near-hyperbolic limit of negative curvature JT gravity.

1. Generalities

Let us first note that, for an arbitrary polygon, there is no unique canonical notion of area; see, e.g., Eq. (3.10). For this reason, the case of the RP model is a bit special; see Sec. IV B 4. For the time being, we limit ourselves to models like CP, SAP, or SOP for which the notion of area is unambiguous.

We note $W_{2n,p}^P$ the number of polygons of length $2n$ and area p in the polygon model under consideration. The numbers $W_{2n,p}^P$ may include a multiplicity factor, as in the case of SOP; see Sec. III E. We introduce, as usual, the generating functions at fixed length and area

$$W_{2n}^P(t) = \sum_{p \geq 1} W_{2n,p}^P t^p, \quad \tilde{W}_p^P(g) = \sum_{n \geq 1} W_{2n,p}^P g^{2n} \quad (4.46)$$

and the “grand canonical” generating function

$$W^P(t, g) = \sum_{n \geq 1} W_{2n}^P(t) g^{2n} = \sum_{p \geq 1} \tilde{W}_p^P(g) t^p. \quad (4.47)$$

The parameter t , or more precisely its logarithm, plays the rôle of a “bare bulk cosmological constant.” Note that, in the context of random polygon models, one may rather talk about “pressure,” with

$$P = -\frac{\Lambda}{16\pi}. \quad (4.48)$$

This terminology is more natural when one uses the polygons to model pressurized vesicles, cell membranes, or polymer chains.

To develop our analysis, we assume that the following properties are satisfied.

Hypothesis 4.1. There exist strictly positive constants a and b such that $W_{2n,p}^P = 0$ unless $2bn \leq p \leq a(2n)^2$.

In our models, this follows from the isoperimetric inequality, which ensures that for a given n there is a maximum possible value $p = p_{\max}(n) = [\frac{1}{4}n^2]$, and from

Eq. (2.18), which ensures that for a given p there is a maximum value $n = n_{\max}(p) = p + 1$.

A simple important consequence is that $W_{2n}^P(t)$ is a polynomial in t . In particular, the number

$$w_{2n}^P = W_{2n}^P(1) = \sum_{p \geq 1} W_{2n,p}^P \quad (4.49)$$

of polygons of a given length is finite. *These properties, as well as other crucial consequences described below, are not valid in Liouville gravity.*

An important physical feature is the typical “size” of the polygons. This can be measured by studying the area distribution, whose expected moments are given in terms of the derivatives of the generating function as

$$\langle p^r \rangle_{2n}(t) = \frac{1}{W_{2n}^P} \frac{d^r W_{2n}^P}{d(\ln t)^r}. \quad (4.50)$$

Let

$$w_{2n;r}^P = \sum_{p \geq 1} p^r W_{2n,p}^P, \quad (4.51)$$

generalizing (4.49). We assume that, at large n ,

$$w_{2n;r}^P \underset{n \rightarrow \infty}{\sim} a_r g_*^{-2n} (2n)^{2\nu(r-\vartheta)-1}, \quad (4.52)$$

for all $r \geq 0$, for certain strictly positive constants a_r and g_* , and critical exponents ν and ϑ . In particular,

$$\langle p^r \rangle_{2n}(t=1) = \frac{w_{2n;r}^P}{w_{2n}^P} \underset{n \rightarrow \infty}{\sim} \frac{a_r}{a_0} (2n)^{2\nu r}. \quad (4.53)$$

The asymptotic scaling (4.52) is found in many models, by analytical methods or numerically. We conjecture that this large- n scaling will be valid for SOPs as well, modulo possible logarithmic corrections.

Observables other than the area can also be used to measure the size of polygons. A widely used quantity is the radius of gyration R_γ . For an arbitrary polygon γ made of $\mathcal{N}$ points located at $w_i \in \mathbb{C}$, noting $w_\gamma = \frac{1}{\mathcal{N}} \sum_i w_i$ the “center of mass,” it is defined by

$$R_\gamma^2 = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} |w_i - w_\gamma|^2 = \frac{1}{\mathcal{N}^2} \sum_{i < j} |w_i - w_j|^2, \quad (4.54)$$

where the second equality can be checked by an elementary calculation. One could also consider the diameter $D_\gamma = \max_{i,j} |w_i - w_j|$ and several other quantities of the same type. It is natural to expect that the moments of these quantities will have scalings similar to (4.53), e.g.,

$$\langle R^{2r} \rangle \underset{n \rightarrow \infty}{\propto} (2n)^{2\nu r}. \quad (4.55)$$

If the notion of size provided by the area and the radius of gyration are the same, then $\nu' = \nu$, which, as far as we know, happens in all known models for which both exponents have been studied.

From the above, one can easily derive a few important general properties that the generating functions must satisfy.

- (i) The constant g_* appearing in (4.52) is the radius of convergence of the series $\sum_{n \geq 1} w_{2n}^P g_*^{2n}$. It is such that $0 < g_* \leq 1$. The quantity $1/g_*$ is usually called the *connective constant* of the model. Its precise value is not universal (it depends on the lattice) and is, thus, not of great interest in physics, but computing it is a challenging and interesting problem in mathematics; see, e.g., [65] for the case of SAP.⁵²
- (ii) For $t \leq 1$, the series $\sum_{n \geq 1} W_{2n}^P(t) g_*^{2n}$ has a strictly positive radius of convergence $g_*(t)$, which is a decreasing function of t , with $g_*(t=1) = g_*$. One shows that $g_*(t) \geq g_* t^{-b}$, where b is the constant appearing in Hypothesis 4.1. In our models, with $n_{\max}(p) = p + 1$, we actually have $g_*(t) \geq g_*/\sqrt{t}$. In particular, $\lim_{t \rightarrow 0} g_*(t) = +\infty$. When $t > 1$, the radius of convergence is $g_*(t) = 0$.
- (iii) The series $\sum_{p \geq 1} \tilde{W}_p^P(g_*) t^p$, as well as the derivatives of this series with respect to t , to any order, has radius of convergence $t_* = 1$ and, moreover, $\lim_{t \rightarrow 1^-} \partial_t^r W^P(t, g_*) = \partial_t^r W^P(1, g_*)$ for any $r \geq 0$.
- (iv) The critical exponent ν is the inverse of the Hausdorff fractal dimension of the curves, $d_H = 1/\nu$, and satisfies

$$\frac{1}{2} \leq \nu \leq 1. \quad (4.56)$$

For instance, the values $\nu = 1/2$, $\nu = 3/4$, and $\nu = 1$ corresponds to purely random loops, self-avoiding loops, and smooth loops (realized for convex polygons), respectively. More generally, SLE $_\kappa$ curves have $\nu = \max((1 + \kappa/8)^{-1}, 1/2)$. For JT quantum gravity coupled to a matter CFT of central charge $c \leq 0$, the continuum formulation developed in [16,28] implies that

$$\nu = \frac{1}{2} \left[1 + \sqrt{\frac{c}{c-24}} \right]. \quad (4.57)$$

This formula predicts that the exponent is a strictly decreasing function of c that goes to one when $c \rightarrow -\infty$, consistently with the idea that this limit is semiclassical [29,64], and is equal to $1/2$ for the

⁵²For SAPs on a square lattice, $1/g_* \simeq 2.638$.

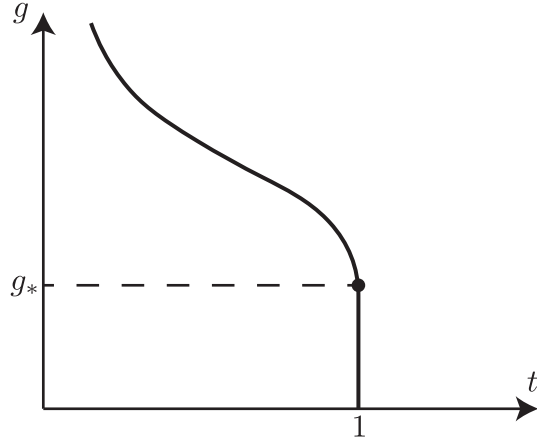


FIG. 16. Phase diagram of a random closed paths model in the grand canonical ensemble (from [46]). The representative curve of the function $g_*(t)$ is plotted. Above and on the right of the curve, the grand canonical generating function $W^P(t, g)$ does not exist. The thick dot represents the critical point ($t_* = 1, g_*$) which plays a crucial rôle in defining the continuum limit of the model.

pure gravity theory, $c = 0$, on which we focus in the present paper. So the SOP model has the same critical exponent ν , or, equivalently, the same Hausdorff dimension, as the random walks.

2. Phases

The phase diagram in the (t, g) plane is depicted in Fig. 16. If we work at fixed length n , we can distinguish three regimes.

- (i) The *deflated regime* occurs when $t < 1$ (positive cosmological constant, negative pressure) and corresponds to a large- n linear growth of the free energy $\ln W_{2n}^P(t) \sim -2n \ln g_*(t)$. In particular, the average area

$$\langle p \rangle_{2n}(t) = \frac{d \ln W_{2n}^P}{d \ln t} \underset{n \rightarrow \infty}{\sim} -2n \frac{t g'_*(t)}{g_*(t)} \quad (4.58)$$

is proportional to n . This indicates that the typical polygon is treelike or, equivalently, that the deflated regime is a branched polymer phase.

- (ii) The *critical point* occurs at $t = 1$ (zero cosmological constant or pressure). It is characterized by the scaling laws (4.52) and (4.53). As we shall see below, the critical point is instrumental in defining the continuum limit and the continuum limit is entirely determined by the properties at the critical point.
- (iii) The *inflated regime* occurs when $t > 1$ (negative cosmological constant, positive pressure). Using Hypothesis 4.1, one finds that, if $t > 1$,

$$t^{a(2n)^2} \leq W_{2n}^P(t) \leq w_{2n}^P t^{a(2n)^2}. \quad (4.59)$$

Using (4.52) for $r = 0$ [or the weaker condition $\ln w_{2n}^P = o(n^2)$], we get

$$\lim_{n \rightarrow \infty} \frac{1}{(2n)^2} \ln W_{2n}^P(t) = a \ln t. \quad (4.60)$$

In particular, recalling that $a = 1/16$ for our models,

$$\langle p \rangle_{2n}(t) \underset{n \rightarrow \infty}{\sim} 4an^2 = \frac{n^2}{4} \quad (4.61)$$

matches with the maximum area allowed by the isoperimetric inequality: The polygons are inflated “balloons” in this phase.

The very simple result (4.60) is valid when the limit $n \rightarrow \infty$ is taken at fixed $t > 1$. However, if $t - 1$ is scaled appropriately with n at large (but finite) n , a more interesting behavior is generally expected, based, for example, on numerical simulations; see, e.g., [66] for a study of the case of SAPs. It is useful to distinguish two ranges of parameters in the inflated phase.

- (a) The region $0 < n^{2\nu}(t - 1) \ll n^{2\nu-1}$ corresponds to an inflated “crumpled” phase. If $\nu > 1/2$, one can distinguish a weakly inflated regime, corresponding to $n^{2\nu}(t - 1)$ of the order of one, and a strongly inflated regime, for $1 \ll n^{2\nu}(t - 1) \ll n^{2\nu-1}$.
- (b) For $\nu > 1/2$ and $n^{2\nu}(t - 1) \geq n^{2\nu-1}$, we have a fully inflated phase, characterized by $\langle p \rangle \propto n^2$.

The existence of the inflated crumpled phase plays an important rôle; see the next subsection on the continuum limit. The fact that the case $\nu = 1/2$ is special also has important qualitative consequences; see Sec. IV B 4 for an example and the conjecture is Sec. IV C 4.

Remark. The transition at $t = t_* = 1$ is reminiscent of a deconfining or of a Hagedorn phase transition, in the sense that the large- n scaling of the free energy changes, here from $\ln W_{2n}^P(t) \propto n$ for $t < 1$ to $\ln W_{2n}^P(t) \propto n^2$ for $t > 1$.

3. Continuum limit

A very important property of the random paths models is that their continuum limit can be defined starting from the critical point at $t = t_* = 1$, i.e., for zero cosmological constant, and then possibly deform away from it, by turning on the renormalized cosmological constant.

Critical theory. To define the continuum limit, we introduce as usual a cutoff ℓ_0 . The dimensionful area is then given by $A = p\ell_0^2$ and Eq. (4.53) yields

$$\langle A^r \rangle = \frac{a_r}{a_0} (2n\ell_0^{1/\nu})^{2\nu r}. \quad (4.62)$$

The continuum limit is then defined by taking

$$n \rightarrow \infty, \quad \ell_0 \rightarrow 0, \quad 2n\ell_0^{1/\nu} = \beta_q \text{ fixed.} \quad (4.63)$$

The parameter β_q is the “quantum length” of the polygons in the continuum limit, to which we have already referred several times in previous sections. Note that it is proportional to the microscopic length n , and yet its engineering length dimension is $1/\nu$. The scaling (4.63) yields well-defined moments for the area:

$$\langle A^r \rangle = \frac{a_r}{a_0} \beta_q^{2\nu r}. \quad (4.64)$$

Under mild assumptions (see the remarks at the end of this subsection), these moments determine uniquely the area distribution law at zero cosmological constant, noted $\rho_{\beta_q}^P$, which is such that

$$\langle A^r \rangle = \int_0^\infty A^r \rho_{\beta_q}^P(A) dA. \quad (4.65)$$

The continuum generating function at zero cosmological constant, $W^P(\beta_q, \Lambda = 0)$, is determined by the asymptotics of w_{2n}^P , given by (4.52) for $r = 0$, together with the scaling (4.63). Considering, as usual [compare with Eq. (4.26)], $g^{2n} w_{2n}^P$, for a bare parameter g adjusted in such a way that

$$g \rightarrow g_*, \quad \lambda = \frac{1}{\ell_0^{1/\nu}} \left(1 - \frac{g}{g_*}\right) \text{ fixed}, \quad (4.66)$$

and discarding an overall factor $\ell_0^{2\vartheta+1/\nu}$ of the cutoff, we get

$$W^P(\beta_q, \Lambda = 0) = a_0 e^{-\beta_q \lambda} \beta_q^{-1-2\nu\vartheta}. \quad (4.67)$$

The contribution $e^{-\beta_q \lambda}$ is a “boundary cosmological counterterm” that may be adjusted at will. It is important to emphasize here that the logarithm of this term is proportional to the *quantum* length β_q . In particular, the boundary cosmological constant λ has dimension of length to the power $-1/\nu$. This observation may be unremarkable from the point of view of a standard random path model, but it is nontrivial for the SOP model, which has a gravitational interpretation: It shows that the naive gravitational boundary length counterterm, which is ill defined in the quantum theory because the boundary is fractal, should be replaced by a quantum boundary length counterterm.

Turning on Λ . The generating function for nonzero cosmological constant is obtained by using the area distribution as

$$W^P(\beta_q, \Lambda) = W^P(\beta_q, 0) \int_0^\infty e^{-\frac{\Lambda}{16\pi} A} \rho_{\beta_q}^P(A) dA. \quad (4.68)$$

This can be put in a form similar to (4.33):

$$W^P(\beta_q, \Lambda) = a_0 e^{-\beta_q \lambda} \beta_q^{-1-2\nu\vartheta} f_P\left(\frac{\Lambda \beta_q^{2\nu}}{16\pi}\right), \quad (4.69)$$

for

$$f_P(z) = \frac{1}{a_0} \sum_{r \geq 0} \frac{(-1)^r a_r}{r!} z^r. \quad (4.70)$$

The continuum generating function $W^P(\beta_q, \Lambda)$ may also be obtained directly from $g^{2n} w_{2n}^P(t)$ defined in (4.46), in a way similar to what is done in Liouville theory, by using the scalings (4.63) and (4.66) and

$$t \rightarrow t_* = 1, \quad \Lambda = \frac{16\pi}{\ell_0^2} (1 - t) \text{ fixed}, \quad (4.71)$$

which is exactly as in (4.24).

The representation (4.68) shows that the models are always well defined for positive cosmological constant, $\Lambda \geq 0$. On the other hand, the existence for negative cosmological constant is not *a priori* ensured. It depends on the asymptotic behavior of the area probability distribution $\rho_{\beta_q}^P(A)$ for large A .

The discussion at the end of Sec. IV B 2, which applies to the regime $\Lambda < 0$ for models defined on a flat lattice, shows that when $1/2 < \nu < 1$ the continuum limit selects the crumpled phase. It is then natural to conjecture that such models will exist for all values of Λ , positive or negative. When $\nu = 1/2$, the situation is less clear, and one may expect an upper bound on the possible value of $-\Lambda$; see, for example, the example in Sec. IV B 4 below.

In the case of JT gravity, the asymptotic behavior of the density ρ_{β_q} and, thus, the existence of the models for negative Λ is expected to strongly depend on whether the curvature of space-time is positive, zero, or negative; see Secs. IV C 4 and VI and [40].

Grand canonical ensemble and scaling form. It is also useful to discuss the grand canonical continuum generating function $W^P(\lambda, \Lambda)$, which is the continuum limit of the generating function $W^P(t, g)$ defined in (4.47). One has⁵³

$$W^P(\lambda, \Lambda) = \frac{1}{2} \int_{2\ell_0^{1/\nu}}^\infty W^P(\beta_q, \Lambda) d\beta_q. \quad (4.72)$$

We have reintroduced the cutoff ℓ_0 in⁵⁴ (4.72) in order to regulate the divergence of the integral at $\beta_q = 0$ which, from (4.69), occurs when $\nu\vartheta \geq 0$. From (4.69) and (4.70),

⁵³Note that the factor $e^{-\beta_q \lambda}$ is conventionally included in $W^P(\beta_q, \Lambda)$; see Eq. (4.69). Note also that the sum over n should be replaced by an integral over β_q with the measure $d\beta_q / (2\ell_0^{1/\nu})$, according to the definition (4.63). We also discard any overall power of ℓ_0 , as usual.

⁵⁴The way we do it is natural because, from (4.63), $2\ell_0^{1/\nu}$ is the minimal value of β in the microscopic theory.

we get

$$W^P(\lambda, \Lambda) = \frac{1}{2} \left(\frac{\Lambda}{16\pi} \right)^\vartheta \mathcal{F} \left(\left(\frac{16\pi}{\Lambda} \right)^{\frac{1}{2\nu}} \lambda \right) + \text{diverging piece}, \quad (4.73)$$

where

$$\mathcal{F}(z) = \sum_{r \geq 0} \mathcal{F}_r z^{-2\nu(r-\vartheta)}, \quad \mathcal{F}_r = \frac{(-1)^r}{r!} \Gamma(2\nu(r-\vartheta)) a_r. \quad (4.74)$$

The diverging piece can be straightforwardly evaluated; if α is not a negative integer, it is a polynomial in λ of degree $[2\nu\vartheta]$. Using (4.66) and (4.71), one sees that Eq. (4.73) follows from the scaling form:

$$W^P(t, g) = W_{\text{reg}}^P(t, g) + \frac{1}{2} (1-t)^\vartheta \mathcal{F} \left(\frac{1-g/g_*}{(1-t)^{\frac{1}{2\nu}}} \right), \quad (4.75)$$

which is valid near the critical point $(t, g) = (1, g_*)$, when $t \rightarrow 1^-$ and $g \rightarrow g_*^-$. The regular part is a nonuniversal piece that plays no role in the continuum limit. Using the expansion (4.74) in (4.75), we also get the critical behavior

$$\frac{1}{r!} \frac{\partial^r W^P}{\partial t^r} (t=1, g) \underset{g \rightarrow g_*^-}{\sim} \frac{\frac{1}{2} (-1)^r \mathcal{F}_r}{(1-g/g_*)^{2\nu(r-\vartheta)}}. \quad (4.76)$$

If ϑ is a positive integer and $r = \vartheta$, the above power-law behavior is replaced by a logarithmic singularity.

Remarks.

- (i) Necessary and sufficient conditions for the moments of a random variable to fix uniquely the associated probability distribution are well known; see, e.g., [67] and references therein. A simple sufficient condition is the existence of the theory at negative Λ .
- (ii) In the case of the Liouville theory, it is straightforward to check from (4.29) that the moments $\langle A^r \rangle$ all diverge at the critical point $\Lambda = 0$, except for $r = 1$. In particular, there is no convergent expansion as in Eqs. (4.69) and (4.70), in Liouville gravity. Physically, this is related to the fact that Liouville gravity exists only for positive cosmological constant, as the exact results (4.29) or (4.34) clearly show. A similar qualitative behavior is expected for the positive curvature JT gravity theory; see Sec. VI and [40].

4. Example 1: Purely random polygons

As a first example, let us discuss the case of the RP model: general, unconstrained, random polygons, with the uniform measure. This is the flat space version of the model studied by Kitaev and Suh in negative curvature [32].

For general polygons, there is no canonical notion of area. We use Lévy's algebraic area defined in Eq. (3.10) to define the model. This notion matches the ordinary notion of area for CP, SAP, and SOP. For arbitrary polygons, the algebraic area may be positive or negative and there is a symmetry $A \mapsto -A$.

The model was solved as soon as 1950 by Lévy himself [53], and the results were rediscovered later by physicists using the link with the quantum mechanics of a particle in a uniform imaginary magnetic field [61].

On the square lattice, the number of rooted polygons is $(2n)!^2/n!^4$, and, thus, the asymptotics of the number of unrooted polygons, that is to say, polygons with no marked (privileged) point, is given by

$$w_{2n}^{\text{RP}} \underset{n \rightarrow \infty}{\sim} \frac{1}{2n} \frac{(2n)!^2}{n!^4} \underset{n \rightarrow \infty}{\sim} \frac{4^{2n}}{2\pi n^2}. \quad (4.77)$$

Comparing with (4.52) and using the well-known fact that $\nu = 1/2$ for random walks, we get $\vartheta = 1$ and $a_0 = 2/\pi$. In the continuum, with β_q given as in (4.63), the model is described by the particle action

$$S = \frac{1}{2} \int_0^{\beta_q} \dot{\mathbf{r}}^2 d\tau + \frac{\Lambda}{32\pi} \oint (x dy - y dx), \quad (4.78)$$

where $\mathbf{r} = (x, y)$ is the particle position vector. The area distribution is found to be

$$\rho_{\beta_q}^{\text{RP}}(A) = \frac{\pi}{2\beta_q} \frac{1}{\cosh^2 \frac{\pi A}{\beta_q}} \quad (4.79)$$

and the generating function at fixed β_q is

$$\begin{aligned} W^{\text{RP}}(\beta_q, \Lambda) &= W^{\text{RP}}(\beta_q, 0) \int_{-\infty}^{+\infty} e^{-\frac{\Lambda}{16\pi} A} \rho_{\beta_q}^{\text{RP}}(A) dA \\ &= \frac{2e^{-\lambda\beta_q}}{\pi\beta_q^2} \frac{\frac{\Lambda\beta_q}{32\pi}}{\sin \frac{\Lambda\beta_q}{32\pi}}. \end{aligned} \quad (4.80)$$

It is important to realize that the peculiar symmetry $A \mapsto -A$, or, equivalently, $\Lambda \mapsto -\Lambda$, of the model, implies that we are always in the “inflated crumpled phase,” described at the end of Sec. IV B 2, for the critical exponent $\nu = 1/2$. In particular, this model does not allow one to probe the $\Lambda > 0$ regime of the models for which the area is positive. However, it is plausible that it can capture, at least at the qualitative level, some physical features of flat JT gravity in the negative cosmological constant regime.

With this in mind, it is very interesting to highlight the following property, which illustrates nicely, on a toy model, the discussion we made in Sec. II E 4.

Naively, the limit of large dimensionless parameter $|\Lambda|\beta_q$ is a semiclassical limit, akin to the $\Lambda \rightarrow -\infty$ limit at fixed ℓ discussed in JT gravity in Sec. II E 4. In this limit, the area

term dominates in the action (4.78) and one may expect, in the same way as this might have been naively expected in JT, to have small fluctuations around a circle configuration, which maximizes the area.⁵⁵ But this is *not* what happens. The exact formula (4.80) for the generating function shows that, actually, there is a maximal value

$$(|\Lambda|\beta_q)_{\max} = 32\pi^2. \quad (4.81)$$

There is, thus, no large- $|\Lambda|\beta_q$ limit in this model. Instead, the loop “explodes” at the critical value (4.81), for which the average area

$$\langle A \rangle = -16\pi \frac{\partial \ln W^{\text{RP}}}{\partial \Lambda} = \frac{\beta_q}{2} \left[\frac{1}{\tan \frac{\Lambda\beta_q}{32\pi}} - \frac{32\pi}{\Lambda\beta_q} \right] \quad (4.82)$$

diverges.

The divergence of the expected area in the model at fixed quantum boundary length β_q is a nice illustration of the fact that the isoperimetric inequality gives no constraint in the continuum limit, because the closed polygon becomes a fractal in this limit, of infinite geometric length. Let us also mention that if, instead of taking the strict continuum limit, we simply consider n large but finite, the sharp bound (4.81) becomes a crossover scale between the crumpled regime for which $\langle A \rangle \propto n^{2\nu} \ell_0^2 = n \ell_0^2$, which characterizes the continuum theory, to a fully inflated “smooth” regime with $\langle A \rangle \propto n^2 \ell_0^2$. As we shall argue in Sec. VI, this crossover is similar to the transition from the full-fledged microscopic description to the effective Schwarzian description at large negative Λ in the negative curvature JT theory; see also [40].

As a last comment, let us mention that it is also possible to compute, when $\Lambda = 0$, the average area A_ν of the sectors that are enclosed by the loop and that have a winding number ν , as defined in Eq. (3.6); see Comtet *et al.* in [61,68]. The result is

$$\langle A_0 \rangle = \frac{\pi\beta_q}{30}, \quad \langle A_\nu \rangle = \frac{\beta_q}{2\pi\nu^2} \quad \text{if } \nu \neq 0. \quad (4.83)$$

In particular, the average arithmetic area defined in Eq. (3.10) is

$$\langle A_{\text{arith}} \rangle = \sum_{\nu=-\infty}^{+\infty} \langle A_\nu \rangle = \frac{\pi\beta_q}{5}, \quad (4.84)$$

whereas the average winding area

⁵⁵On the square lattice, it is the square configuration that maximizes the area. In the continuum formulation based on the action (4.78), it is the circle.

$$\langle A_{\text{wind}} \rangle = \sum_{\nu=-\infty}^{+\infty} |\nu| \langle A_\nu \rangle = +\infty \quad (4.85)$$

diverges. This is an interesting piece of information, to put into perspective with the model of self-overlapping polygons, for which the winding area coincides with the actual area of the associated distorted disk.

5. Example 2: Random self-avoiding polygons

The model of self-avoiding polygons (polygons that do not self-intersect), with the uniform measure, was proposed by Stanford and Yang in [34] as a possible microscopic definition of JT. We have explained in previous sections that such a model takes into account only a measure zero subset of the allowed configurations in JT and is, thus, expected to be a rather mediocre minisuperspace approximation to the full theory. Yet, as for the simple RP model, it is illuminating to describe its main features.

Note that relatively little is actually proven rigorously for this model, in spite of decades of intense study. In particular, the numbers $W_{2n,p}^{\text{SAP}}$ counting SAPs of length $2n$ bounding a distorted disk of area p are not known for arbitrary values of n and p . Explicit formulas that would allow one to determine these numbers recursively are not known either. But general properties can be derived, and well-motivated conjectures exist about many important quantities [37,45,46]. In particular, the critical exponents ν and θ are

$$\nu = \frac{3}{4}, \quad \theta = 1 \quad \text{for SAP.} \quad (4.86)$$

The (nonuniversal) connective constant that governs the exponential growth in (4.52) is $1/g_* \simeq 2.638$ on the square lattice.⁵⁶ The parameter a_0 that enters in the asymptotic law (4.52) for $r = 0$ is not known exactly but can be determined numerically:

$$a_0 \simeq 0.562301. \quad (4.87)$$

On the square lattice, the numbers (4.49) are, thus, given by

$$w_{2n}^{\text{SAP}} \underset{n \rightarrow +\infty}{\sim} a_0 g_*^{-2n} (2n)^{-5/2} \simeq 0.562 \times \frac{2.638^{2n}}{(2n)^{5/2}}. \quad (4.88)$$

The most impressive result available for the model is a conjecture for the exact solution in the continuum limit [69]. This conjecture is supported by analogies with the simpler, exactly solvable, model of staircase polygons and has been confirmed to a very high degree of precision by numerical analysis. To present the result, we introduce the logarithmic derivative of the Airy function

⁵⁶See [65] for an exact calculation on the hexagonal lattice.

$$\varphi(x) = -\frac{\text{Ai}'(x)}{\text{Ai}(x)}. \quad (4.89)$$

This function has an asymptotic large- x expansion of the form

$$\varphi(x) = \sum_{r \geq 0} \varphi_r x^{\frac{1-3r}{2}}. \quad (4.90)$$

Using the differential equation $\text{Ai}''(x) = x\text{Ai}(x)$, which implies that φ satisfies a Riccati equation, one finds that the coefficients φ_r can be computed recursively from

$$\begin{aligned} \varphi_0 &= 1, & \varphi_1 &= \frac{1}{4}, & \varphi_2 &= -\frac{5}{32}, \\ \varphi_{r+1} &= -\frac{3r}{4}\varphi_r - \frac{1}{2} \sum_{r'=2}^{r-1} \varphi_{r'} \varphi_{r-r'+1} \quad \text{for } r \geq 2. \end{aligned} \quad (4.91)$$

The conjecture states that the derivative of the scaling function defined in (4.74) is given by

$$\mathcal{F}'(z) = (2a_0)^{2/3} \varphi((2a_0)^{2/3} \pi z). \quad (4.92)$$

The parameter a_0 is given in Eq. (4.87). The coefficients a_r can be straightforwardly deduced from (4.92) by using (4.74):

$$a_r = 2\sqrt{\pi}a_0 \frac{(-1)^{r+1} r!}{\Gamma(\frac{3}{2}(r - \frac{1}{3}))} \frac{\varphi_r}{(2\pi^{3/2}a_0)^r}. \quad (4.93)$$

Equations (4.69) and (4.70) then yield the exact SAP generating function:

$$W^{\text{SAP}}(\beta_q, \Lambda) = -2\sqrt{\pi}a_0 \frac{e^{-\beta_q \Lambda}}{\beta_q^{5/2}} \sum_{r=0}^{\infty} \frac{\varphi_r}{\Gamma(\frac{3}{2}(r - \frac{1}{3}))} \left(\frac{\Lambda \beta_q^{3/2}}{32\pi^{5/2}a_0} \right)^r. \quad (4.94)$$

The series on the right-hand side of the above equation has an infinite radius of convergence; see Appendix B. The continuum SAP generating function is, thus, an entire function, well defined for any value, positive or negative, of the cosmological constant Λ .

It is interesting to study the limit for which the dimensionless parameter $\Lambda \beta_q^{3/2} \rightarrow -\infty$. In contrast to the case of random polygons discussed in the previous subsection, this limit is well defined, since the generating function (4.94) is an entire function. We show in Appendix B that⁵⁷

⁵⁷This result does not seem to have been derived before.

$$\begin{aligned} W^{\text{SAP}}(\beta_q, \Lambda) &= \frac{e^{-\beta_q \Lambda}}{\beta_q^{5/2}} \frac{\Lambda^2 \beta_q^3}{2^{10} \pi^5 a_0} e^{\frac{\Lambda^2 \beta_q^3}{3 \cdot 2^{12} \pi^5 a_0^2}} (1 + O(1/(\Lambda^2 \beta_q^3))). \end{aligned} \quad (4.95)$$

In particular, the average area is

$$\langle A \rangle_{\Lambda \beta_q^{3/2} \rightarrow -\infty} = \frac{|\Lambda| \beta_q^{3/2}}{3 \cdot 2^7 \pi^4 a_0^2} \beta_q^{3/2} \left[1 + \frac{3 \cdot 2^{12} \pi^5 a_0^2}{\Lambda^2 \beta_q^3} + O(1/(\Lambda^4 \beta_q^6)) \right]. \quad (4.96)$$

We see that there is a well-defined $\Lambda \rightarrow -\infty$ expansion, but it is not a standard semiclassical expansion around a dominant configuration. In particular, the expected area diverges when $\Lambda \rightarrow -\infty$. As in the case of ordinary random polygons reviewed in the previous subsection, this is possible because the microscopic length $\ell = 2n\ell_0 = \beta/\ell_0^{1/3}$ diverges in the continuum limit and, consequently, the isoperimetric inequality implies no constraint. We can also note that the term governing the exponential growth in (4.95) is proportional to Λ^2 , not to Λ , as might have been suggested by the fact that the cosmological constant term is proportional to Λ ; moreover, the corrections are governed by the expansion parameter $1/(\Lambda^2 \beta^3)$, not $1/(\Lambda \beta^{3/2})$.

6. Example 3: Random convex polygons

As a last example, let us present the model of convex polygons. Convex polygons form a subset of SAP polygons for which the intersections of the associated distorted disk with horizontal rows and vertical columns of aligned occupied lattice tiles are all connected. For instance, the polygon on the left in Fig. 1 is convex, but the polygon in the center of this figure is not. The model is, thus, an approximation to the approximation from the point of view of JT, but it has the advantage of being exactly solvable [46,70]. Moreover, it displays qualitative features reminiscent of those expected in the Schwarzian limit of negative curvature JT.

The universal critical exponents ν and ϑ are, respectively,

$$\nu = 1, \quad \vartheta = -1, \quad (4.97)$$

whereas the nonuniversal connective constant and coefficient a_0 are, on the square lattice, $1/g_* = 2$ and $a_0 = \frac{1}{256}$. The number of convex polygons, thus, grows as

$$w_{2n}^{\text{CP}} \underset{n \rightarrow +\infty}{\sim} \frac{n 2^{2n}}{128}. \quad (4.98)$$

The moments of the area at zero cosmological constant are given by

$$\langle A^r \rangle = \frac{r!^2}{2^{2r}(2r+1)!} \beta_q^{2r}. \quad (4.99)$$

Equivalently, the exact area distribution law is

$$\rho_{\beta_q}^{\text{CP}}(A) = \frac{8}{\beta_q^2} \frac{1}{\sqrt{1 - 16A/\beta_q^2}}. \quad (4.100)$$

One can check that this very simple distribution matches the distribution obtained in the continuum limit of random rectangles, which is elementary to solve.⁵⁸

The upper bound on the area, $A_{\text{max}} = (\beta_q/4)^2$, corresponds to the area of a square of perimeter β_q . The existence of the bound is a consequence of the isoperimetric inequality on the square lattice. Unlike the examples of RP and SAP reviewed in the previous subsections, this inequality applies to convex polygons in the continuum, because the value of the critical exponent $\nu = 1$ implies that the polygon length $\ell = 2n\ell_0 = \beta_q$ has a finite limit in the continuum; the “quantum length” is simply the usual geometric “classical” length in this example.

By using Eqs. (4.68)–(4.70), the fixed length continuum generating function is found to be

$$W^{\text{CP}}(\beta_q, \Lambda) = \frac{1}{256} e^{-\beta_q \lambda} \beta_q e^{-\frac{\Lambda \beta_q^2}{256\pi}} \int_0^1 e^{\frac{\Lambda \beta_q^2}{256\pi} x^2} dx. \quad (4.101)$$

It is then straightforward to study the $\Lambda \beta_q^2 \rightarrow -\infty$ limit. One finds

$$W^{\text{CP}}(\beta_q, \Lambda) \xrightarrow{\Lambda \beta_q^2 \rightarrow -\infty} e^{-\beta_q \lambda} \frac{\pi}{16|\Lambda|^{1/2}} e^{\frac{|\Lambda| \beta_q^2}{256\pi}} \left[1 + O\left(e^{-\frac{|\Lambda| \beta_q^2}{256\pi}}\right) \right]. \quad (4.102)$$

In particular, the average area is

$$\langle A \rangle \xrightarrow{\Lambda \beta_q^2 \rightarrow -\infty} \frac{\beta_q^2}{16} \left[1 - \frac{128\pi}{|\Lambda| \beta_q^2} + O\left(e^{-\frac{|\Lambda| \beta_q^2}{256\pi}}\right) \right]. \quad (4.103)$$

These results show that the $\Lambda \beta_q^2 \rightarrow -\infty$ limit in the convex polygon model is a standard “semiclassical” limit. It is dominated by the classical configuration maximising the area, which is the square of perimeter β_q and area $A_{\text{max}} = (\beta_q/4)^2$ on the square lattice. The leading exponential in (4.102) is obtained by evaluating the “classical action” $S_{\text{cl}} = \frac{\Lambda}{16\pi} A$ on this configuration. Corrections to the leading

⁵⁸Of course, the asymptotic law (4.98) is not valid for rectangles, which simply have $w_{2n} = n - 1$. But, in the continuum limit, the area moments (4.99) are the same for both rectangles and convex polygons. This kind of unexpected simplification, or universality, in the limit distributions, is at the basis of the conjecture for the solution of the SAP model, discussed in the previous subsection.

classical term are governed by $1/(|\Lambda| \beta_q^2)$, as expected for a standard semiclassical expansion. We also observe that the perturbative series stops at one loop: There is no term of the order of $1/(|\Lambda| \beta_q^2)^L$ for $L \geq 2$ in the asymptotic expansion (4.102) of the generating function.

This behavior is clearly reminiscent of the effective Schwarzian description of the negative curvature JT model, including the perturbative one-loop exactness. This may have been expected, since convex loops are similar to the smoothly wiggling boundaries described by the reparametrization *Ansatz*. Note, however, that, even for this very simple model of convex loops, the model has nonperturbative corrections of the order of $e^{-\frac{|\Lambda|}{16\pi} A_{\text{max}}}$.

7. Model comparison and summary

We have reviewed three models of closed random paths. In Fig. 17, the expected area as a function of the cosmological constant is represented for each model, to illustrate some similarities and differences between the three models.

The R model of purely random Brownian paths is likely to be the best toy model for pure JT quantum gravity when the cosmological constant Λ is negative (it does not allow one to probe the positive cosmological constant regime). Even though it is not a model of random metrics, the “boundary” (boundary in quotation marks because a purely random loop does not bound a disk, in general) has the same Hausdorff dimension as for pure JT. The negative curvature version of this model was studied by Kitaev and Suh in [32]. In zero curvature, there is a maximum allowed value for $|\Lambda|$ beyond which the loop “explodes”. The model can also be studied in positive curvature, yielding

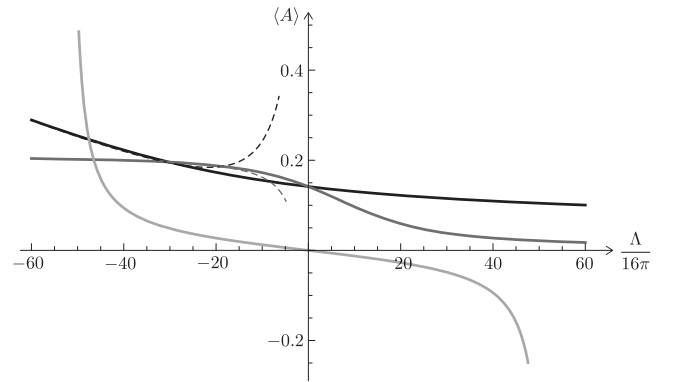


FIG. 17. Area expectation values in the SAP (black line), CP (gray line), and R (light gray line) models, as a function of the cosmological constant. Dashed lines correspond to the asymptotic forms (4.96) and (4.103). The parameter β_q for SAP is chosen to be $\beta_{q,\text{SAP}} = 1$. The parameter $\beta_q = \beta_{q,\text{CP}} = \frac{6}{\pi a_0} \simeq 1.843$ for CP is chosen in such a way that the expected area for $\Lambda = 0$ matches with SAP. The parameter $\beta_q = \beta_{q,\text{R}} = \frac{\sqrt{10-3\pi}}{2\pi^{3/2} a_0} \simeq 0.1211$ for R is chosen in such a way that the quadratic fluctuations ΔA for $\Lambda = 0$ matches with SAP.

interesting insights for positive curvature JT as well; see Sec. VI and [40].

The SAP model of self-avoiding loops has the advantage to describe random metrics on the disk. It was proposed by Stanford and Yang in [34] as a microscopic definition of JT gravity. It takes into account only a measure zero subset of the space of constant curvature metrics on the disk. It predicts that the boundary is a nontrivial fractal but with the wrong Hausdorff dimension. The model does not have a $\Lambda \rightarrow -\infty$ semiclassical expansion and the area for a given β_q is not bounded above, features shared by the SOP model describing JT gravity.

The CP model of convex loops is the simplest of all three models we have reviewed. It has the advantage of being exactly solvable. It describes random metrics on the disk but an even smaller subset than that taken into account in the SAP model. The boundaries in the continuum limit of the convex loop model are smooth. It is a rather terrible model for flat space JT but shares some basic properties with the Schwarzian description of negative curvature JT in the limit $\Lambda \rightarrow -\infty$.

C. Random self-overlapping polygons

We now focus on the model of random self-overlapping curves, which provides the correct microscopic definition of JT quantum gravity on the disk. We shall use the intuition developed in the previous subsections. Our presentation, and the conjectures we propose, should be seen as a springboard for further research in the future.

There are actually many random SOP models one may wish to consider. The model most relevant to us is the one corresponding to the definition (3.12), which counts SOP with the multiplicity factor μ_γ . This model provides the microscopic definition of pure JT gravity. When we talk about *the* SOP model, we refer to this one. One could also consider models associated with the coupling of JT gravity with a bulk QFT, which amounts to counting SOPs with a much more complicated measure. For instance, precise predictions for the case of a coupling to an arbitrary CFT were made in [16]. Another obvious model to contemplate is the model with the uniform measure, for which we count each SOP once without taking into account the multiplicity factor μ_γ . The generating function, noted with a hat, is then

$$\hat{W}(t, g) = \sum_{\gamma \in \text{SOP}} t^{F(\gamma)} g^{v(\gamma)} = \sum_{n, p \geq 1} \hat{W}_{2n, p} t^p g^{2n}. \quad (4.104)$$

We call this model SOP_u . It is not very natural from the metric point of view, since it amounts to identifying metrics that are not diffeomorphism equivalent, but, as a model of random paths, SOP_u is interesting in its own right and some properties are easier to prove for SOP_u than for SOP. We discuss SOP and SOP_u in parallel below.

1. Basic results and simple proofs

A crucial basic property we would like to prove is the exponential growth of the numbers w_{2n} and $\hat{w}_{2n}$ of self-overlapping polygons counted with and without multiplicity, respectively. This amounts to proving the existence of the connective constant $1/g_*$ for these models. A simple argument, adapting well-known ideas, can be made in the case of the SOP_u model.

Theorem 4.1. Let $\hat{w}_{2n}$ be the number of self-overlapping polygons of length $2n$ in the SOP_u model (for which polygons are counted with the uniform measure). Then, the connective constant

$$\lim_{n \rightarrow \infty} (\hat{w}_{2n})^{\frac{1}{2n}} = 1/\hat{g}_* \quad (4.105)$$

exists and is finite. Moreover, $1/\hat{g}_* \leq 3$.

Note that the connective constants of the various models we have studied automatically satisfy

$$1/g_*^R \geq 1/\hat{g}_* \geq 1/g_*^{\text{SAP}} \geq 1/g_*^{\text{CP}}, \quad (4.106)$$

simply because there are more arbitrary polygons than SOPs, more SOPs than SAPs, and more SAPs than CPs. These inequalities are true on any lattice, in spite of the fact that the precise values of the connective constants depend on the lattice. On the square lattice, $1/g_*^R = 4$, $1/g_*^{\text{SAP}} \simeq 2.638$, and $1/g_*^{\text{CP}} = 2$. Combined with $1/g_* \leq 3$, we see that the first and the third inequalities in (4.106) are strict. It is natural to conjecture that all the inequalities are strict, not only on the square lattice but on the triangular and hexagonal lattices as well.

In the case of SOPs, a similar result is true, but the proof will require more work.

Theorem 4.2. Let w_{2n} be the number of self-overlapping polygons of length $2n$ in the SOP model, i.e., counted with the multiplicity factor μ_γ . Then, the connective constant

$$\lim_{n \rightarrow \infty} (w_{2n})^{\frac{1}{2n}} = 1/g_* \quad (4.107)$$

exists and is finite. Moreover, $1/\hat{g}_* \leq 1/g_* \leq 4 = 1/g_*^R$.

Note that the inequality $1/\hat{g}_* \leq 1/g_*$ is trivially satisfied.

The proofs of Theorems 4.1 and 4.2 rely on the supermultiplicativity of the sequence $(\hat{w}_{2n})_{n \geq 2}$ and $(w_{2n})_{n \geq 2}$. More generally, one has the following.

Proposition 4.1. If $\mathcal{W}_{2n, p, \mu}$ denotes the number of SOPs of boundary length $2n$, area p , and multiplicity μ , then

$$\mathcal{W}_{2(n_1+n_2), p_1+p_2+1, \mu_1\mu_2} \geq \mathcal{W}_{2n_1, p_1, \mu_1} \mathcal{W}_{2n_2, p_2, \mu_2}. \quad (4.108)$$

The sequences $(\hat{W}_{2n, p})_{n \geq 2, p \geq 1}$ and $(\hat{w}_{2n})_{n \geq 2}$ satisfy

$$\hat{W}_{2(n_1+n_2), p_1+p_2+1} \geq \hat{W}_{2n_1, p_1} \hat{W}_{2n_2, p_2}, \quad \hat{w}_{2(n_1+n_2)} \geq \hat{w}_{2n_1} \hat{w}_{2n_2}. \quad (4.109)$$

Similarly, the sequences $(W_{2n,p})_{n \geq 2, p \geq 1}$ and $(w_{2n})_{n \geq 2}$ satisfy

$$\begin{aligned} W_{2(n_1+n_2), p_1+p_2+1} &\geq W_{2n_1, p_1} W_{2n_2, p_2}, \\ w_{2(n_1+n_2)} &\geq w_{2n_1} w_{2n_2}. \end{aligned} \quad (4.110)$$

Proof. We adapt a well-known gluing argument used to prove the supermultiplicativity property in the case of SAPs.

We introduce a Cartesian coordinate system $(x, y) \in \mathbb{Z}^2$ on the square lattice. For any self-overlapping polygon on the lattice, we call “upper boundary piece” any segment of the polygon made of horizontal edges having the maximum possible value of y , and we call similarly “lower boundary piece” any segment of the polygon made of horizontal edges having the minimum possible value of y .⁵⁹ The code for an upper or a lower boundary piece is necessarily of the form $LS \cdots SL$, with $q \geq 0$ straight vertices appearing in between the two left turns. We call a boundary edge upper (or lower) if it is a horizontal edge belonging to a upper (lower) boundary piece.⁶⁰ Note that any polygon has at least one upper and one lower edge.

We then define a notion of gluing two self-overlapping polygons γ_1 and γ_2 along an upper boundary edge e_u of γ_1 and a lower boundary edge e_l of γ_2 as depicted in Fig. 18. The new polygon obtained after the gluing is noted $\gamma = \gamma_1(e_u) \sqcup \gamma_2(e_l)$. We also use the ambiguous notation $\gamma = \gamma_1 \sqcup \gamma_2$ when the precise choice of upper and lower boundary edges is irrelevant.

Crucially, if γ_1 and γ_2 are self-overlapping, the new polygon $\gamma = \gamma_1 \sqcup \gamma_2$ is self-overlapping. This is obvious from the definition given in Sec. II: If γ_1 and γ_2 bound distorted disks, then γ also bounds a distorted disk. Moreover, if we note $2n_\gamma$, p_γ , and μ_γ the boundary length, area, and multiplicity, respectively, of a SOP γ , then clearly

$$\begin{aligned} n_{\gamma_1 \sqcup \gamma_2} &= n_{\gamma_1} + n_{\gamma_2}, & p_{\gamma_1 \sqcup \gamma_2} &= p_{\gamma_1} + p_{\gamma_2} + 1, \\ \mu_{\gamma_1 \sqcup \gamma_2} &= \mu_{\gamma_1} \mu_{\gamma_2}. \end{aligned} \quad (4.111)$$

If we denote by $\text{SOP}_{2n, p, \mu}$ the set of SOPs of length $2n$, area p , and multiplicity μ , we can use the gluing to construct a map

$$\sqcup : \text{SOP}_{2n_1, p_1, \mu_1} \times \text{SOP}_{2n_2, p_2, \mu_2} \rightarrow \text{SOP}_{2(n_1+n_2), p_1+p_2+1, \mu_1 \mu_2}. \quad (4.112)$$

There are, of course, many possible maps, depending on the detailed choice of upper and lower boundary edges used in

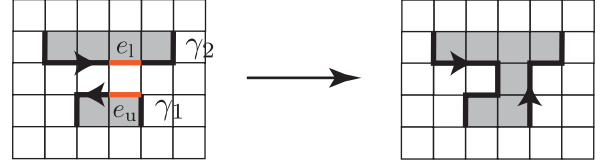


FIG. 18. Gluing of two SOPs γ_1 and γ_2 along an upper boundary edge e_u of γ_1 and a lower boundary edge e_l of γ_2 (outlined in orange) to form a new SOP $\gamma_1 \sqcup \gamma_2$. Some faces of the distorted disks bounded by γ_1 , γ_2 , and $\gamma_1 \sqcup \gamma_2$ are in gray. This construction is used in the proof of the superadditivity inequalities (4.109).

the gluing of each pair of elements of $\text{SOP}_{2n_1, p_1, \mu_1} \times \text{SOP}_{2n_2, p_2, \mu_2}$. We work with a given choice. This map is injective. This is checked by noting that, in $\gamma_1 \sqcup \gamma_2$, the distorted disk associated with γ_1 is entirely below the distorted disk associated with γ_2 . The equality $\gamma_1 \sqcup \gamma_2 = \gamma'_1 \sqcup \gamma'_2$, together with the fact that the lengths of γ_i and γ'_i are given by $2n_i$ and are, thus, the same, then implies that $\gamma_i = \gamma'_i$. The existence of the injective map $\sqcup$ implies the inequality (4.108).

Moreover, denoting by $\text{SOP}_{2n, p}$ and SOP_{2n} the set of SOPs of length $2n$ and area p , or the set of SOPs of length $2n$, respectively, the gluing restricts to injective maps

$$\sqcup : \text{SOP}_{2n_1, p_1} \times \text{SOP}_{2n_2, p_2} \rightarrow \text{SOP}_{2(n_1+n_2), p_1+p_2+1}, \quad (4.113)$$

$$\sqcup : \text{SOP}_{2n_1} \times \text{SOP}_{2n_2} \rightarrow \text{SOP}_{2(n_1+n_2)}. \quad (4.114)$$

The existence of these maps imply the inequalities (4.109).

Finally, let us now denote by $\text{Disk}_{2n, p}$ and Disk_{2n} the set of discretized flat disks of boundary length $2n$ and area p , or the set of distorted disks of boundary length $2n$, respectively. The same gluing operation as above yields injective maps acting on the distorted disks themselves:

$$\sqcup : \text{Disk}_{2n_1, p_1} \times \text{Disk}_{2n_2, p_2} \rightarrow \text{Disk}_{2(n_1+n_2), p_1+p_2+1}, \quad (4.115)$$

$$\sqcup : \text{Disk}_{2n_1} \times \text{Disk}_{2n_2} \rightarrow \text{Disk}_{2(n_1+n_2)}, \quad (4.116)$$

from which the inequalities (4.110) follow. $\blacksquare$

By Fekete's lemma, the supermultiplicativity property of the sequences $(\hat{w}_{2n})$ and (w_{2n}) implies that the limits in (4.105) and (4.107) exist. However, *a priori*, this limit may be infinite.

In the case of the sequence $(\hat{w}_{2n})$, this loophole can be easily overcome, because we have the trivial bound $\hat{w}_{2n} \leq 4 \times 3^{2n}$ by the total number of possible boundary codes times the factor of 4 due to lattice rotations. Theorem 4.1, including the bound $1/\hat{g}_* \leq 3$, is thus proven.

The case of SOP is more subtle. Because the polygons are now counted with a multiplicity μ_γ for which, to the best of our knowledge, no bound is known as a function of the polygon length $2n$, we cannot prove a bound on w_{2n} in a

⁵⁹Note that distinct lower or upper boundary pieces may be located on top of each other, as in the example in Fig. 14.

⁶⁰Note that we could similarly introduce the notions of rightmost and leftmost boundary pieces and rightmost and leftmost edges.

completely elementary way. To fully appreciate the nature of the problem, let us make a few comments.

It is not difficult to construct SOPs for which the multiplicity grows exponentially with the boundary length.⁶¹ For example, by gluing r times the smallest Milnor SOP depicted in Fig. 14, we get a SOP of length $2n = 48r$ that has multiplicity $2^r = 2^{n/24}$. If we could prove that the multiplicity were always exponentially bounded above as a function of the SOP length, i.e., $\mu_\gamma \leq AB^{2n}$ for some constants A and B , then we would deduce that $w_{2n} \leq AB^{2n} \hat{w}_{2n}$. The existence of g_* would follow, with $1/g_* \leq B/\hat{g}_*$. However, finding a direct proof of the bound $\mu_\gamma \leq AB^{2n}$ is nontrivial.

Furthermore, while a bound of the form $\ln w_{2n} \leq O(n)$ is natural for a random path model, it is rather nontrivial for a random metric model. From the metric point of view, we can make the following direct reasoning. The total number of quadrangulations of the disk, with boundary length $2n$ and p faces, is $W_{2n,p}$, given in Eq. (4.15). It is easy to check that $W_{2n,p+1} > W_{2n,p}$. Moreover, quadrangulations corresponding to flat disks necessarily have $n-1 \leq p \leq \lfloor n^2/4 \rfloor$; see, e.g., the discussion after Hypothesis 4.1 in Section IV B 1. Thus,

$$w_{2n} \leq \sum_{p=n-1}^{\lfloor n^2/4 \rfloor} W_{2n,p} \leq (\lfloor n^2/4 \rfloor - n + 2) W_{2n, \lfloor n^2/4 \rfloor}. \quad (4.117)$$

The large- n asymptotics of $W_{2n, \lfloor n^2/4 \rfloor}$ then yields a bound of the form $\ln w_{2n} \leq O(n^2)$. This is much weaker than the bound $\ln w_{2n} \leq O(n)$ that we seek.

In the next subsection, by using a more elaborate approach, using explicitly the flatness condition, we shall prove the following theorem.

Theorem 4.3. The number of distorted disks having boundary length $2n$ is bounded above as

$$w_{2n} \leq \frac{4(2n+1)}{n(n+1)^2(n+2)} \frac{(2n)!^2}{n!^4} \underset{n \rightarrow \infty}{\sim} \frac{8}{\pi} \frac{4^{2n}}{n^4}. \quad (4.118)$$

The right-hand side of the inequality is simply $2/n$ multiplied by the number of excursions in the positive quadrant; see the proof of Theorem 4.4 below. Combining (4.118) with the second inequality in (4.110) proves Theorem 4.2.

Corollary 4.1. The multiplicity of a SOP of length $2n$ is bounded above by

$$\mu_\gamma \leq \frac{4(2n+1)}{n(n+1)^2(n+2)} \frac{(2n)!^2}{n!^4} \underset{n \rightarrow \infty}{\sim} \frac{8}{\pi} \frac{4^{2n}}{n^4}. \quad (4.119)$$

Proof. This is a direct consequence of the trivial inequality

$$w_{2n} = \sum_{\gamma \in \text{SOP}_{2n}} \mu_\gamma \geq \max\{\mu_\gamma | \gamma \in \text{SOP}_{2n}\}. \quad (4.120)$$

More generally, we conjecture that the SOP model fits in the general framework discussed in Sec. IV B. In particular,

Conjecture 4.1. The critical exponents ν and ϑ , which govern the way the continuum limit is defined, Eq. (4.63), and the partition function at zero cosmological constant, Eq. (4.67), are given by

$$\nu = \frac{1}{2}, \quad \vartheta = 2. \quad (4.121)$$

This conjecture is motivated by the continuum approach to JT gravity introduced in [16]. Note that logarithmic corrections to the asymptotics given in Eq. (4.52) cannot be excluded and are suggested by the fact that pure JT gravity seems to share similarities with Liouville theory coupled to $c = 1$ matter [16]. For instance,

Conjecture 4.2. The large- n asymptotic behavior of the number w_{2n} of SOP, counted with multiplicity, is of the form

$$w_{2n} \underset{n \rightarrow \infty}{\sim} a_0 g_*^{-2n} (2n)^{-3} (\ln n)^\varphi \quad (4.122)$$

for a connective constant $1/g_* > 1/g_*^{\text{SAP}}$ and a new critical exponent φ that may be nonzero.

2. The bound on w_{2n}

Let us present the proof of the inequality (4.118).⁶² We start by introducing a few elementary notions that we need to present the argument in a transparent way.

Strips (columns and rows), rails, cuts, chords, and rays. A *row* (or *column*) is a maximum collection of squares aligned horizontally (or vertically) on the discretized flat disk. A *strip* is either a row or a column. A row (or column) starts and ends on vertical (or horizontal) boundary edges, called end edges of the row (or column). Conversely, any vertical (horizontal) boundary edge belongs to a unique row (or column) and is, thus, paired, in a unique way, with another vertical (horizontal) boundary edge, corresponding to the other end of the strip. Boundary edges paired in this way by a row (or a column) are said to face each other horizontally (vertically). The paths corresponding to the two sides of a strip lengthwise are called *rails*. A column has a right rail and a left rail; a row has a top rail and a bottom rail. Note that any square in the disk is at the intersection of a unique row and a unique column.

⁶¹I thank Peter Shor for pointing out this possibility.

⁶²The core of this proof was offered to us by the referee, who should take all the credit for it.

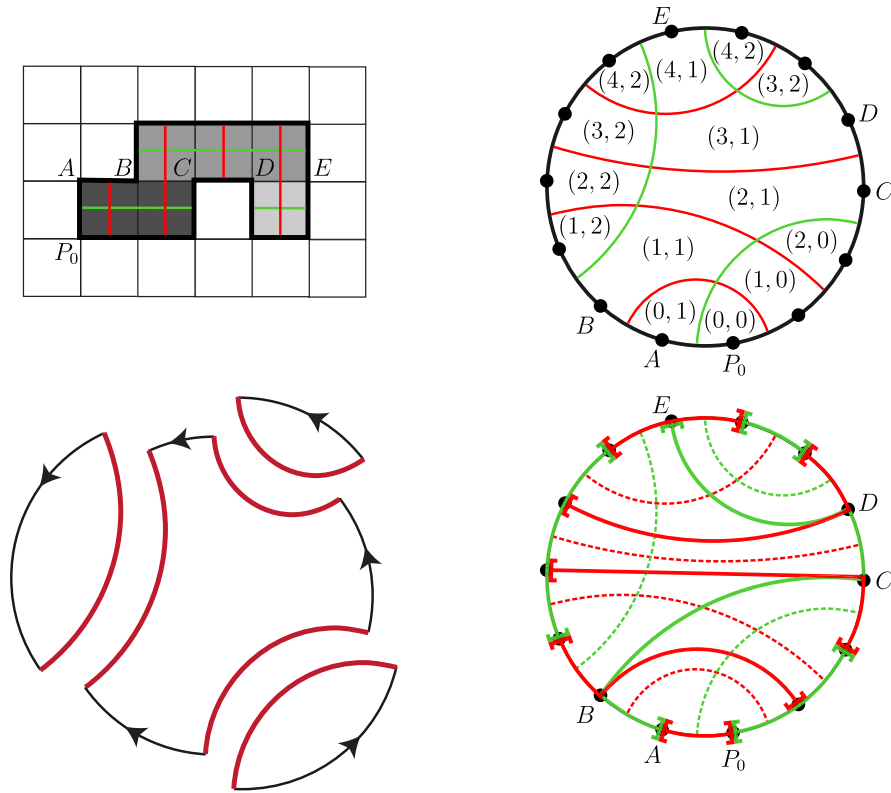


FIG. 19. Upper-left inset: illustration of the notions of strips, rails, strip cuts, chords, and rays for a simple discretized disk. The three rows are colored in three different shades of gray. Row cuts and column cuts, chosen to be geodesics of the flat disk, are green and red lines, respectively. (A, B, C) and (D, E) are top rails; (B, C, D, E) is a bottom rail. (A, B, C, D, E) is a chord. (B, C) and (D, E) are rays. Upper-right inset: convenient representation of the strip structure of the same disk. The boundary corresponds to the unit circle, with boundary vertices depicted by dots. For greater clarity, the strip cuts are now represented by geodesics of the hyperbolic disk. The intersections between the row cuts and the column cuts correspond to the squares that make up the disk. Each region delimited by the cuts is associated with a vertex, for which we have indicated the horizontal and vertical coordinates, as defined after Eqs. (4.124) and (4.125), choosing the root vertex to be $P = P_0$. Lower-left inset: Blank cut decomposition of the disk associated with the rows. We get four horizontal Blank domains. Compare, for example, with the right inset in Fig. 10. Lower-right inset: chord structure. The chords end points are depicted as brackets. The strip structure is also shown in dotted lines. Note that there is exactly one horizontal (vertical) chord in each horizontal (vertical) Blank domain; this is a general property, and, thus, Blank domains can be identified with chords.

A cut that separates a strip into two pieces lengthwise is called a *strip cut*. It is a Blank cut, as defined in Sec. III C 3. Cutting in this way all the rows (or all the columns) yields a Blank decomposition of the disk. The Blank domains associated with the row cuts (column cuts) are called horizontal (vertical) Blank domains. The strip cuts may be chosen to be geodesics of the flat disk, but any set of smooth curves in the disk joining the open interiors of the edges that face each other encode the same combinatorial data.

These notions, together with the notions of chords and rays introduced below, are illustrated in Fig. 19.

Definition 4.1. The strip structure of a discretized flat disk is given by the ordered set of edges $(e_0, e_1, \dots, e_{2n-1})$ forming its boundary and by the two lists of pairs of edges that are facing each other, horizontally and vertically. By convention, we choose the first boundary edge $e_0 = (P_0, P_1)$ to be a lower boundary edge starting at a boundary vertex P_0 of type L.

Lemma 4.1. A discretized flat disk is entirely determined by its strip structure.

Proof. The strip structure of a discretized disk can be conveniently pictured by depicting the ordered set of boundary edges $(e_0, e_1, \dots, e_{2n-1})$ by a sequence of arcs of equal lengths on the unit circle and by representing the pairings by the associated horizontal or vertical strip cuts. By convention, the first edge is chosen to be horizontal, oriented to the right. This fixes the orientation of the disk. This construction is illustrated in Fig. 19 (upper-right inset) and also in Fig. 21 below.

Let us first show that the strip structure determines uniquely the boundary SOP. Recall that the boundary SOP is given by an ordered sequence of lattice points P_k , $0 \leq k \leq 2n-1$, with $e_k = (P_k, P_{k+1})$ for $0 \leq k \leq 2n-2$ and $e_{2n-1} = (P_{2n-1}, P_0)$. The SOP is determined by its boundary code, i.e., the types of the boundary vertices P_k , which can be left, straight, or right. We chose P_0 to be L.

The edge e_0 is lower, thus necessarily horizontal. More generally, column cuts end on horizontal edges and row cuts end on vertical edges; we thus know the nature, horizontal or vertical, of each boundary edge. The type of any vertex P_k is then determined in the following way. If the cuts surrounding P_k , which are the cuts ending on the edges e_{k-1} and e_k , are both row cuts or both column cuts, then the boundary does not change direction at P_k . This means that P_k must be of type S. If one cut surrounding P_k is a column cut and the other a row cut, the boundary changes direction. The vertex P_k is then of type L if the cuts intersect and of type R if the cuts do not intersect.

The full discretized disk is then reconstructed by gluing together the set of rows, or, equivalently, the set of columns, according to the boundary edge pairing given by the strip structure. Note that the length of a given strip (i.e., the number of squares in it) is equal to the number of intersection points that the associated strip cut has with other “perpendicular” strip cuts. The total number of intersections found in the bulk of the strip structure is, thus, equal to the total number of squares making up the disk, i.e., its area. ■

Dual to the notion of strip is the notion of *chord*. Chords are maximal purely horizontal or vertical paths on the discretized disk. Chords may have both bulk pieces (called rays; see below) and boundary pieces. Their end points are on the boundary, with two perpendicular chords ending on L vertices, one ending on S vertices, and none ending on R vertices. Any vertex, bulk or boundary, is at the intersection of two perpendicular chords. Bulk and R vertices are at intersections that are not chord end points; S vertices are at intersections that are the end point of one chord but not of the other; L vertices are at the end points of the two chords that intersect.

Definition 4.2. The chord structure of a discretized flat disk is given by the ordered set of edges $(e_0, e_1, \dots, e_{2n-1})$ forming its boundary and by the two lists of horizontal and vertical chords, a chord being encoded by an ordered set of boundary points through which it passes. By convention, we choose the first boundary edge $e_0 = (P_0, P_1)$ to be a lower boundary edge starting at a boundary vertex P_0 of type L.

Lemma 4.2. A discretized flat disk is entirely determined by its chord structure.

Proof. The proof is very similar to that of Lemma 4.1. One notices that the number of chords ending on a boundary vertex determines its type; the chord structure thus determines the boundary SOP. Furthermore, since each bulk vertex is at the intersection of two chords, and each edge corresponds to a chord section joining two vertices, the full graph is encoded in the chord structure. ■

Finally, one may also introduce the notion of *ray*, following the recent work [71]. Rays are maximal purely horizontal or vertical paths on the discretized disk that visit only bulk vertices except for their end points. To a

boundary vertex of type R is always attached two rays, one vertical and one horizontal; to a boundary vertex of type S is attached a single ray, perpendicular to the surrounding boundary edges; no ray is attached to a boundary vertex of type L. Rays are geodesics, ending on the disk boundaries, that intersect at the bulk vertices. They are Blank cuts. One can define the *ray structure* of a discretized disk to be the ordered set of boundary vertices $(P_0, \dots, P_{2n-1})$ together with the two lists of pairs of boundary vertices that are the end points of the same horizontal rays and of the same vertical rays. It is straightforward to prove that the ray structure determines uniquely the discretized disk, exactly as in the cases of the strip and chord structures. However, for our purposes, using strips and/or chords is more convenient than using rays.

Horizontal and vertical distances. Let P and Q be two vertices (bulk or boundary) of the discretized disk. Let $\Gamma(P, Q)$ be the set of oriented paths from P to Q . If $\gamma \in \Gamma(P, Q)$, we note h_γ and v_γ its horizontal and vertical lengths, respectively, which are equal to the total number of horizontal and vertical edges in γ . The total length is $d_\gamma = h_\gamma + v_\gamma$. The usual distance, the horizontal distance, and the vertical distance between P and Q are then defined by

$$d(P, Q) = \min\{d_\gamma, \gamma \in \Gamma(P, Q)\}, \quad (4.123)$$

$$h(P, Q) = \min\{h_\gamma, \gamma \in \Gamma(P, Q)\}, \quad (4.124)$$

$$v(P, Q) = \min\{v_\gamma, \gamma \in \Gamma(P, Q)\}. \quad (4.125)$$

The set $\Gamma(P, Q)$ being finite, there is always at least one path γ such that $h(P, Q) = h_\gamma$, etc. Such paths are said to realize the distance.

Note that, by definition, vertices belonging to a path never repeat themselves along the path. When no such constraint is imposed, we refer to it as a walk instead of a path. Using walks in the above definitions does not change the associated notions of distances.

From now on, we fix a root vertex P (which may be a bulk or a boundary vertex), and we note $h_P = h(P, \cdot)$ and $v_P = v(P, \cdot)$ the horizontal and vertical distance functions, respectively, from P .

Lemma 4.3. Let $(x, y) \in \mathbb{Z}^2$ be a Cartesian coordinate system on the square lattice, centered on P . We note (x_Q, y_Q) the coordinates of a vertex Q on a discretized disk. Then,

$$h_P(Q) \geq |x_Q|, \quad v_P(Q) \geq |y_Q| \quad (4.126)$$

and

$$x_Q \equiv h_P(Q) \pmod{2}, \quad y_Q \equiv v_P(Q) \pmod{2}. \quad (4.127)$$

Proof. This is the direct consequence of an elementary counting. Denote by right_γ , left_γ , up_γ , and down_γ the number of right, left, upward, and downward moves along a path $\gamma \in \Gamma(P, Q)$. On the one hand,

$$x_Q = \text{right}_\gamma - \text{left}_\gamma, \quad y_Q = \text{up}_\gamma - \text{down}_\gamma, \quad (4.128)$$

and, on the other hand,

$$h_\gamma = \text{right}_\gamma + \text{left}_\gamma, \quad v_\gamma = \text{up}_\gamma + \text{down}_\gamma. \quad (4.129)$$

Combining the above equations, we get

$$h_\gamma \geq |x_Q|, \quad v_\gamma \geq |y_Q| \quad (4.130)$$

and

$$\begin{aligned} x_Q &= h_\gamma - 2\text{left}_\gamma \equiv h_\gamma \pmod{2}, \\ y_Q &= v_\gamma - 2\text{down}_\gamma \equiv v_\gamma \pmod{2}. \end{aligned} \quad (4.131)$$

Equations (4.126) and (4.127) follow by applying (4.130) and (4.131) to paths that realize the distances $h_P(Q)$ and $v_P(Q)$. ■

Lemma 4.4. Assume that Q and Q' are two vertices of a discretized disk. If Q and Q' are joined by a horizontal edge, then $v_P(Q) = v_P(Q')$ and $h_P(Q') = h_P(Q) \pm 1$. If Q and Q' are joined by a vertical edge, then $h_P(Q) = h_P(Q')$ and $v_P(Q') = v_P(Q) \pm 1$.

Proof. Assume that Q and Q' are joined by a horizontal edge (Q, Q') . By considering a path that realizes $v_P(Q)$, followed by (Q, Q') , we deduce that $v_P(Q') \leq v_P(Q)$. By considering a path that realizes $v_P(Q')$, followed by (Q', Q) , we deduce that $v_P(Q) \leq v_P(Q')$. Thus, $v_P(Q) = v_P(Q')$. By considering a path that realizes $h_P(Q)$, followed by (Q, Q') , we deduce that $h_P(Q') \leq h_P(Q) + 1$. Similarly, by symmetry, $h_P(Q) \leq h_P(Q') + 1$. Thus, $h_P(Q') = h_P(Q)$ or $h_P(Q') = h_P(Q) \pm 1$. The first possibility is ruled out by Eq. (4.127) and the fact that $x_{Q'} = x_Q \pm 1$. The proof of $h_P(Q) = h_P(Q')$ and $v_P(Q') = v_P(Q) \pm 1$ if Q and Q' are joined by a vertical edge is strictly similar. ■

Corollary 4.2. The connected components of the level sets of h_P and v_P are the set of vertices on vertical and horizontal chords (or, by abuse of language, the vertical and horizontal chords themselves), respectively.

Proof. The set of edges on vertical (horizontal) chords are sets of constant h_P (v_P) by Lemma 4.4. Two vertical (horizontal) chords having the same value of h_P (v_P) are disconnected. Indeed, if not, they would be joined by a horizontal (vertical) edge, which is impossible by Lemma 4.4. ■

Strip structure and level sets. Consider the Blank cut decomposition associated with column cuts (the case of row cuts is strictly similar). Assume that the disk has c columns, and thus c column cuts. These cuts delimit $c + 1$

Blank domains, as, for instance, in the lower-left inset in Fig. 19. These domains are level sets of h_P . More precisely, we have the following.

Lemma 4.5. Blank domains and chords are in one-to-one correspondence, each Blank domain associated with the row cut (column cut) Blank decomposition being made of a unique horizontal (vertical) chord. In particular, horizontal (vertical) Blank domains are level sets for v_P (h_P), and the number of horizontal (vertical) chords is equal to the number of rows (columns) plus one.

Proof. Consider one horizontal Blank domain associated with the row cuts (the case of vertical Blank domains is strictly similar). By definition, this domain is bounded by at least one row cut. Either the top or the bottom rail of the associated row is in the Blank domain. Extending this rail horizontally in a maximal way to the left and to the right, we get a chord. This chord is entirely included in the Blank domain. Indeed, when extending the rail, we cannot leave the Blank domain before we reach the disk boundary, since the row cuts are horizontal [see, e.g., the extension of the top rail (A, B, C) to the chord (A, B, C, D, E) in Fig. 19, upper-left inset]. Assume now that a given horizontal Blank domain contains two distinct horizontal chords. Since a Blank domain is connected, this would imply that the horizontal Blank domain contains a vertical edge. But this is impossible, since a vertical edge always crosses a row cut and, thus, joins two distinct Blank domains. The counting of the chords then follows from the counting of the Blank domains. ■

Corollary 4.3. The vertices of the discretized disk are in one-to-one correspondence with the intersections of the horizontal and the vertical Blank domains, which correspond to the intersection of the associated chords.

Consider now a strip cut, that we denote by $\mathbf{C}$, cutting a strip $\mathbf{S}(\mathbf{C})$. The cut divides the disk in two complementary regions $R_P(\mathbf{C})$ and $\bar{R}_P(\mathbf{C})$, the region $R_P(\mathbf{C})$ containing the root vertex P . One rail of $\mathbf{S}(\mathbf{C})$ is in $R_P(\mathbf{C})$; it is a piece of a chord that defines a Blank domain that we denote $\mathbf{C}_P$. The other rail of $\mathbf{S}(\mathbf{C})$ is in $\bar{R}_P(\mathbf{C})$; it is a piece of a chord that defines a Blank domain $\bar{\mathbf{C}}_P$. If $\mathbf{C}_P$ and $\bar{\mathbf{C}}_P$ are horizontal, we denote by $v_P(\mathbf{C})$ and $\bar{v}_P(\mathbf{C})$ their associated constant values of v_P ; if they are vertical, we denote by $h_P(\mathbf{C})$ and $\bar{h}_P(\mathbf{C})$ their associated constant values of h_P .

Lemma 4.6. If $\mathbf{C}$ is a column cut (a row cut), then $\bar{h}_P(\mathbf{C}) = h_P(\mathbf{C}) + 1$ [$\bar{v}_P(\mathbf{C}) = v_P(\mathbf{C}) + 1$]. Moreover, all the vertices in $\bar{R}_P(\mathbf{C})$ have $h_P > h_P(\mathbf{C})$ [$v_P > v_P(\mathbf{C})$].

Proof. Let us make the reasoning in the case of a column cut $\mathbf{C}$. The situation is depicted in Fig. 20. Let Q' be a vertex in $\bar{R}_P(\mathbf{C})$. Any path γ going from P to Q' must cross $\mathbf{C}$. Just before crossing, the path goes through a vertex in $\mathbf{C}_P$; taking into account the horizontal edge used for crossing, it thus contains at least $h_P(\mathbf{C}) + 1$ horizontal edges. This shows that $h_P(Q') > h_P(\mathbf{C})$. Let us now pick a vertex Q in $\mathbf{C}_P$ and a vertex Q' in $\bar{\mathbf{C}}_P$ such that (Q, Q') is a horizontal edge; such vertices always exist. From what we

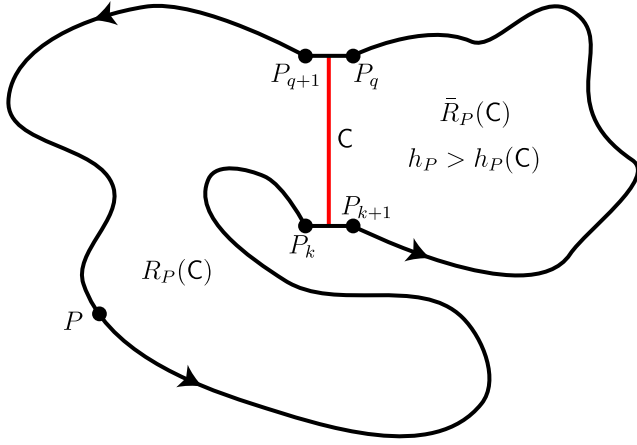


FIG. 20. Schematic representation of a column cut C (in red) dividing the disk in two regions R_P and $\bar{R}_P$. The end edges of the column are noted (P_k, P_{k+1}) and (P_q, P_{q+1}) , as in the proof of Theorem 4.4. We have $h_P(P_k) = h_P(P_{q+1}) = h_P(C)$ and $h_P(P_{k+1}) = h_P(P_q) = \bar{h}_P(C) = h_P(C) + 1$. All vertices in $\bar{R}_P$ have $h_P > h_P(C)$.

have just shown, $h_P(Q') = \bar{h}_P(C) \geq h_P(C) + 1$; Lemma 4.4 implies that this inequality must be an equality. ■

Lemma 4.6 makes it easy to compute the horizontal and vertical “coordinates” (h, v) of any vertex of the discretized disk or, equivalently, of the intersections of vertical and

horizontal Blank domains. These coordinates are indicated in Fig. 19, upper-right inset, and in Fig. 21, in the interesting example of the two distinct disks that are bounded by the same minimal Milnor SOP depicted in Fig. 14. Note that the lemma also shows that rows and column have a natural transverse orientation pointing away from the root vertex, in the direction of increasing v_P or h_P ; this is an important property used in the recent work [71].

Discretized disks and discrete excursions in the quadrant. We endow the square lattice with Cartesian coordinates $(x, y) \in \mathbb{Z}^2$. Let Exc_{2n} be the set of excursions of length $2n$ in the positive quadrant, that is to say, the set of walks on the square lattice, starting and ending at $(0,0)$, of length $2n$ and confined in the quadrant $x \geq 0, y \geq 0$. Let Disk_{2n}^* be the set of discretized flat disks of boundary length $2n$ with a marked vertex on the boundary. We are going to build an injection

$$\Psi: \text{Disk}_{2n}^* \rightarrow \text{Exc}_{2n}. \quad (4.132)$$

The map Ψ associates an excursion $\Psi(D) = (\tilde{P}_0, \dots, \tilde{P}_{2n-1}, \tilde{P}_0)$ to a disk $D \in \text{Disk}_{2n}^*$ in the following way. We note $(P_0, P_1, \dots, P_{2n-1})$ the rooted boundary SOP of D , P_0 being the root. To the boundary vertex P_k we associate the vertex $\tilde{P}_k$ on the square lattice having Cartesian coordinates $\tilde{x}_k = h_{P_0}(P_k)$ and $\tilde{y}_k = v_{P_0}(P_k)$. Lemma 4.4 ensures that $(\tilde{P}_0, \tilde{P}_1, \dots, \tilde{P}_{2n-1}, \tilde{P}_0)$ is a closed walk, and it is trivially confined to the positive quadrant, $\tilde{x}_k \geq 0, \tilde{y}_k \geq 0$.

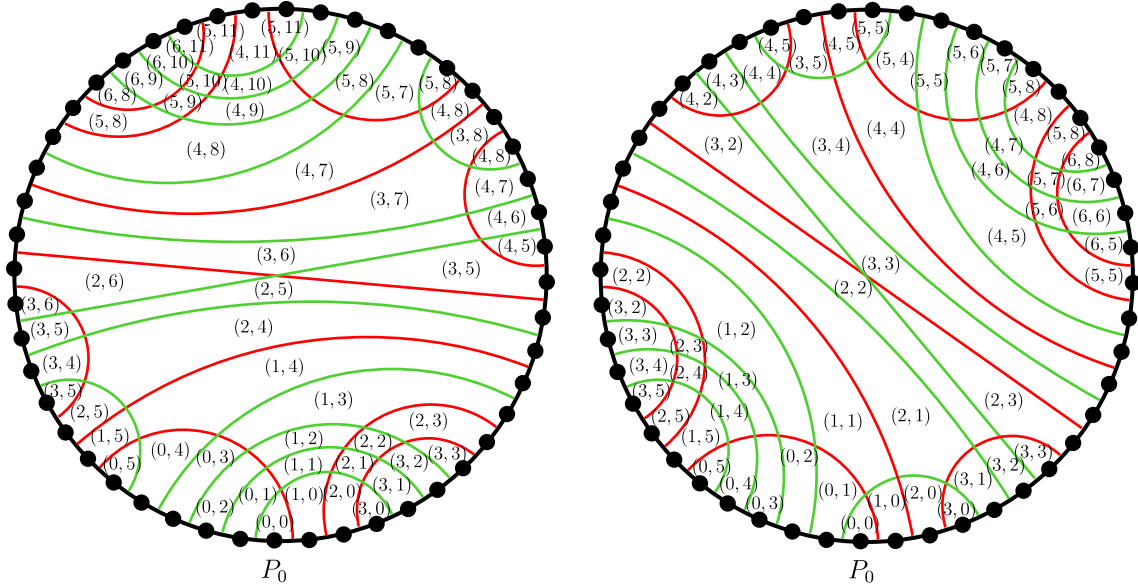


FIG. 21. Strip structures of the two distinct disks bounded by the same Milnor SOP depicted in Fig. 14. The upper-right inset in this figure corresponds to the disk on the left-hand side. The vertex P_0 is as in Fig. 14. Row cuts and column cuts are drawn in green and red, respectively. For greater clarity, we have chosen to represent these cuts by geodesics of the hyperbolic disk. As explained in the main text, the strip structure determines uniquely the discretized disk. For instance, the area of the disk is equal to the number of intersection points between the row cuts and the column cuts. We let the reader work out the dual chord structure of these examples; there is a single vertical (horizontal) chord in each vertical (horizontal) Blank domain. Choosing the root vertex to be $P = P_0$, we have indicated the horizontal and vertical coordinates at the intersections of Blank domains, which are in one-to-one correspondence with the vertices of the underlying graph.

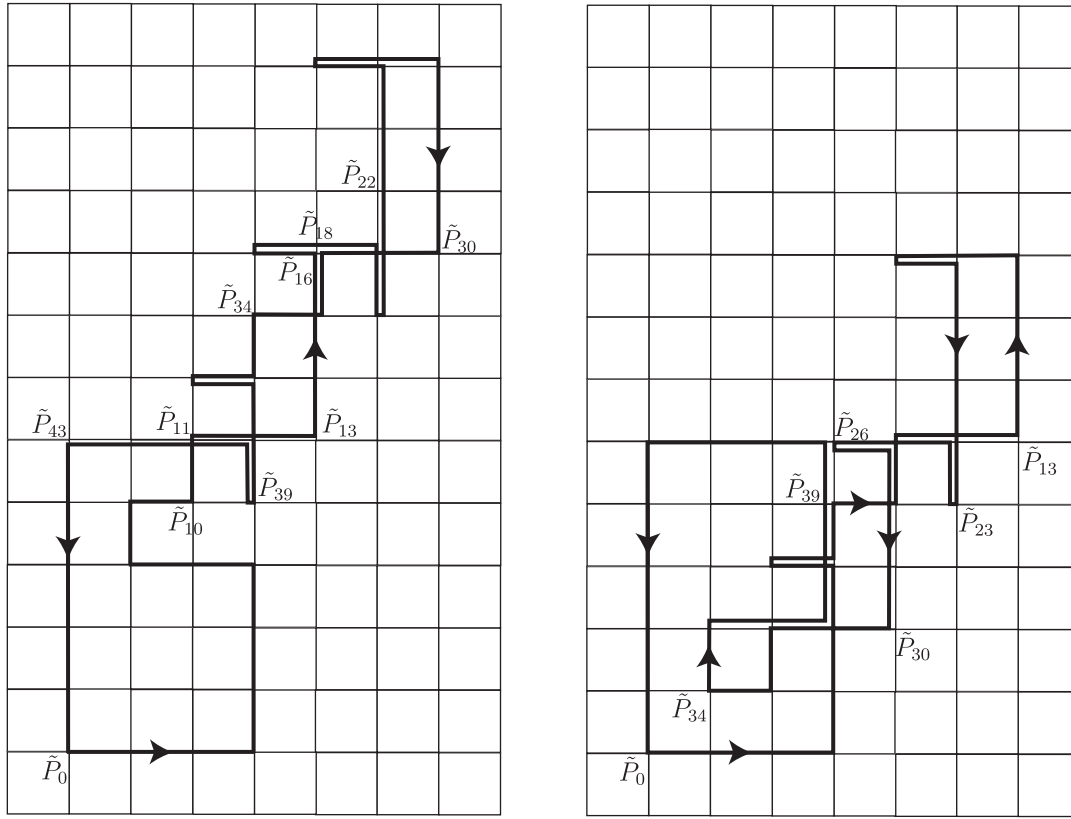


FIG. 22. The two distinct excursions in the positive quadrant, associated with the two disks bounded by the Milnor SOP in Fig. 14. The excursions are straightforwardly built from the coordinates indicated in Fig. 21. Theorem 4.4 ensures that the two distinct discretized disks can be reconstructed from the excursions alone.

The map Ψ is, thus, well defined. The excursions associated with the two disks bounded by the Milnor SOP depicted in Fig. 14 are represented in Fig. 22, using the data given in Fig. 21.

Theorem 4.4. The map Ψ is an injection.

Proof. Let us show that we can reconstruct the strip structure of the disk from the excursion. We can then conclude by using Lemma 4.1.

Let us first note that Lemma 4.4 implies that if an edge $e_k = (P_k, P_{k+1})$ of the boundary SOP is horizontal (vertical), then the associated edge $\tilde{e}_k = (\tilde{P}_k, \tilde{P}_{k+1})$ in the excursion is also horizontal (vertical).

Assume that the edge e_k , and, thus, also $\tilde{e}_k$, is horizontal. The edge e_k is then the end edge of a column $\mathbf{S}(\mathbf{C})$, with associated cut $\mathbf{C}$. To reconstruct the strip structure of the disk, all we need is to find the edge $e_q = (P_q, P_{q+1})$ that faces e_k vertically, that is to say, the other end edge of $\mathbf{S}(\mathbf{C})$.

In the pair of edges (e_k, e_q) that face each other vertically, one edge will be oriented from $\mathbf{C}_{P_0}$ to $\tilde{\mathbf{C}}_{P_0}$ and the other from $\tilde{\mathbf{C}}_{P_0}$ to $\mathbf{C}_{P_0}$, since the disk must be on the left of the boundary edges. Without loss of generality, we can thus assume that e_k is oriented from $\mathbf{C}_{P_0}$ to $\tilde{\mathbf{C}}_{P_0}$. Equivalently, by Lemma 4.6, $\tilde{x}_{k+1} = \tilde{x}_k + 1$. We are then in the situation depicted in Fig. 20, with $P = P_0$. The vertex

P_k is in $\mathbf{C}_{P_0}$, with $h_P(P_k) = \tilde{x}_k = h_P(\mathbf{C})$, whereas the vertex P_{k+1} is in $\tilde{\mathbf{C}}_{P_0}$, with $\tilde{x}_{k+1} = h_P(P_{k+1}) = \tilde{h}_P(\mathbf{C}) = h_P(\mathbf{C}) + 1 = \tilde{x}_k + 1$. By Lemma 4.6, all the boundary vertices that follow P_{k+1} must be in $\tilde{R}_{P_0}(\mathbf{C})$, and, thus, have $h_P > h_P(\mathbf{C})$, as long as the boundary does not cross $\mathbf{C}$ again. The crucial remark is then that the boundary crosses $\mathbf{C}$ precisely at the edges e_k and e_q . The vertex P_{q+1} is thus characterized by being the first boundary vertex after P_k that has $\tilde{x}_{q+1} = \tilde{x}_k$.

A strictly similar reasoning can be made in the case of vertical edges, using the vertical distance instead of the horizontal distance. ■

The existence of the injection Ψ proves that the cardinal of Disk_{2n}^* , which is, according to the discussion at the end of Sec. III B 2, $\frac{n}{2}w_n$, is bounded above by the number of excursions in the quadrant, which is well known:

$$|\text{Exc}_{2n}| = \frac{2(2n+1)}{(n+2)(n+1)^2} \frac{(2n)!^2}{n!^4}. \quad (4.133)$$

This yields the inequality (4.118).

3. Observables

Beyond the standard boundary length, area,⁶³ and radius of gyration discussed before, there are several interesting observables to study. A nonexhaustive list is as follows (many variants are possible).

- (i) The arithmetic area A_{arith} , defined in (3.10), which provides an intuitive measure of the size of the SOP drawn on the Euclidean plane. This is expected to yield a notion of size which is equivalent to the radius of gyration, with a scaling exponent ν' ; see Eq. (4.55). It is interesting to recall that, in the case of the Brownian path model, the arithmetic area and the winding area behave very differently; see Eqs. (4.84) and (4.85). In the case of the SOP model, it is natural to conjecture that $\nu' = \nu$.
- (ii) The number ι_γ of bounded simply connected components of $\mathbb{R}^2 \setminus \gamma$ or, equivalently, the number of self-intersections $\iota_\gamma - 1$ of the boundary curve.⁶⁴ The average area of the bounded simply connected components is $a = \frac{1}{\iota_\gamma} A_{\text{arith}}$.
- (iii) The average number of overlaps $\nu_\gamma = \frac{1}{\iota_\gamma} \sum_i \nu_i$, where ν_i is the number of overlaps in the i th bounded simply connected component of $\mathbb{R}^2 \setminus \gamma$, as defined in Sec. III C 2.
- (iv) The hull of the SOP is the closed curve that bounds the region occupied by the bounded simply connected components of $\mathbb{R}^2 \setminus \text{Im } \gamma$. In the case of Brownian loops, a famous conjecture by Mandelbrot [72,73] states the hull is the continuum limit of the SAP model, with Hausdorff dimension $4/3$. It is natural to conjecture that the hull of the SOP model will also be a SLE_κ curve, for a certain value of κ .
- (v) The multiplicity μ_γ .

It is interesting to study all these observables in the SOP and in the SOP_u models as well as in the natural generalizations of SOP associated with the coupling to a CFT of arbitrary central charge c .

The most thought-provoking observable of all is undoubtedly the multiplicity, which is a very subtle and surprising feature of the quantum gravity model, that has no counterpart in other random path models. Are the SOP and SOP_u models equivalent in the continuum limit, with the same critical exponents, partition functions, etc.? This would be the case if the space of SOPs having nontrivial multiplicities were of measure zero in the continuum limit. Or, on the contrary, does multiplicity play a crucial role in

the quantum gravity models, the set of configurations having a nontrivial multiplicity being of measure strictly greater than zero? Is it actually essential to ensure, for instance, the consistency with quantum mechanics? Aesthetically speaking, the possibility that an effect as subtle could play a fundamental role in quantum gravity is, of course, extremely attractive.

4. The area distribution law

The area distribution law $\rho_{\beta_q}^{\text{SOP}}$ of the SOP model in the continuum limit governs the Λ dependence of the partition function [Eq. (4.68)]. Its large area asymptotics provides an important qualitative information on the theory and governs, in particular, the behavior of the theory when Λ is negative and large. We conjecture the following.

Conjecture 4.3. The area distribution of the SOP model, i.e., pure JT gravity in zero curvature, decays exponentially at large area, with

$$\ln \rho_{\beta_q}^{\text{SOP}}(A) \underset{A \rightarrow +\infty}{\sim} -\kappa A / \beta_q, \quad (4.134)$$

where κ is a strictly positive numerical constant.

This conjecture implies that JT gravity in zero curvature is defined only for $\Lambda > \Lambda_c$, for a strictly negative critical cosmological constant $\Lambda_c = -16\pi\kappa/\beta_q$.

The motivation for this conjecture comes from the general discussion in Sec. IV B 2. At the end of this subsection, it is explained that, when $\nu = 1/2$, which is the expected value for SOPs, the so-called inflated crumpled phase *a priori* exists only for a finite range of the parameter $n^{2\nu}(t-1)$. In the continuum limit, this suggests that the theory exists only for a finite range of negative cosmological constant. This is also observed in the purely random polygon model, which has $\kappa = 2\pi$; see Eq. (4.79).

When JT gravity is coupled to $c < 0$ conformal matter, the exponent ν was predicted in [16] to be as in Eq. (4.57), which implies that $1/2 < \nu < 1$, the limit $\nu = 1$ being reached in the semiclassical regime $c \rightarrow -\infty$ studied in [29]. In these cases, based on the discussion in Sec. IV B 2, we expect the theory in flat space to exist for arbitrary values of the cosmological constant. We make the following conjecture.

Conjecture 4.4. The area distribution of JT gravity in zero curvature coupled to $c \leq 0$ conformal matter, which has a critical exponent ν given by Eq. (4.57), has the asymptotic behavior

$$\ln \rho_{\beta_q}^0(A) \underset{A \rightarrow +\infty}{\sim} -\kappa_\nu (A/\beta_q^{2\nu})^{\chi_\nu}, \quad (4.135)$$

where κ_ν is a strictly positive numerical constant and χ_ν a critical exponent, which is a strictly increasing function of ν satisfying $\lim_{\nu \rightarrow 1/2^+} \chi_\nu = 1$ and $\lim_{\nu \rightarrow 1^-} \chi_\nu = +\infty$.

⁶³Note that, for SOPs, the notions of Lévy and winding areas defined in (3.10) coincide and match with the area of the disk computed with the associated metric.

⁶⁴In the continuum limit, a generic curve will self-intersect at double points only. Such a curve drawn on the plane, thus, yields a four-regular planar graph, whose number of vertices, which equals the number of self-intersections, must be the number of faces, given by $\iota_\gamma + 1$, minus 2.

The fact that $\chi_\nu > 1$ for $\nu > 1/2$ ensures the existence of the theory for all Λ . The conjectured limit when $\nu \rightarrow \frac{1}{2}^+$ is consistent with Eq. (4.134). The conjectured limit when $\nu \rightarrow 1^-$ is consistent with the semiclassical analysis in [29]. Actually, a straightforward saddle-point analysis, using Eq. (4.68), shows that Eq. (4.135) is equivalent to the following asymptotic behavior of the JT gravity partition function:

$$\ln W^0(\beta_q, \Lambda) \underset{\Lambda\beta_q^{2\nu} \rightarrow -\infty}{\sim} \frac{\chi_\nu - 1}{\chi_\nu} (\chi_\nu \kappa_\nu)^{-\frac{1}{\chi_\nu - 1}} \left(\frac{|\Lambda| \beta_q^{2\nu}}{16\pi} \right)^{\frac{\chi_\nu}{\chi_\nu - 1}}, \quad (4.136)$$

or, equivalently, of the expected area:

$$\langle A \rangle \underset{\Lambda\beta_q^{2\nu} \rightarrow -\infty}{\sim} \left(\frac{|\Lambda| \beta_q^{2\nu}}{16\pi \chi_\nu \kappa_\nu} \right)^{\frac{1}{\chi_\nu - 1}} \beta_q^{2\nu}. \quad (4.137)$$

Comparing with the $c \rightarrow -\infty$ results in [29], we find that

$$\chi_\nu \underset{c \rightarrow -\infty}{\sim} \frac{|c|}{12} \underset{\nu \rightarrow 1^-}{\sim} \frac{1}{2(1 - \nu)}. \quad (4.138)$$

It is tempting to conjecture that this asymptotic law as a function of ν is actually exact.

Conjecture 4.5. The exponent χ_ν governing the asymptotic behavior of the area distribution, Eq. (4.135), is given by

$$\chi_\nu = \frac{1}{2(1 - \nu)}. \quad (4.139)$$

This is consistent with the limit $\lim_{\nu \rightarrow \frac{1}{2}^+} \chi_\nu = 1$. Moreover, the formula predicts that $\chi_{3/4} = 2$ and, thus, $\langle A \rangle \propto |\Lambda|$ in the limit $\Lambda \rightarrow -\infty$ in this case. This is the behavior found for the SAP model; see Eq. (4.96). It is natural to expect that flat space JT for $\nu = 3/4$ (which corresponds to $c = -8$) and the SAP model will have similar properties in the large geometry, $\Lambda \rightarrow -\infty$, limit.

V. THE MATRIX MODEL FORMULATION

A. Introduction

Matrix model techniques are extremely useful in the case of Liouville gravity, and it is, thus, natural to look for a matrix model formulation of JT gravity. We construct such a matrix model in this section, focusing on flat JT. The possible relevance of the construction to the cases of positive and negative curvature will be discussed in Sec. VI.

To avoid confusion, let us immediately emphasize that the matrix model we are discussing below is conceptually distinct from the matrix model proposed by Saad, Shenker, and Stanford in [27]. Our matrix model is designed to compute the generating function $W(t, g)$ or, more generally,

the generalized generating function $\mathcal{W}$ defined in (3.13). It is, thus, a reformulation of the combinatorial problem of counting self-overlapping curves, taking into account the multiplicity index or, more generally, discretized flat metrics on an arbitrary surface with boundaries. This is along the lines of the matrix model formulation of Liouville gravity, but it is very different from [27]. The matrix model in [27] is supposed to describe a random ensemble of quantum Hamiltonian to which JT gravity is dual in a certain limit.

The basic idea underlying our construction, the so-called dually weighted graphs, dates back to a remarkable work of Di Francesco and Itzykson [74]. Our contribution is to explain how to correctly introduce boundaries in this picture, which amounts to constructing the JT gravity loop operators.

Models of dually weighted graphs are much more difficult to study than the usual Hermitian one-matrix model or other similar models, but they have been solved in some interesting cases [75]. The model relevant to JT gravity cannot be solved straightforwardly with the existing technology, but one may hope that a solution will eventually be found by appropriately extending the powerful formalism developed in [75].

As a preparation for what follows, let us recall that the disk generating function at fixed boundary length in Liouville gravity is obtained by considering the expectation value of the so-called macroscopic loop operator $\text{tr} X^{2n}$ in the matrix integral [Eq. (4.18)]. By summing, the full generating function (4.19) is, thus,

$$\begin{aligned} W(t, g) &= \frac{1}{N} \left\langle \ln \frac{1}{\det(\mathbb{I} - gX)} \right\rangle \\ &= \frac{1}{N} \ln \left\langle \frac{1}{\det(\mathbb{I} - gX)} \right\rangle \quad \text{when } N \rightarrow \infty. \end{aligned} \quad (5.1)$$

The second equality is valid because we consider only the leading large- N approximation, which is given by a saddle point for the matrix X . It is then natural to introduce a pair of complex vector variables $(\tilde{Q}_i, Q^i)_{1 \leq i \leq N}$ and consider the free energy F defined by

$$e^{F(t, g)} = \frac{\int dX d\tilde{Q} dQ e^{-N \text{tr}(\frac{1}{2}X^2 - \frac{1}{4}X^4) - N \tilde{Q}(\mathbb{I} - gX)Q}}{\int dX e^{-N \text{tr}(\frac{1}{2}X^2 - \frac{1}{4}X^4)}}. \quad (5.2)$$

Using the identity

$$\frac{1}{\det(\mathbb{I} - gX)} = \int d\tilde{Q} dQ e^{-N \tilde{Q}(\mathbb{I} - gX)Q}, \quad (5.3)$$

we get the large- N expansion

$$F(t, g) = N^2 W^{(s)}(t) + N W(t, g) + O(N^0), \quad (5.4)$$

where $W^{(s)}(t)$ is the sphere generating function defined in (4.1).

So boundaries can be introduced either by inserting macroscopic loop operators $\text{tr} X^{2n}$ or by introducing vector variables.

B. Dually weighted graphs

1. Basic idea

When we write (5.1), we use the dual point of view, for which an interaction term $\text{tr} X^4$ in the matrix model action yields a discretized surface built out of squares and the insertion of $\text{tr} X^{2n}$ corresponds to a marked face of length $2n$. More generally, with an action

$$S(X; \{t\}) = N \text{tr} \left(\frac{1}{2} X^2 - \sum_p \frac{t_p}{p} X^p \right), \quad (5.5)$$

we get discretized surfaces built out of p -gons, the number of p -gons being counted by the power of the coupling t_p in the corresponding Feynman graph.

To build a matrix model for flat JT gravity, we want to fix the bulk curvature of the discretized surfaces to zero. According to Eq. (2.5), we thus need to have control both on the order of the vertices and on the lengths of the faces that can appear in the matrix model graphs. Associated with the three regular tessellations of the flat Euclidean plane, we have three possibilities: If we use square tiles, $p = 4$, only faces of length 4 are allowed; if we use hexagonal tiles, $p = 6$, only faces of length 3 are allowed; and if we use triangular tiles, $p = 3$, only faces of length 6 are allowed.

The trick to do this is to introduce a new set of coupling constants $\tilde{t}_q$ in the model, in such a way that each face of length q in the Feynman graph produces a factor $\tilde{t}_q$, in exactly the same way as each vertex of degree p produces a factor t_p . This is the idea of “dually weighted graphs,” developed by Di Francesco, Itzykson [74], and Kostov. One considers an external background Hermitian $N \times N$ matrix A and sets

$$\tilde{t}_q = \frac{1}{N} \text{tr} A^q. \quad (5.6)$$

The action (5.5) is generalized to

$$S(X; \{t\}, \{\tilde{t}\}) = N \text{tr} \left(\frac{1}{2} X^2 - \sum_p \frac{t_p}{p} (AX)^p \right). \quad (5.7)$$

It is then straightforward to check that each face of a Feynman graph now produces a factor of $\text{tr} A^q = N \tilde{t}_q$ instead of the usual factor of N , exactly as looked for. If

$$e^{\mathcal{F}(\{t\}, \{\tilde{t}\})} = \int dX e^{-S(X; \{t\}, \{\tilde{t}\})}, \quad (5.8)$$

one has the large- N expansion

$$\mathcal{F}(\{t\}, \{\tilde{t}\}) = \sum_{h \geq 0} N^{2-2h} \sum_{\{p_i, q_j\}} \mathcal{W}_{\{p_i, q_j\}}^{(h)} \prod_{i,j} t_i^{p_i} \tilde{t}_j^{q_j}, \quad (5.9)$$

where $\mathcal{W}_{\{p_i, q_j\}}^{(h)}$ counts the number of genus h Feynman graphs with p_i vertices of degree i and q_j faces of length j or, equivalently, in the dual picture, the number of discretized surfaces of genus h made of p_i i -gons with q_j vertices of degree j .⁶⁵

It is natural to introduce another background Hermitian matrix $\tilde{A}$ and write

$$t_p = \frac{1}{N} \text{tr} \tilde{A}^p. \quad (5.10)$$

The free energy $\mathcal{F}$ is then a function of the external matrices A and $\tilde{A}$. Standard manipulations show that

$$\begin{aligned} e^{\mathcal{F}(A, \tilde{A})} &= \int dX dV dV^\dagger e^{-\frac{N}{2} \text{tr} X^2 - N \text{tr} V^\dagger V + N \text{tr} V \tilde{A} V^\dagger A X} \\ &= \int d\tilde{X} dV dV^\dagger e^{-\frac{N}{2} \text{tr} \tilde{X}^2 - N \text{tr} V^\dagger V + N \text{tr} V^\dagger A V \tilde{A} \tilde{X}} \\ &= \int dV dV^\dagger e^{-N \text{tr} V^\dagger V + \frac{N}{2} \text{tr} (V^\dagger A V \tilde{A})^2}, \end{aligned} \quad (5.11)$$

where V is a complex matrix and $\tilde{X}$ a “dual” Hermitian matrix. These formulas make manifest the duality

$$A \leftrightarrow \tilde{A}, \quad X \leftrightarrow \tilde{X}, \quad V \leftrightarrow V^\dagger \quad (5.12)$$

which exchanges vertices and faces. One has

$$\mathcal{F}(A, \tilde{A}) = \mathcal{F}(\tilde{A}, A). \quad (5.13)$$

Equation (5.11) also makes manifest an underlying $U(N) \times U(N)$ covariance:

$$\mathcal{F}(A, \tilde{A}) = \mathcal{W}(\Omega A \Omega^\dagger, \tilde{\Omega} \tilde{A} \tilde{\Omega}^\dagger), \quad (5.14)$$

under which the variables transform as

$$\begin{aligned} A &\mapsto \Omega A \Omega^\dagger, & \tilde{A} &\mapsto \tilde{\Omega} \tilde{A} \tilde{\Omega}^\dagger, & X &\mapsto \Omega A \Omega^\dagger, \\ \tilde{X} &\mapsto \tilde{\Omega} \tilde{X} \tilde{\Omega}^\dagger, & V &\mapsto \Omega V \tilde{\Omega}^\dagger, & V^\dagger &\mapsto \tilde{\Omega} V^\dagger \Omega^\dagger, \end{aligned} \quad (5.15)$$

⁶⁵As usual, the matrix integrals are normalized to one when the couplings are zero, and the counting is made modulo the order of the automorphism groups of the graphs. See below for a discussion of these symmetry factors in the cases of interest for JT gravity.

where Ω and $\tilde{\Omega}$ are independent unitary matrices. The two independent $U(N)$ symmetries are associated with the strands forming the faces in the direct and dual representations of the Feynman graphs, respectively.

Let us remark that, at finite N , the definitions (5.6) and (5.10) imply that only $(\tilde{t}_1, \dots, \tilde{t}_N)$ and $(t_1, \dots, t_N)$ are independent couplings. But in the formal large- N expansion of the matrix integrals we are interested in, it is consistent to consider that all the couplings are independent. This is important for our purposes, since, to impose the flatness condition in the bulk, we choose

$$\begin{aligned} t_p &= \delta_{p,p_0} t, & \tilde{t}_q &= \delta_{q,q_0} \text{ for} \\ (p_0, q_0) &= (4, 4), (6, 3), \quad \text{or} \quad (3, 6). \end{aligned} \quad (5.16)$$

2. The Di Francesco–Itzykson formula

The integral (5.8) can be expressed explicitly as a sum over Young tableaux, or partitions [74]. Explicitly, for a partition $\lambda = (\lambda_1, \dots, \lambda_N)$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N \geq 0$, we define $h_i = N - i + \lambda_i$. We note Λ_N the set of partitions such that if $N = 2n$ is even, n of the h_i 's are even (noted h_i^e) and n are odd (noted h_i^o); if $N = 2n - 1$ is odd, n of the h_i 's are even and $n - 1$ are odd. We then have

$$e^{\mathcal{F}(A, \tilde{A})} = (-1)^{n(n-1)/2} \sum_{\lambda \in \Lambda_N} \frac{\prod_i (h_i^e - 1)!! \prod_j h_j^o!!}{\prod_{i,j} (h_j^o - h_i^e)} s_\lambda(A) s_{\tilde{\lambda}}(\tilde{A}), \quad (5.17)$$

where s_λ is the Schur polynomial. One can show that summing over the partitions of a given size $|\lambda| = \sum_i \lambda_i$ yields the sum over the Feynman graphs that have a given number of edges $E = |\lambda|/2$.

The main advantage of this formula is that the number of degrees of freedom is reduced from N^2 in the original matrix integral to N in the sum over partitions $\lambda = (\lambda_1, \dots, \lambda_N)$. This is technically similar to the usual diagonalization method, the partitions replacing the eigenvalues. It is then tempting to try to evaluate (5.17) in the large- N limit using a saddle-point method, assuming that a very large partition will dominate the sum. This idea is at the basis of the results presented in [75].

Let us emphasize, however, that the assumption that the sum is dominated by a single partition is very nontrivial in itself, in particular, because the sum in (5.17) involves terms of positive and negative signs. In relation to this, there are cases for which the free energy is easy to evaluate directly from the Feynman graphs, whereas the matrix integral or the sum (5.17) seem to be very complicated to compute directly. For instance, if we use flat quadrangulated surfaces, $(p_0, q_0) = (4, 4)$, then by Gauss-Bonnet we know that only the torus topology can contribute to $\mathcal{F}$ and one finds

$$e^{\mathcal{F}} = \frac{1}{\prod_{k \geq 1} (1 - t^k)^{1/4}} \quad (5.18)$$

by examining the possible Feynman graphs. We have checked that this formula matches with (5.17) up to the order of t^{12} in the small- t expansion, but we do not know a simple derivation of the identity (5.18) starting from (5.17), nor do we know how to obtain this result from a saddle point.

3. No-go theorems

The flatness condition in the bulk is enforced by choosing the couplings as in (5.16). The resulting models are rather trivial, as Eq. (5.18) shows. To make the link with JT gravity, we need to introduce boundaries.

It is very natural to try to do so by inserting macroscopic loop operators, which is the standard tool in Liouville gravity as we have reviewed in Sec. IV A 2. There are four natural loop operators that one may consider in the models, which are also the observables studied in [75]:

$$W_L^{(1)}(t) = \frac{1}{L} \left\langle \frac{1}{N} \text{tr} X^L \right\rangle, \quad W_L^{(2)}(t) = \frac{1}{L} \left\langle \frac{1}{N} \text{tr} (AX)^L \right\rangle, \quad (5.19)$$

together with their duals. The aim of this subsection is to show that these standard loop operators are trivial in our case.

Proposition 5.1. $W_L^{(1)}(t)$ is independent of t and is, thus, equal to the $t = 0$, Gaussian matrix model value

$$W_{2n}^{(1)}(t) = \frac{(2n)!}{2n(n+1)n!^2}, \quad W_{2n+1}^{(1)}(t) = 0. \quad (5.20)$$

Proof. Let us analyze the possible graphs contributing to $W_L^{(1)}(t)$. We work in the direct representation. We have V vertices of degree p_0 , one marked vertex of degree L , E edges, and F faces. It is useful to decompose $F = f + f_\bullet + f'_\bullet$, where f is the number of faces that do not go through the marked vertex, $f_\bullet$ the number of faces that go only through the marked vertex, and $f'_\bullet$ the number of faces that go through the marked vertex and through other vertices as well. We also denote by $f_{\bullet,k}$ the number of faces of length k that go only through the marked vertex and by $f'_{\bullet,k}$ the number of faces that go through the marked vertex k times and through other vertices as well. Note that the faces that do not go through the marked vertex must all have the same length q_0 and that the faces that go through the marked vertex k times and also through other vertices must have length $k + q_0$. Standard arguments yield the relations

$$2E = p_0 V + L = q_0 f + \sum_{k \geq 1} k f_{\bullet,k} + \sum_{k \geq 1} (q_0 + k) f'_{\bullet,k}, \quad (5.21)$$

$$\sum_{k \geq 1} k(f_{\bullet,k} + f'_{\bullet,k}) = L, \quad (5.22)$$

$$F - E + V = 1 \text{ (Euler)}. \quad (5.23)$$

Combining (5.21) and (5.22) yields

$$q_0(F - f_{\bullet}) = p_0 V. \quad (5.24)$$

We then get, using (5.23),

$$\frac{4 - (p_0 - 2)(q_0 - 2)}{2q_0} V + f_{\bullet} = \frac{L}{2} + 1. \quad (5.25)$$

But the prefactor in front of V is proportional to the curvature (2.5) and, thus, vanishes for the allowed values of p_0 and q_0 . This yields

$$f_{\bullet} = \frac{L}{2} + 1. \quad (5.26)$$

This formula proves that L is even, $L = 2n$, and that the number of faces that go through the marked vertex is $n + 1$. Proposition 5.1 then follows from the following elementary lemma.

Lemma 5.1. Assume that a connected planar graph has a vertex of degree $2n$ such that there exist $n + 1$ distinct faces that go through this vertex and only through this vertex. Then the planar graph has only this vertex.

The key to this result is to realize that there is always at least one face that goes only through the marked vertex and that has length one. Indeed, from (5.22), we get

$$\begin{aligned} 2n &\geq \sum_{k \geq 1} k f_{\bullet,k} \geq f_{\bullet,1} + 2 \sum_{k \geq 2} f_{\bullet,k} \\ &= -f_{\bullet,1} + 2f_{\bullet} = -f_{\bullet,1} + 2n + 2 \end{aligned} \quad (5.27)$$

and, thus, $f_{\bullet,1} \geq 2$. One then proves the lemma straightforwardly from a recursion on n , by using a simple operation of deletion of one of the faces of length one. ■

Proposition 5.2. $W_L^{(2)}(t) = 0$.

Proof. Working again in the direct representation, one observes that the faces of the graphs potentially contributing to $W_L^{(2)}$ all have length q_0 and that all the vertices have degree p_0 except one that has degree L . This marked vertex introduces a potential curvature defect on the otherwise flat discretized surface, with local Ricci scalar $R_{q_0,L}$ given by (2.5). The Gauss-Bonnet formula then yields $(2 - q_0)L = 2q_0$ or, equivalently, $L = -p_0$, a condition that cannot be satisfied. ■

C. JT gravity matrix models

1. Vector degrees of freedom

As explained at the end of Sec. V A, in ordinary Liouville gravity, boundaries may be introduced either by using loop operators or by adding vector variables. The no-go theorems of the previous section show that the usual loop operators are not adequate for JT gravity. The simplest and most pedagogical way to proceed is then to introduce vector degrees of freedom. This will also produce automatically the correct loop operators.

It is convenient to work in the direct formulation, in which the Feynman graphs of the model correspond to the discretized surfaces.⁶⁶ The degrees of freedom are a $N \times N$ Hermitian matrix X_j^i and a complex N -vector (Q^i, Q_i^*) . We introduce a polynomial

$$P(x) = \sum_{s \geq 2} g_s x^{s-2} \quad (5.28)$$

and define the free energy as

$$e^{F(A, \tilde{A}, \{g_s\})} = \int dX dQ dQ^* e^{-S(X, Q, Q^*; A, \tilde{A})}, \quad (5.29)$$

for an action

$$\begin{aligned} S(X, Q, Q^*; A, \tilde{A}, \{g_s\}) \\ = N \text{tr} \left(\frac{1}{2} X^2 - \sum_q \frac{\tilde{t}_q}{q} (\tilde{A} X)^q \right) + N Q^* (\mathbb{I} - P(\tilde{A} X) \tilde{A}) Q. \end{aligned} \quad (5.30)$$

The external matrices A and $\tilde{A}$ are as in Eqs. (5.6) and (5.10), with the couplings (5.16).

The matrix propagators form the bulk edges and the vector propagators form the boundary edges. The bulk vertices correspond to the interactions $\text{tr}(\tilde{A} X)^q$ and the boundary vertices correspond to $Q^*(\tilde{A} X)^{s-2} \tilde{A} Q$. The background matrix $\tilde{A}$ is inserted precisely in such a way that a face of length p is weighted by the factor $t_p = \frac{1}{N} \text{tr} \tilde{A}^p$, both in the case of “bulk” faces, whose edges are all made of matrix propagators, and in the case of faces touching the boundary, that have one or more edges made of vector propagators.

For the choices (5.16), the graphs correspond to discretized surfaces for which there is no curvature at the bulk vertices and for which, according to (2.7), the boundary has extrinsic curvature $k_{p_0,q}$ at a vertex $Q^*(\tilde{A} X)^{q-2} \tilde{A} Q$. This is

⁶⁶Strict consistency with the notations used so far would require one to put tildes on top of our matrix variable X , but this would be unnecessarily cumbersome.

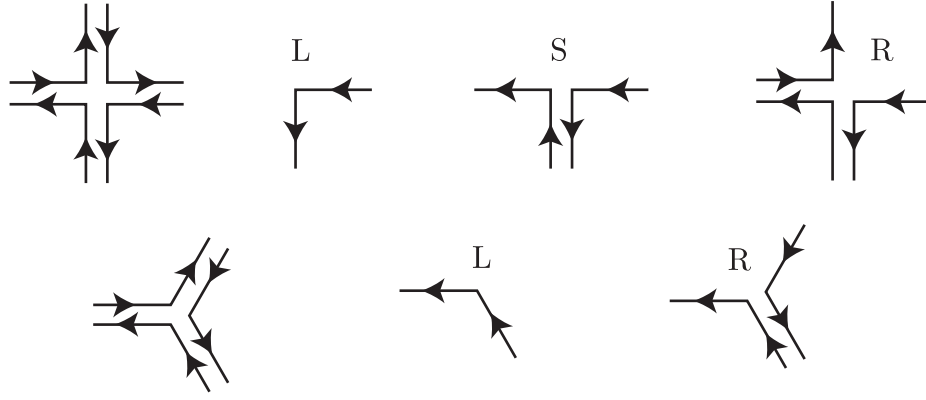


FIG. 23. Feynman rules for the $(p_0, q_0) = (4, 4)$ and $(p_0, q_0) = (6, 3)$ models (5.32) and (5.35). Double lines correspond to the matrix degrees of freedom X and simple lines, that form the boundary, to the vector degrees of freedom (Q, Q^*) . The orientation is such that the faces that form the discretized surface are to the left. The vertices are drawn isometrically. For each model, there is a unique bulk vertex (to the left). The $(4, 4)$ model has three boundary vertices L, S, and R (upper right). The $(6, 3)$ model has two boundary vertices L and R (lower right). At leading large- N order, the associated Feynman graphs are distorted disks, as in Fig. 1.

precisely what we need to build the distorted disk contributing to the JT gravity theory.

Explicitly, if we use quadrangulations, the generalized disk partition function $\mathcal{W}$ defined in (3.13), in the case where the variable u counting the multiplicity is set to one, is given by the leading large- N term in the large- N expansion of the free energy F :

$$F = \frac{1}{4} N \mathcal{W}(t, g_L, g_S, g_R, 1) + O(N^0), \quad (5.31)$$

for the model with action

$$S = N \text{tr} \left(\frac{1}{2} X^2 - \frac{1}{4} (\tilde{A} X)^4 \right) + N Q^* (\mathbb{I} - (g_L + g_S \tilde{A} X + g_R (\tilde{A} X)^2) \tilde{A}) Q \quad (5.32)$$

and background matrix $\tilde{A}$ satisfying

$$\text{tr} \tilde{A}^p = N t \delta_{p,4}. \quad (5.33)$$

The factor $1/4$ in (5.31) comes from the fact that, by convention, we count configurations modulo lattice translations but not modulo lattice rotations, as explained in Secs. II C and III B 2. See also the next subsection for a discussion of the symmetry factors.

If we use a honeycomb lattice instead, $p_0 = 6$, we have only two types of boundary vertices, left or right, and Eq. (2.11) is replaced by

$$k_{6,q} = \begin{cases} \pi/3 & \text{if } q = 2(\text{left turn}), \\ -\pi/3 & \text{if } q = 3(\text{right turn}). \end{cases} \quad (5.34)$$

The factor $1/4$ in (5.31) is then replaced by a factor $1/6$, and the action is

$$S = N \text{tr} \left(\frac{1}{2} X^2 - \frac{1}{3} (\tilde{A} X)^3 \right) + N Q^* (\mathbb{I} - (g_L + g_R \tilde{A} X) \tilde{A}) Q \quad (5.35)$$

with a background matrix $\tilde{A}$ satisfying

$$\text{tr} \tilde{A}^p = N t \delta_{p,6}. \quad (5.36)$$

One may similarly consider the case of triangulations, $p_0 = 3$. The Feynman rules are depicted in Fig. 23.

Finally, note that considering the case of a general polynomial P amounts to allowing branch points. These are clearly interesting models to consider as well, with potentially a variety of critical points and associated continuum limits. As briefly discussed in Sec. III, these models correspond to a generalization for which the boundaries are not standard self-overlapping curves but more general “interior boundaries.” Let us mention that an interesting model counting a subset of interior boundaries, called “fighting fishes,” is presented in [76].

2. Symmetry factors

Symmetry factors associated to a Feynman graph correspond to the order of the automorphism group of the graph. We would like to show that this matches precisely with the counting rules explained in Secs. II C and III B.

Let us deal with the case of quadrangulations and assume we have a graph with v_L , v_S , and v_R left, straight, and right boundary vertices, respectively, and V_B bulk vertices. The contribution of such a graph to the free energy F is computed by adding up all the possible Wick contractions of the monomial

$$\frac{g_L^{v_L} g_S^{v_S} g_R^{v_R}}{v_L! v_S! v_R! V_B! 4^{V_B}} (Q^* \tilde{A} Q)^{v_L} (Q^* \tilde{A} X \tilde{A} Q)^{v_S} (Q^* \tilde{A} X \tilde{A} X \tilde{A} Q)^{v_R} \times (\text{tr}(\tilde{A} X)^4)^{V_B} \quad (5.37)$$

that produce the graph under consideration. The prefactor comes from the expansion of e^{-S} , where S is given by (5.32), at the desired order.

Let us first consider the Wick contractions of the vector variables. The structure of the possible contractions is completely determined by the path code of the boundary, as defined in Sec. III B. Concretely, imagine that we start from a given left term $Q^* \tilde{A} Q$ in (5.37). Let us call this term the “root” [note that there are always at least four such terms; see Eq. (2.16)]. This left term must correspond to a particular left vertex in the code. Let us pick one, that we denote by v . As explained in Sec. III B, to v is associated a rooted code B_v . The allowed contractions are clearly completely fixed by the rooted code. At each step along the boundary, we must choose which vertex we contract with. This yields a total of $(v_L - 1)! v_S! v_R!$ possibilities. Taking into account the $1/(v_L! v_S! v_R!)$ prefactor in (5.37), we get a contribution $1/v_L$. According to Sec. III B, there are then three possibilities. In the generic case, the marked codes associated with the v_L left vertices are all distinct. This means that the root term $Q^* \tilde{A} Q$ we started with can be associated with v_L distinct patterns of contractions all yielding the same unrooted path code. Overall, we get a contribution $\frac{1}{v_L} \times v_L = 1$. In another case, the boundary code has a $\mathbb{Z}_2$ symmetry, which means that there are only $v_L/2$ distinct patterns of contractions and we get the symmetry factor $1/2$. In the last possible case, the boundary code has a $\mathbb{Z}_4$ symmetry, with only $v_L/4$ distinct patterns of contractions, yielding a symmetry factor $1/4$.

It remains to contract the matrix variables. Geometrically speaking, the procedure amounts to picking the vertices $\text{tr}(\tilde{A} X)^4$ one by one and attaching them to the preexisting boundary and bulk vertices, building up the distorted disk. Since the V_B vertices $\text{tr}(\tilde{A} X)^4$ are identical, we get $V_B!$ possible contraction patterns that yield precisely the same distorted disk. Moreover, each vertex can be attached in four equivalent ways, since they have a $\mathbb{Z}_4$ rotational symmetry. This yields an additional 4^{V_B} factor. We, thus, get a $V_B! 4^{V_B}$ factor that precisely cancels the corresponding prefactor in (5.37).

In conclusion, our matrix-vector model counts the distorted disks, or, equivalently, the SOP that form the boundaries, using precisely the same rules as those explained in Secs. II C and III B.

3. JT gravity loop operators and arbitrary topologies

The construction presented above works perfectly well for the disk topology. What about more general topologies?

The large- N expansion of the free energy defined in (5.29) is of the form

$$F = \sum_{k \geq 1} N^{2-k} F_k, \quad (5.38)$$

where F_k gets contributions from Feynman graphs of genus h and b boundaries, with

$$k = 2h + b. \quad (5.39)$$

The leading term selects unambiguously the disk topology, $(h, b) = (0, 1)$, with, according to Eq. (5.31), $F_1 = \frac{1}{4} \mathcal{W}$. However, the higher-order terms pick contributions from several topologies. For example, at $k = 2$, we get contributions from the torus, $(h, b) = (1, 0)$, and from the annulus, $(h, b) = (0, 2)$; at $k = 3$, from $(h, b) = (0, 3)$ and $(h, b) = (1, 1)$; etc. In the case $k = 2$, we can subtract the known torus contribution [Eq. (5.18)] to single out the annulus contribution. But, even in this simple case, we do not get the most general annulus partition function in this way, because the couplings g_L , g_S , and g_R count the vertices on both boundaries without distinction. This implies, for instance, that we cannot extract the partition function for which the lengths of the two boundaries are fixed independently of each other. When $k \geq 3$, the situation is even less favorable, because contributions from various topologies cannot be disentangled.

This problem is easily solved in the case of Liouville theory, by using loop operators. For instance, generalizing Eq. (4.18), the number $\mathcal{W}_{k_1, \dots, k_b, p}^{(h)}$ of quadrangulations of genus h surfaces with b boundaries of lengths $k_1, \dots, k_b$ and area p is obtained by computing the connected correlator of macroscopic loop operators:

$$\begin{aligned} & \left\langle \frac{1}{k_1} \text{tr} X^{k_1} \dots \frac{1}{k_b} \text{tr} X^{k_b} \right\rangle_c \\ &= \sum_{h \geq 0} N^{2-2h-b} \mathcal{W}_{k_1, \dots, k_b}^{(h)}(t) \\ &= \sum_{h \geq 0} N^{2-2h-b} \sum_{p \geq 0} \mathcal{W}_{k_1, \dots, k_b, p}^{(h)} t^p \end{aligned} \quad (5.40)$$

in the one-matrix model with action $N \text{tr}(\frac{1}{2} X^2 - \frac{t}{4} X^4)$.

For JT gravity, we have explained in Sec. V B 3 that the naive *Ansätze* for the loop operators do not work. But we now have all the ingredients to construct the correct loop operators.

To motivate the construction, let us start from the matrix-vector integral (5.29) and let us integrate out the vector variables explicitly. For concreteness, we focus on the case of quadrangulations, with action (5.32) and $\text{tr} \tilde{A}^p = N t \delta_{p,4}$ and $\text{tr} A^p = N \delta_{p,4}$ (the case of hexagons or triangles is very similar). We get

$$e^F = \int dX \frac{1}{\det(\mathbb{I} - P(\tilde{A}X)\tilde{A})} e^{-N\text{tr}[\frac{1}{2}X^2 - \frac{1}{4}(\tilde{A}X)^4]} \\ = \int dX e^{-N\text{tr}[\frac{1}{2}X^2 - \frac{1}{4}(\tilde{A}X)^4] + \sum_{k \geq 1} \frac{1}{k} \text{tr} \phi(X)^k}, \quad (5.41)$$

where $\phi(X)$ is the Hermitian matrix defined by

$$\phi(X) = P(\tilde{A}X)\tilde{A} = g_L \tilde{A} + g_S \tilde{A}X\tilde{A} + g_R \tilde{A}X\tilde{A}X\tilde{A}. \quad (5.42)$$

At leading large- N order, we can write

$$e^{NF_1 + O(N^0)} = \frac{\int dX e^{-N\text{tr}[\frac{1}{2}X^2 - \frac{1}{4}(\tilde{A}X)^4] + \sum_{k \geq 1} \frac{1}{k} \text{tr} \phi(X)^k}}{\int dX e^{-N\text{tr}[\frac{1}{2}X^2 - \frac{1}{4}(\tilde{A}X)^4]}} \\ = \left\langle e^{\sum_{k \geq 1} \frac{1}{k} \text{tr} \phi^k} \right\rangle = e^{\sum_{k \geq 1} \frac{1}{k} \langle \text{tr} \phi(X)^k \rangle + O(N^0)}, \quad (5.43)$$

where in the last equality we have used the standard large- N factorization of correlation functions, $\langle \text{tr} \phi^{k_1} \text{tr} \phi^{k_2} \rangle = \langle \text{tr} \phi^{k_1} \rangle \langle \text{tr} \phi^{k_2} \rangle + O(1/N)$. This yields

$$F_1 = \frac{1}{4} \mathcal{W} = \sum_{k \geq 1} \frac{1}{k} \left\langle \frac{1}{N} \text{tr} \phi(X)^k \right\rangle \quad (5.44)$$

at leading large- N order. The disk generating function at fixed boundary length is, thus,

$$\frac{1}{4} \mathcal{W}_k = \frac{1}{k} \left\langle \frac{1}{N} \text{tr} \phi(X)^k \right\rangle \quad \text{when } N \rightarrow \infty. \quad (5.45)$$

This is the exact analog of (4.18), with a loop operator for JT gravity identified with $\frac{1}{k} \text{tr} \phi(X)^k$.

More generally, we can write the analog of Eq. (5.40) for JT gravity:

$$\left\langle \frac{1}{k_1} \text{tr} \phi_1(X)^{k_1} \cdots \frac{1}{k_b} \text{tr} \phi_b(X)^{k_b} \right\rangle_c = \sum_{h \geq 0} N^{2-2h-b} \mathcal{W}_{k_1, \dots, k_b}^{(h)}, \quad (5.46)$$

where $\mathcal{W}_{k_1, \dots, k_b}^{(h)}$ is the generating function for the number of connected flat discretized surfaces of genus h with b boundaries. The operators $\phi_a(X)$, $1 \leq a \leq b$, are given by

$$\phi_a(X) = g_{La} \tilde{A} + g_{Sa} \tilde{A}X\tilde{A} + g_{Ra} \tilde{A}X\tilde{A}X\tilde{A}. \quad (5.47)$$

If we do not want to keep track of the number of left, straight, and right boundary vertices but only on the lengths of the boundary components, Eqs. (5.46) and (5.47) simplify to

$$\left\langle \frac{1}{k_1} \text{tr} \phi(X)^{k_1} \cdots \frac{1}{k_b} \text{tr} \phi(X)^{k_b} \right\rangle_c = \sum_{h \geq 0} N^{2-2h-b} W_{k_1, \dots, k_b}^{(h)}(t), \quad (5.48)$$

for an operator

$$\varphi(X) = \tilde{A} + \tilde{A}X\tilde{A} + \tilde{A}X\tilde{A}X\tilde{A}. \quad (5.49)$$

In the case of the honeycomb lattice, we work with $\text{tr} \tilde{A}^p = N t \delta_{p,6}$, $\text{tr} A^p = N \delta_{p,3}$, and a simpler loop operator $\varphi = \tilde{A} + \tilde{A}X\tilde{A}$.

4. A matrix formula for the multiplicity

An interesting application of the loop operator formalism is to derive a matrix integral formula for the multiplicity index μ_γ associated with an arbitrary polygon γ . Recall that μ_γ is a positive integer that counts the number of distinct distorted disks bounded by γ ; see Sec. III D. In particular, μ_γ is nonzero if and only if γ is a self-overlapping polygon.

As usual, for concreteness, we work with the square lattice, the generalization to the hexagonal and triangular lattices being straightforward. We introduce

$$\varphi_L = \tilde{A}, \quad \varphi_S = \tilde{A}X\tilde{A}, \quad \varphi_R = \tilde{A}X\tilde{A}X\tilde{A}. \quad (5.50)$$

Let γ be an arbitrary polygon, characterized by a certain path code, as defined in Sec. III B. Call γ_v the polygon γ rooted at a certain vertex v , with associated rooted code B_v , as in Sec. III B 2. To γ_v , we associate the matrix φ_{γ_v} , which is obtained by taking the product of φ_L , φ_S , and φ_R , ordered according to the rooted path code. For instance, for $B_v = \text{LSLSLLRL}$, we have $\varphi_{\gamma_v} = \varphi_L \varphi_S \varphi_L \varphi_S \varphi_L \varphi_L \varphi_R \varphi_L$. The operator $\text{tr} \varphi_{\gamma_v}$ does not depend on the marked vertex, thanks to the invariance of the trace under cyclic permutations.

Definition 5.1. The loop operator associated with a polygon γ is defined to be $\text{tr} \varphi_{\gamma_v}$, where v is an arbitrary marked vertex on γ . Since the loop operator does not depend on the choice of v , we denote it by $\text{tr} \varphi_\gamma$.

Theorem 5.1. We consider the large- N Hermitian matrix model with action

$$S = N\text{tr} \left(\frac{1}{2} X^2 - \frac{1}{4} (\tilde{A}X)^4 \right), \quad (5.51)$$

where the external Hermitian matrix $\tilde{A}$ is such that $\text{tr} \tilde{A}^p = N t \delta_{p,4}$. The expectation value of the loop operator associated with a polygon γ , satisfying the constraint $v_L - v_R = 4$ [Eq. (2.16)], in this matrix model, is given by

$$\langle \text{tr} \varphi_\gamma \rangle = N \mu_\gamma t^{p_\gamma}, \quad (5.52)$$

where μ_γ and p_γ are the multiplicity of γ and the area of the distorted disks it bounds, respectively (recall from Sec. III C 2, Corollary 3.2, that the area depends only on γ , even in the cases where $\mu_\gamma > 1$).

Proof. The proof is based on a slight refinement of the discussion made in Sec. V C 2. A boundary code γ being

given, it was shown that the number of Wick contractions of the vector variables in (5.37) associated with γ is $v_L!v_S!v_R!/\sigma_\gamma$, where σ_γ is the symmetry factor associated with γ (one for a generic polygon and two or four for a $\mathbb{Z}_2$ or $\mathbb{Z}_4$ symmetric polygon). The new piece of information we need is that these contractions precisely yield the loop operator $\text{tr}\varphi_\gamma(X)$. This is straightforward to check. The subsequent contractions of the matrix variables and sum over the number V_B of $\text{tr}(\tilde{A}X)^4$ insertions amounts, on the one hand, to computing $\frac{1}{\sigma_\gamma}\langle\text{tr}\varphi_\gamma\rangle$ and, on the other hand, to reconstructing all the distorted disks bounded by γ , yielding the contribution $Nt^{p_\gamma}\mu_\gamma/\sigma_\gamma$ to the free energy F . We conclude. ■

Equation (5.52) is exact at any order in the large- N expansion; it, thus, implies that the expectation value of the loop operator gets contributions only from the disk topology. If we waive the constraint $v_L - v_R = 4$, Eq. (2.14) predicts that $\langle\text{tr}\varphi_\gamma\rangle = 0$ if $v_L - v_R$ is not of the form $4(1 - 2h)$ for a positive integer h and gets a contribution only at genus h if $v_L - v_R = 4(1 - 2h)$. Similarly, we could compute connected correlators $\langle\text{tr}\varphi_{\gamma_1}\cdots\text{tr}\varphi_{\gamma_b}\rangle_c$ in the matrix model to decide whether the curves γ_i bound a flat surface.

Investigating these questions and more generally finding a solution to JT gravity, on the disk and possibly on higher topologies as well, using the matrix model, is clearly a challenging and fascinating problem for the future.

Note. Very recently, the problem of counting so-called *rigid quadrangulations* of the flat disk was solved [71,77]. The rigid quadrangulations, defined in [71], form a strict subset of the set of discretized disks considered in our paper, for which the allowed rays must join R and S vertices (the notion of ray was defined in Sec. IV C 2). It is very plausible that the continuum limit of the model of rigid quadrangulations may be identical to that of the more general model of discretized disks. As mentioned in [71], in a rigid quadrangulation, the bulk edges are oriented from R to S, suggesting a formulation in terms of a complex matrix model. Using the ideas introduced in the present section, it is indeed straightforward to build the matrix model that counts the rigid quadrangulations. The degrees of freedom are a $N \times N$ complex matrix X and a complex N vector (Q, Q^*) , with a complex action

$$S = N\text{tr}(X^\dagger X - (\tilde{A}X)^2(\tilde{A}X^\dagger)^2) + NQ^*(\mathbb{I} - (g_L + g_S\tilde{A}X + g_R(\tilde{A}X^\dagger)^2)\tilde{A})Q \quad (5.53)$$

and background matrix $\tilde{A}$ as in Eq. (5.33). The analysis presented in [71,77] gives the exact disk partition function of this model for $t = 1$ [one can show that only the combination $g_L^4(g_L g_S^2 g_R)^{k-3}$, for $k \geq 3$, can enter in the Feynman graphs; one can, thus, set $g_L = g_S = g_R = g$ without loss of generality]. The counting at $t \neq 0$, that is

to say, with a control on the area, remains an outstanding open problem.

VI. THE CASES OF NONZERO CURVATURE

We have already briefly discussed the case of nonzero curvature in Sec. III E 4. We can now go into more detail and attempt to paint a general qualitative picture of the properties expected for negative and positive curvatures.

A. Qualitative features

There are two important properties that we would like to emphasize.

The microscopic structure of the models cannot depend on the curvature of space-time. If L is the curvature length scale, defined by $R = \pm 2/L^2$, then both the positive and the negative curvature models will be well approximated by the flat model on distance scales much less than L . When $\Lambda \geq 0$, this will be the case as soon as the quantum length of the boundary β_q is much less than L^2 .⁶⁷ In the negative curvature model, if Λ is negative, one must also impose that $\beta_q|\Lambda|$ is not too large. We shall argue that the positive curvature model is actually well defined only for $\Lambda \geq 0$. In terms of the partition functions, which were formally defined in Eq. (2.49), one must have

$$\lim_{L \rightarrow +\infty} W^\pm(\beta_q, \Lambda, L) = W^0(\beta_q, \Lambda), \quad (6.1)$$

the limit being taken for $\Lambda \geq 0$ in positive curvature or for either $\Lambda \geq 0$ or $\Lambda < 0$ and $\beta_q|\Lambda|$ not too large in negative curvature.

The fact that all three models have the same short-distance behavior is equivalent to saying that they have the same UV completion.

The models in negative, zero, and positive curvature differ in very important ways on large-distance scales. These differences will be made very precise below, in particular, by offering conjectures on the large area asymptotics of the area probability densities. At a qualitative level, we can understand the very different behavior of the three models at long distances by referring to the existence, or nonexistence, of isoperimetric inequalities. In negative curvature, for a disk having a smooth boundary of length ℓ , the area is bounded above by the area of the round disk:

$$A \leq 2\pi L^2 \left[-1 + \sqrt{1 + \frac{\ell^2}{4\pi^2 L^2}} \right] \quad \text{when } R = -2/L^2. \quad (6.2)$$

For large configurations, the bound grows linearly with the boundary length, $A \leq L\ell$. In zero curvature, the upper

⁶⁷Recall that β_q has the dimension of a length square, because the exponent $\nu = 1/2$ in pure JT gravity; see Eq. (4.57) and [16].

bound is still given by the round disk:

$$A \leq \frac{\ell^2}{4\pi} \quad \text{when } R = 0. \quad (6.3)$$

It now grows quadratically with the boundary length. In positive curvature, there is no bound. This will be illustrated explicitly below, in Sec. VI C.

If the boundaries were smooth in the quantum theory, the above bounds would imply that the area probability densities in negative and zero curvature would have compact support. Instead, the boundary is a fractal, with a fixed quantum boundary length β_q but an infinite smooth length ℓ . There is, thus, no bound on the possible areas in the quantum theory, and the probability densities can have tails that extend to infinity, even in negative or zero curvature. It is natural to expect that the densities will decay much faster in negative curvature than in zero curvature, since the smooth bound as a function of the boundary length is stricter in negative curvature, and much faster in zero curvature than in positive curvature, since there is no smooth bound in the latter case. The decay laws of the area probability densities can be interpreted as quantum versions of the classical isoperimetric inequalities. Conjectures about their explicit form are given below; see Conjectures 4.3, 4.4, 6.1, and 6.5.

The rapid decay in negative curvature implies that large areas are strongly suppressed; this ensures the existence of the theory for any value of the cosmological constant, including the regime of large negative Λ where the effective Schwarzian description is emerging. In zero curvature, we conjecture an exponential decay, implying that the pure gravity model exists for a range of cosmological constant $\Lambda \geq \Lambda_c$, where Λ_c is finite and negative. In positive curvature, we conjecture a power-law decay. Large area configurations are then ubiquitous and the theory exists only for positive cosmological constants.

B. Negative curvature

1. Area distribution law

Conjecture 6.1. The area distribution of JT gravity in negative curvature coupled to $c \leq 0$ conformal matter, with critical exponent ν given by Eq. (4.57), has the asymptotic behavior

$$\ln \rho_{\beta_q}^-(A) \underset{A \rightarrow +\infty}{\sim} -\kappa_\nu^- \left(\frac{A}{\beta_q^\nu L} \right)^{\chi_\nu^-}, \quad (6.4)$$

where κ_ν^- is a strictly positive numerical constant and χ_ν^- a critical exponent, which is a strictly increasing function of ν satisfying $\lim_{\nu \rightarrow \frac{1}{2}^+} \chi_\nu^- = 2$ and $\lim_{\nu \rightarrow 1^-} \chi_\nu^- = +\infty$.

The fact that $\chi_\nu^- \geq 2 > 1$ ensures the existence of the theory for all Λ , including in the case of the pure gravity

theory, for which χ_ν^- is conjectured to be 2. The limit when $\nu \rightarrow 1^-$ follows from the semiclassical analysis of [29]. The limit $\nu \rightarrow \frac{1}{2}^+$ is motivated in part by a similar behavior found for purely random Brownian paths in hyperbolic space [40]; see also the discussion in Sec. VI B 2 below.

In a very similar way to the discussion in Sec. IV C 4, we observe that Eq. (6.4) is equivalent to the asymptotic behavior

$$\ln W^-(\beta_q, \Lambda, L) \underset{\Lambda \rightarrow -\infty}{\sim} \frac{\chi_\nu^- - 1}{\chi_\nu^-} (\chi_\nu^- \kappa_\nu^-)^{-\frac{1}{\chi_\nu^- - 1}} \left(\frac{|\Lambda| \beta_q^\nu L}{16\pi} \right)^{\frac{\chi_\nu^-}{\chi_\nu^- - 1}} \quad (6.5)$$

for the JT gravity partition function in negative curvature or, similarly, to the asymptotic behavior

$$\langle A \rangle \underset{\Lambda \rightarrow -\infty}{\sim} \left(\frac{|\Lambda| \beta_q^\nu L}{16\pi \chi_\nu^- \kappa_\nu^-} \right)^{\frac{1}{\chi_\nu^- - 1}} \beta_q^\nu L \quad (6.6)$$

for the expected area.

These formulas are consistent with the $c \rightarrow -\infty$ analysis presented in [29], with

$$\chi_\nu^- \underset{c \rightarrow -\infty}{\sim} \frac{|c|}{6} \underset{\nu \rightarrow 1^-}{\sim} \frac{1}{1 - \nu}. \quad (6.7)$$

They are also consistent with the discussion in Sec. VI B 2 below. We conjecture that the asymptotic law (6.7), as a function of ν , is exact.

Conjecture 6.2. The exponent χ_ν^- governing the asymptotic behavior of the area distribution [Eq. (6.4)] is given by

$$\chi_\nu^- = \frac{1}{1 - \nu}. \quad (6.8)$$

2. Area distribution, ballistic behavior, and the Schwarzian limit

As already mentioned repeatedly, e.g., in Sec. II E 1, one expects that the negative curvature JT gravity theory has an effective description in the $\Lambda \rightarrow -\infty$ limit, valid on scales much larger than the curvature length scale L , in terms of near-hyperbolic geometries described by the reparametrization *Ansatz* and the Schwarzian action. These effective geometries have a smooth boundary, with a fixed length ℓ . This length is an effective macroscopic parameter which can be expressed, in principle, in terms of the microscopic parameters β_q , Λ , and L . We now have the necessary ingredients to derive this relation.

But first things first, in order to get an intuitive understanding of why and how the Schwarzian description may emerge, it is useful to recall the behavior of Brownian motion in hyperbolic space; see, e.g., [78]. If we let a Brownian particle move for a diffusion (“quantum”) time

β_q , then the average geodesic distance between the particle and its initial position will start to grow as the square root of β_q , as it does in flat space. Precisely, with the conventions used in Sec. IV B 4 to define the Brownian bridge, this average distance is $\sqrt{\pi\beta_q}/2$. The square root behavior is directly related to the fact that the exponent ν is $1/2$ for Brownian motion. In flat space, this behavior remains valid for all β_q . But in hyperbolic space, the situation is drastically different. The square root growth is valid as long as $\beta_q \ll L^2$, but when $\beta_q \gg L^2$ one finds a completely different “ballistic” regime for which the average geodesic distance grows linearly, as $\beta_q/(2L)$. The linear growth with β_q indicates that, at large distances, an effective description emerges, with an effective exponent $\nu_{\text{eff}} = 1$ corresponding to a smooth trajectory.

We claim that the transition from the flat space description to the Schwarzian description in negative curvature JT gravity is of a similar nature.⁶⁸ The long-distance, large geometry limit is obtained when the cosmological constant $\Lambda \rightarrow -\infty$. The emergent macroscopic length parameter ℓ can be naturally defined by the asymptotic expected area, consistently with the leading term in Eq. (2.64):

$$\langle A \rangle_{\Lambda \rightarrow -\infty} \sim \ell L. \quad (6.9)$$

The fact that the boundary will be effectively smooth means that ℓ will be proportional to β_q , exactly as for the ballistic regime of Brownian motion. Assuming a simple power-law dependence in Λ and L , dimensional analysis implies that

$$\ell = a\beta_q |\Lambda|^\epsilon L^{1-1/\nu+2\epsilon}, \quad (6.10)$$

where a is a numerical constant of proportionality and ϵ an exponent. The results of [29] yield the $c \rightarrow -\infty$ behavior:

$$\epsilon = \frac{6}{|c|} + O(|c|^{-2}). \quad (6.11)$$

The linearity of ℓ with β_q also has the following nice consequence. At the microscopic level, there is a boundary cosmological constant counterterm that is proportional to the quantum boundary length β_q .⁶⁹ This means that the logarithm of the partition function of the theory is defined modulo the addition of terms that are proportional to β_q . In the effective Schwarzian description of the theory, such a boundary cosmological counterterm is usually used to

cancel the diverging leading ℓL term in the Schwarzian area expansion [Eq. (2.64)]. But a term proportional to the macroscopic length parameter ℓ in the effective description is actually an allowed counterterm from the fundamental, microscopic point of view only if ℓ is itself proportional to β_q . The ballistic property $\ell \propto \beta_q$ ensures that this is indeed the case.

Note that Eqs. (6.9) and (6.10) are consistent with (6.6), with $\epsilon = 1/(\chi_\nu^- - 1)$. Using the conjecture (6.8), we get the prediction

$$\epsilon = \frac{1}{\nu} - 1. \quad (6.12)$$

In the pure gravity case, $\nu = 1/2$, $\epsilon = 1$, and Eq. (6.10) yields $\ell \propto \beta_q |\Lambda| L$. In particular, the dimensionless Schwarzian limit parameter β_s defined in Eq. (2.50) turns out to be simply proportional to β_q/L^2 , in pure gravity.

3. Lattice formulation and the Schwarzian limit

Up to now, we have (extensively) discussed the lattice formulation of JT gravity only in zero curvature. It is, of course, natural to look for a similar formulation with nonzero curvature and, in particular, for negative curvature, which is the case we are interested in in the present section. This turns out to be more subtle and interesting than one might have expected and sheds an unexpected light on the Schwarzian regime.

A straightforward generalization of the lattice formulation in zero curvature, that uses regular tessellations of the Euclidean plane, is to consider the self-overlapping polygon model on regular tessellations of hyperbolic space, also called “hyperbolic lattices.” As in flat space, these tessellations are characterized by the number of edges p of the regular polygons forming the lattice tiles and the degree q of the vertices at which the tiles meet. The integers p and q must satisfy the inequality (3.14), already mentioned in Sec. III E 4, that follows from the formula (2.5) for the curvature and the constraint of negative curvature.

A crucial difference with the case of flat space is that the regular tessellations of hyperbolic space cannot be scaled down to take a continuum limit. The edge length and polygon area are indeed fixed in terms of the curvature length scale and the values of p and q . For instance, in the construction in Sec. II, for which the polygons are flat, of edge length ℓ_0 , and the curvature is localized at the vertices, the area of a regular p -gon is

$$A_0 = \frac{p\ell_0^2}{4 \tan(\pi/p)} \quad (6.13)$$

and the area per vertices is qA_0/p , since a given vertex belongs to q different p -gons. If we denote by $-2/L^2$ the average curvature, the integral of the curvature per vertex is, thus, $-2qA_0/(pL^2)$. On the other hand, this integral is also given by $R_{p,q}$, as defined in (2.5). This yields

⁶⁸And a similar mechanism is undoubtedly responsible for the fact that the Kitaev-Suh model for JT gravity, which is based on Brownian motion in hyperbolic space, also has a Schwarzian description at long distances [32,40].

⁶⁹At a fundamental level, this comes from the exponential growth of the number of configurations with the discretized boundary length n and the fact that β_q is proportional to n ; see Sec. IV B 3 and, in particular, Eq. (4.63).

$$A_0 = \frac{\pi L^2}{q} ((p-2)(q-2) - 4). \quad (6.14)$$

Thus, the area of the basic building block of the lattice is proportional to L^2 , and a continuum limit in which this area goes to zero does not exist. Of course, the edge length ℓ_0 itself can be made very small when p is very large at fixed q and L , Eqs. (6.13) and (6.14) implying that $\ell_0^2 \sim 4\pi^2 \frac{q-2}{q} \frac{1}{p} L^2$, but at the same time A_0 goes to infinity, proportionally to p . Thus, this is not suitable for a continuum limit. It is actually easy to check that the polygon area is bounded below by $\frac{1}{7}\pi L^2$, a bound saturated for $(p, q) = (3, 7)$.

The conclusion is that a lattice model for negative curvature JT gravity based on regular tessellations (or similar tessellations, like the semiregular tessellations, for which the microscopic length and area are also fixed by L) cannot yield a microscopic description of the theory, valid on arbitrarily small length scales, but at best an effective description, valid on length scales much greater than L . Obviously, we have a very natural candidate for this effective theory, for which the “UV” cutoff is L : It is the Schwarzian theory.

Conjecture 6.3. The Schwarzian effective description of negative curvature JT gravity is the continuum limit of lattice models of self-overlapping polygons, defined on regular or semiregular tessellations of hyperbolic space. This continuum limit corresponds to taking the curvature length scale $L \propto \ell_0$ to zero and the polygon length n to infinity, with the combination nL held fixed and proportional to the macroscopic length parameter ℓ . In particular, the critical exponent ν and the Hausdorff dimension $1/\nu$ are one in this case.

One may actually conjecture that the Schwarzian description is universal and is not limited to the SOP model or JT gravity. Any model of random closed curves in hyperbolic space having a suitable macroscopic limit will likely be described by the Schwarzian on length scales much greater than the curvature length scale. Strong evidence for this in the case of purely random curves was discussed above and given in [32]. Self-avoiding polygons on hyperbolic lattices have also been studied in the literature [58]. The ballistic property was proven in [59], consistently with the idea that a Schwarzian description will also emerge for these models in the limit $\Lambda \rightarrow -\infty$.

The intuitive reason explaining why ballistics seems to be generic in hyperbolic space is that the number of lattice vertices accessible after n steps from a given vertex grows exponentially with n for hyperbolic lattices instead of quadratically for flat space lattices.⁷⁰ A random path starting from a given point has, thus, many more

possibilities to extend far away in hyperbolic space than in flat space.

To obtain a microscopic definition of JT gravity in negative curvature by using a lattice formulation, one thus needs to refine the regular tessellations in one way or another, in order to create a lattice mesh of arbitrarily small length scales. There are many ways to do this, cutting the polygons of the regular tessellations into smaller pieces; see, e.g., the discussion in Sec. V of the first reference in [57].

4. Matrix model

The matrix model formulation described in Sec. V in flat space can be straightforwardly generalized to the case of negative curvature. One simply chooses the matrices A and $\tilde{A}$ such that

$$\text{tr} \tilde{A}^k = N t \delta_{k,p}, \quad \text{tr} A^k = N \delta_{k,q} \quad (6.15)$$

for integers p and q satisfying the inequality (3.14), $(p-2)(q-2) > 4$. The coupling to the vector variables is as in Eq. (5.30), and the loop operator is $\phi(X) = P(\tilde{A}X)\tilde{A}$, along the lines of Sec. VC 3. The polynomial P can be naturally chosen to be of degree $q-2$. The matrix model partition function is then the generating function for self-overlapping polygons on the hyperbolic lattice (p, q) , the allowed polygons being defined as in Sec. III. The fact that $\deg P = q-2$ ensures that the only constraint on successive points P_k and P_{k+1} along the polygon is that the edge joining P_k and P_{k+1} is not the same as the edge joining P_{k-1} and P_k .

As mentioned in the previous subsection, such a model has a short-distance cutoff set by the curvature length scale L . Its “continuum” limit is expected to yield the Schwarzian theory.

However, it might be the case that the matrix model is more powerful than this and could give access to the full UV-complete theory in curved space. The idea is that there may exist a natural and physically relevant way to analytically continue the matrix model partition function and correlators as a function of the parameters p and q . Actually, only the analytic continuation in one of the parameter is needed, e.g., in q . Let us then set $p = 3, 4$, or 6 and consider q to be very near $6, 4$, or 3 , such that the curvature $R_{p,q}$ [Eq. (2.5)] can be made arbitrarily small. This trick will make it possible to fix the curvature so that it is very small in the lattice spacing units. For concreteness, let us work with a square lattice $p = 4$ as we have done in most of this paper, and let us write

$$q = 4(1 + \varepsilon), \quad (6.16)$$

where ε is small and positive. To leading order in ε , Eqs. (6.13) and (6.14) yield (for $p = 4$)

⁷⁰This is the lattice version of the well-known exponential growth of a round disk area A as a function of its geodesic radius r , for large r , $A \sim \pi L^2 e^{r/L}$.

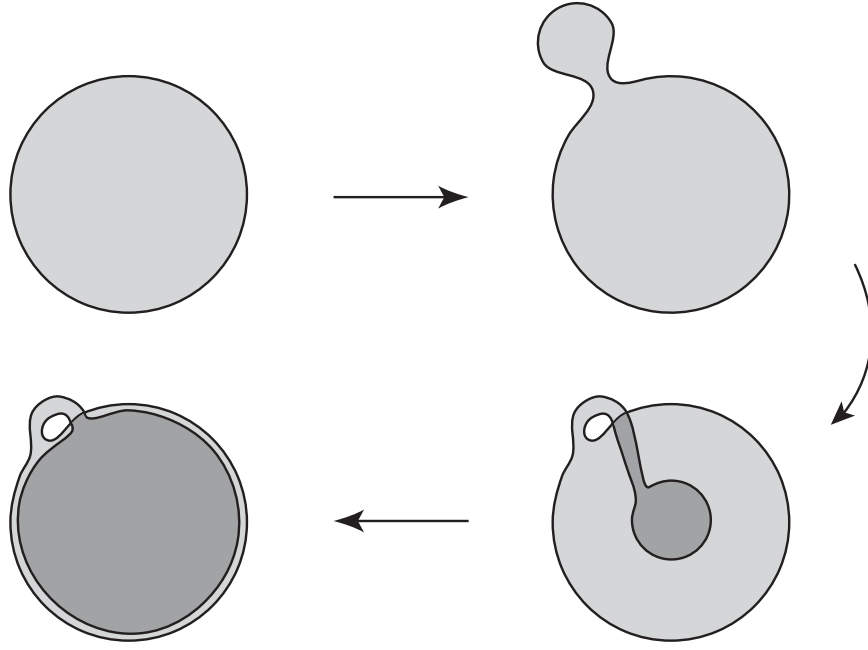


FIG. 24. One step in the construction of a metric of constant positive curvature on the disk having arbitrarily large area and arbitrarily small boundary length. The distorted disks are drawn on the two-sphere in projective coordinates, with the metric given by Eq. (2.39). In the process depicted in the figure, the boundary length and the area are both multiplied by a number that can be made arbitrarily close to two. By repeating this operation n times and making the figure as large as we want in the last step, we construct a metric of area $\sim 4\pi n$ and arbitrarily small boundary length. See the main text for more details.

$$\ell_0^2 = 2\pi L^2 \varepsilon. \quad (6.17)$$

By taking $\varepsilon \rightarrow 0^+$, we can therefore obtain, in principle, a genuine continuous limit for the theory of negative curvature. This is summarized by the following (bold) conjecture.

Conjecture 6.4. The disk partition function

$$W_n(t; p = 4, q) = \frac{4}{N} \left\langle \frac{1}{N} \text{tr} \varphi(X)^n \right\rangle \quad (6.18)$$

of the matrix model with action

$$S = N \text{tr} \left(\frac{1}{2} X^2 - \frac{1}{q} (\tilde{A} X)^q \right) \quad (6.19)$$

for $\text{tr} \tilde{A}^k = N t \delta_{k,4}$ and loop operator

$$\varphi(X) = \sum_{s=0}^{q-2} (\tilde{A} X)^s \tilde{A} \quad (6.20)$$

admits an analytic continuation in the variable q , in the vicinity of $q = 4$, such that the limit

$$n \rightarrow \infty, \quad \ell_0^2 = \frac{\pi L^2}{2} (q - 4) \rightarrow 0^+, \quad t \rightarrow 1, \quad n \ell_0^2 = \beta_q, \quad \text{and} \quad \Lambda = \frac{16\pi}{\ell_0^2} (1 - t) \text{ fixed} \quad (6.21)$$

yields the UV-complete disk partition function of pure JT gravity in negative curvature $R = -2/L^2$.

Of course, similar conjectures can be made when $p = 3$ or 6.

C. Positive curvature

1. Area distribution law

A crucial qualitative feature of the positive curvature theory is the existence of smooth geometries that can have arbitrarily large areas and arbitrarily small boundary lengths at the same time. A step in the construction of such a geometry is illustrated in Fig. 24. The drawings depict immersions of the disk on the sphere, which is represented in projective coordinates with metric given by Eq. (2.39) (the units are such that the curvature length scale is $L = 1$). Starting from a round disk of projective coordinates radius R , boundary length $\ell(R)$, and area $A(R)$,

$$\ell = \frac{4\pi R}{1 + R^2} \underset{R \rightarrow \infty}{\sim} \frac{4\pi}{R}, \quad A(R) = R \ell(R) \underset{R \rightarrow \infty}{\sim} 4\pi, \quad (6.22)$$

we get a distorted disk of approximately twice the original boundary length and area. By repeating the same procedure n times, we obtain a distorted disk which overlaps n times with itself over a region covering a fraction of the entire two-sphere. The associated metric has a boundary length ℓ_n which is approximately $n\ell(R)$ and an area A_n which is approximately $nA(R)$. By taking $R \gg 1$, the overlapping region covers almost the entire two-sphere, which implies that $A_n \sim 4\pi n$, whereas the boundary is concentrated in a tiny region around the south pole, with $\ell_n \sim 4\pi n/R$. If we scale R with n in such a way that $n/R \rightarrow 0$ when $n \rightarrow \infty$, for instance, by taking $R \sim n^2$, we obtain a constant positive curvature metric on the disk that has arbitrarily large area and arbitrarily small boundary length.

This example illustrates a fundamental feature of the positive curvature JT gravity theory, that large geometries will be ubiquitous. This yields the following conjecture on the large area behavior of the area distribution law for positive curvature JT.

Conjecture 6.5. The area distribution of JT gravity in positive curvature coupled to $c \leq 0$ conformal matter, with critical exponent ν given by Eq. (4.57), has the asymptotic behavior

$$\rho_{\beta_q}^+(A) \underset{A \rightarrow +\infty}{\sim} \frac{\kappa_\nu^+(\beta_q/L)}{L^2} \left(\frac{L^2}{A}\right)^{\chi_\nu^+}, \quad (6.23)$$

where κ_ν^+ is a strictly positive function of the dimensionless parameter β_q/L and χ_ν^+ a critical exponent, which is a strictly increasing function of ν satisfying $\lim_{\nu \rightarrow \frac{1}{2}^+} \chi_\nu^+ = 2$ and $\lim_{\nu \rightarrow 1^-} \chi_\nu^+ = +\infty$.

The asymptotic power-law behavior (6.4) implies that the positive curvature JT gravity theory exists only for positive cosmological constant. Moreover, only a finite number of moments of the area at zero cosmological constant exist, $\langle A^k \rangle < \infty$ for $k < \chi_\nu^+ - 1$. Note that these properties are shared by Liouville gravity. Actually, in the case of Liouville, it is straightforward to derive the area distribution law ρ^L from the exact partition function given in Eq. (4.34). One finds

$$\rho^L(A) = \frac{1}{\Gamma(1 - \gamma_{\text{str}})} \left(\frac{\tilde{\ell}^2}{4}\right)^{1 - \gamma_{\text{str}}} A^{\gamma_{\text{str}} - 2} e^{-\frac{\tilde{\ell}^2}{4A}} \quad (6.24)$$

for

$$\tilde{\ell} = \frac{\ell}{\sqrt{\sin \frac{\pi}{1 - \gamma_{\text{str}}}}}. \quad (6.25)$$

The large area behavior of ρ^L is, thus, also powerlike. The moments $\langle A^k \rangle$ exist only when $k < 1 - \gamma_{\text{str}}$, and the theory is well defined only in positive cosmological constant.

The conjectured behavior $\lim_{\nu \rightarrow 1^-} \chi_\nu^+ = +\infty$ is consistent with the idea that this limit is semiclassical and all the moments of the area then automatically exist, with $\langle A^k \rangle = A_*^k$, where A_* is the area of the relevant classical background. In fact, this argument is not without its subtleties, because the simple round disk classical background in positive curvature exists only when the boundary length ℓ is less than $2\pi L$. The behavior in the $c \rightarrow -\infty$ limit is not known if this condition is not satisfied [29]. One may, thus, expect a nontrivial behavior of the function κ_ν^+ in Eq. (6.4) in the limit $\nu \rightarrow 1^-$ if $\beta_q > 2\pi L$. The conjectured behavior $\lim_{\nu \rightarrow \frac{1}{2}^+} \chi_\nu^+ = 2$ is motivated by the fact that the Brownian bridge model in positive curvature (that is to say, the generalization of the Kitaev-Suh model [32] to the case of positive curvature) is found to have an area distribution that goes as $1/A^2$ asymptotically [40], and it is plausible that JT gravity will behave similarly for very large areas.

2. Lattice formulation, matrix model

Regular tessellations of the two-sphere must satisfy the inequality $(p - 2)(q - 2) < 4$. The only solutions are the five Platonic solids, $(p, q) = (3, 3), (3, 4), (3, 5), (4, 3)$, or $(5, 3)$, with a finite number of vertices and tiles. As in the case of negative curvature, the area of the p -gons are of the order of L^2 , the precise formula being exactly as in Eq. (6.14) with a global sign change in the right-hand side of this equation. The random SOP model on such lattices seems too trivial to have any interesting physical application; in particular, contrary to the negative curvature model, there is no Schwarzian-like limit in positive curvature. To obtain an interesting lattice model for positive curvature JT, one must thus refine the Platonic solids and discretize the two-sphere with tiles that may be made arbitrarily small—a problem well studied in weather forecasting and climate modeling.

The matrix model may also be conjectured to be relevant, mimicking Conjecture 6.4. The only change with the case of negative curvature is that the small parameter ε in Eq. (6.16) is now taken to be negative and the formula (6.17) for the lattice spacing becomes $\ell_0^2 = -2\pi L^2 \varepsilon = \frac{\pi L^2}{2}(4 - q)$, which goes to 0^+ in the continuum limit.

VII. SUMMARY AND OUTLOOK

A. Upshots

We have proposed a UV-complete definition of Euclidean Jackiw-Teitelboim quantum gravity, starting from the fundamental postulate that two-dimensional quantum gravity is a diffeomorphism-invariant theory of random metrics. The model on the disk has a dual formulation in terms of a theory of random self-overlapping loops that describes the fluctuating boundary. Self-overlapping loops must bound a distorted disk, a highly

nontrivial constraint that can be analyzed algorithmically using Blank cuts. A surprising aspect of this formulation is that some self-overlapping curves bound several distinct disks and, therefore, must be counted with a nontrivial multiplicity, being associated with diffeomorphism-inequivalent metrics on the disk.

At the microscopic level, JT gravity can, thus, be reduced to a purely combinatorial problem, which amounts to counting self-overlapping polygons with the appropriate multiplicity. This is very similar to, but probably harder than, the counting of “maps” (polygonizations) of surfaces in Liouville gravity. We have constructed a matrix model that does the counting, using an upgraded version of the method of dually weighted graphs. All these ideas open the way to studying the JT theories using techniques of enumerative geometry, matrix models, and numerical simulations, which in the past have been instrumental in gaining insight into nontrivial random polygon models, such as the self-avoiding polygon model.

We have learned that the space of constant bulk curvature metrics one must consider is highly singular on the boundary, albeit perfectly smooth in the bulk. Viewed as closed curves immersed in the canonical spaces of constant curvature (hyperbolic space, Euclidean space, or two-sphere), the boundaries are fractals with a nontrivial Hausdorff dimension d_H , which is conjectured to be two in the case of the pure gravity model. In particular, the boundaries do not have a smooth geometric length. Instead, one can associate to them a quantum length β_q , which has geometric length dimension d_H . This quantum length is analogous to the diffusion time, or Euclidean quantum mechanical time, that is associated to Brownian paths. At the microscopic level, it is proportional to the number of discretized, UV lattice length edges that form the boundary, but in the continuum limit it takes on an anomalous dimension, reflecting the very singular nature of the boundary metric. As is made clear elsewhere [16,28], the boundary metric is actually distribution valued.

The short-distance properties of the three models of JT gravity (negative, zero, or positive curvature) are identical: They have the same UV completion. On the other hand, the IR properties drastically differ. The large-distance behavior is governed by quantum versions of the classical isoperimetric inequalities. In negative curvature, the probability density for the area at fixed quantum boundary length β_q is conjectured to decay as fast as $\exp(-\kappa_- A^2/(\beta_q L^2))$, where κ_- is a strictly positive numerical constant and L the curvature length scale, whereas in zero curvature a simple exponential decay is expected. In positive curvature, the absence of a classical isoperimetric inequality is conjectured to imply a power-law decay. Large area configurations become ubiquitous, and the model is well defined only for a positive cosmological constant, as is also the case for the Liouville theory.

The usual formulation of the negative curvature model in terms of the Schwarzian theory emerges when the cosmological constant $\Lambda \rightarrow -\infty$. It is an approximate, IR effective description of JT gravity, valid on distance scales that are much larger than the curvature length scale L . In particular, the usual smooth boundary length ℓ is an effective, long-distance parameter that can be expressed in terms of the microscopic parameters in the model, β_q , Λ , and L , as in Eq. (6.10). The Schwarzian theory, therefore, plays a role, with respect to JT, that is similar to the one it plays with respect to the SYK model. We believe that the Schwarzian theory is a universal large-distance effective theory for many different random loop models in hyperbolic space, including models that do not have a metric or quantum gravity interpretation. Its emergence and universality is tightly related to the asymptotic ballistic property of random paths models in hyperbolic space.

B. Comments on general topologies

In the framework of the Schwarzian description, the construction of the negative curvature JT gravity on a two-dimensional manifold of arbitrary topology was presented in [27]. The results are particularly simple and elegant. They rely on two basic assumptions.

The first assumption is to consider that the contributing geometries in genus h with b boundaries can be built as follows. One considers a “core” geometry of genus h with b geodesic boundaries of lengths $l_1, \dots, l_b$. These are precisely the geometries that are considered in topological gravity. They yield the moduli space volumes $V_{h,b}(l_1, \dots, l_b)$ that are discussed in [10]. On these core geometries are glued “trumpets” that correspond to annulus partition functions for which one of the boundary is a geodesic of fixed length whereas the other is a JT gravity boundary, which means that it has a fixed length but need not be a geodesic. Let us call $W_{\text{Sch}}^{\text{trumpet}}(\beta_S, l)$ the associated partition function, where l is the geodesic boundary length and β_S the Schwarzian limit parameter for the JT boundary, defined in Eq. (2.50). The full JT gravity partition function, that follows from this gluing construction, is given by [27]

$$W_{\text{Sch}}^{(h,b)}(\beta_{S,1}, \dots, \beta_{S,b}) = \int_{[0,+\infty]^b} V_{h,b}(l_1, \dots, l_b) \prod_{i=1}^b W_{\text{Sch}}^{\text{trumpet}}(\beta_{S,i}, l_i) l_i dl_i. \quad (7.1)$$

The case of the annulus, $h = 0$, $b = 2$, is special, because there is no core manifold with this topology and geodesic boundaries. The partition function is then given by the gluing of two trumpets:

$$W_{\text{Sch}}^{(0,2)}(\beta_{S,1}, \beta_{S,2}) = \int_0^\infty W_{\text{Sch}}^{\text{trumpet}}(\beta_{S,1}, l) W_{\text{Sch}}^{\text{trumpet}}(\beta_{S,2}, l) l dl. \quad (7.2)$$

The measure of integration ldl comes from the usual Weil-Petersson measure, the factor of l being associated with the choice of twisting one has to make when one glues the trumpet on the core surface.

The second assumption is that the trumpet partition function, in the Schwarzian limit, is obtained from an argument following closely the derivation of the disk partition function [Eq. (2.67)] in the same limit. It is then found to be one-loop exact and given by

$$W_{\text{Sch}}^{\text{trumpet}}(\beta_S, l) = \beta_S^{-1/2} e^{-\frac{l^2}{16\pi\beta_S}}. \quad (7.3)$$

Using the equivalence between the Mirzakhani recursion relations [10] and the Eynard topological recursion [12], Eqs. (7.1) and (7.2) imply that the JT gravity partition functions in the Schwarzian limit are computed by a standard Hermitian matrix model. The explicit spectral curve for the matrix model is derived straightforwardly from Eq. (7.3).

In the microscopic formulation of JT gravity, the partition function in genus h and b boundaries is a function $W_{h,b}(\beta_{q,1}, \dots, \beta_{q,b})$ of the b quantum boundary lengths $\beta_{q,i}$. In the Schwarzian limit $\Lambda \rightarrow -\infty$, this microscopic partition function should match with $W_{\text{Sch}}^{(h,b)}$, with an appropriate identification between the microscopic parameters $\beta_{q,i}$, Λ , and L and the Schwarzian parameters $\beta_{S,i}$. The discussion at the end of Sec. VIB 2 suggests that $\beta_{S,i}$ is simply proportional to $\beta_{q,i}/L^2$ in the pure gravity theory. Even though it remains to be derived rigorously, we strongly believe that the matching between $W_{h,b}$ and $W_{\text{Sch}}^{(h,b)}$ in the Schwarzian limit is valid, and so is the analysis presented in [27] in the same limit. The physics underlying the emergence of the Schwarzian or matrix model description is the same as for the disk topology, as explained in Sec. VI.

A fundamental open problem is the computation of $W_{h,b}(\beta_{q,1}, \dots, \beta_{q,b})$ for finite values of the microscopic parameters. Can we hope that the Saad-Shenker-Stanford framework can be suitably generalized in this case? Is it possible to maintain the relationship with a matrix model? The answer to these questions seems to be a definite no. Indeed, it is straightforward to construct many geometries contributing to $W_{h,b}$ that do not admit the simple “trumpet decomposition” that yields the formula (7.1), because they do not contain a geodesic on which the gluing can be made. Let us illustrate this on the simplest case of the annulus.

The “double-trumpet” geometries that contribute in (7.2) are characterized by the fact that the holonomy of the associated flat $\text{PSL}(2, \mathbb{R})$ connection along the nontrivial cycle of the annulus is a so-called hyperbolic element, conjugate to a matrix of the form

$$\begin{pmatrix} e^{l/2} & 0 \\ 0 & e^{-l/2} \end{pmatrix}. \quad (7.4)$$

The parameter l is the length of the geodesic along which the gluing of the trumpets is made. These double-trumpet geometries can be explicitly described by an immersion of the annulus into hyperbolic space with an appropriate identification. The simplest way to depict this immersion is to use the strip representation of $\mathbb{H}^2$, with metric, in units for which $L = 1$, given by

$$ds^2 = \frac{d\tau^2 + d\sigma^2}{\cos^2 \sigma}, \quad -\frac{\pi}{2} < \sigma < \frac{\pi}{2}, \quad (7.5)$$

and identify $\tau \equiv \tau + a$. The geodesic of length a is at $\sigma = 0$; see the upper-left inset in Fig. 25.

But there are myriads of annulus geometries that are not of this type, including geometries that can have arbitrarily large smooth boundary lengths. For instance, the annulus represented in the lower-left inset in Fig. 25 does not contain a geodesic, even though its holonomy is hyperbolic. One can also use immersions into $\mathbb{H}^2$ in the Poincaré disk representation, with the metric given by Eq. (2.51), for instance, of the type represented in Fig. 15. The holonomy of such configurations is trivial. Configurations with a more general elliptic holonomy of the form

$$\begin{pmatrix} \cos(2\pi\alpha) & \sin(2\pi\alpha) \\ -\sin(2\pi\alpha) & \cos(2\pi\alpha) \end{pmatrix} \quad (7.6)$$

are obtained if one identifies $\phi \equiv \phi + 2\pi\alpha$ in (2.51). Annulus with a parabolic holonomy can be constructed using similar ideas, considering immersions into $\mathbb{H}^2$ in the Poincaré half-plane representation, with metric $ds^2 = (dx^2 + dz^2)/z^2$ and identification $x \equiv x + \text{const}$. The existence of all these geometries clearly invalidates the decomposition (7.2) away from the Schwarzian limit. Similar constructions similarly invalidate (7.1). Let us also note that the trumpet partition function itself, or similar partition functions involving mixed boundary conditions (some boundaries being geodesics, other being fluctuating JT boundaries), involve rather singular geometries in the microscopic theory, as illustrated in the lower-right inset in Fig. 25. The JT boundaries may indeed fluctuate wildly and touch the geodesics, yielding fluctuating boundaries with several components and degenerate, infinitely thin annulus portions.

C. Future research directions

There are many possible future research directions. Let us briefly outline a few of them.

We have already mentioned the new interesting combinatorial questions in relation with the self-overlapping polygon model, its matrix model formulation, and the relevance of numerical methods. One would like to obtain an exact solution in the continuum limit and answer interesting conceptual questions about the model. For instance, is the measure of the set of configurations having

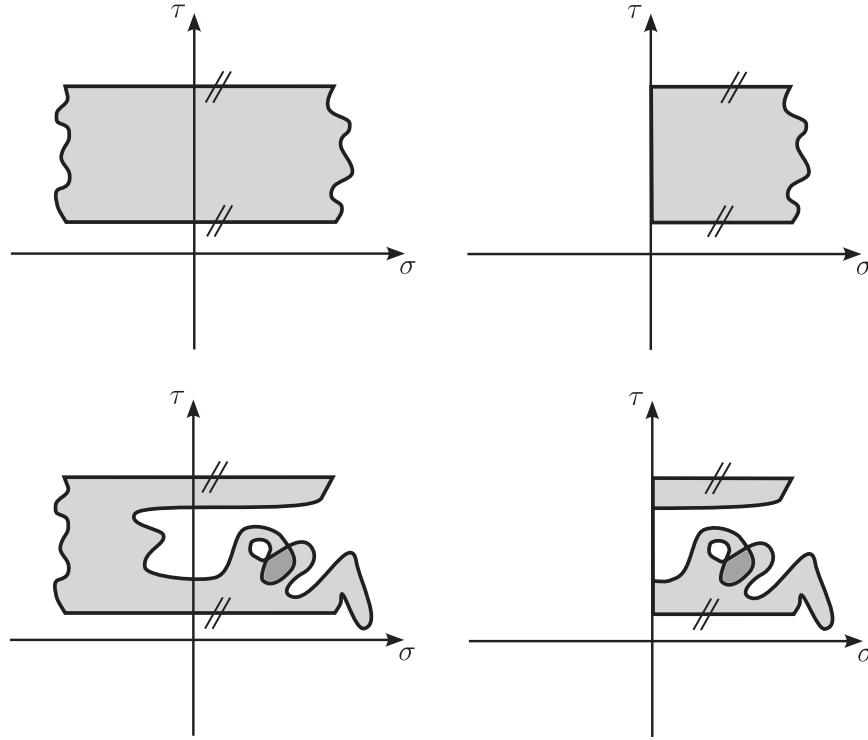


FIG. 25. Annuli of constant negative curvature depicted as immersions into H^2 in the strip representation, with metric (7.5) and identification $\tau \equiv \tau + l$. The identified sides are crossed out twice. Upper-left inset: a double-trumpet geometry, with its smooth and gently wiggling boundaries. It contains a closed geodesic at $\sigma = 0$. Upper-right inset: one of the trumpets used to build the geometry on the upper-left inset. It has a geodesic boundary on the left and a fluctuating JT boundary on the right. Lower-left inset: a geometry with a hyperbolic holonomy but a wildly fluctuating boundary on the right, that also contains an overlapping region for the sake of illustration. The geometry does not contain a closed geodesic. Lower-right inset: a geometry *a priori* contributing to the “trumpet” partition function in the microscopic theory, with a geodesic boundary and a fluctuating JT boundary. The fluctuating boundary may have several components that end on the geodesic. Note that, in the microscopic theory, the fluctuating boundaries are fractal curves, exactly as in the case of the disk topology.

a multiplicity index strictly greater than one zero, or is it strictly positive?

The study of the microscopic formulation of JT gravity on any topology is an outstanding and nontrivial problem. We have emphasized that the usual picture in terms of the trumpet decomposition and the associated matrix model formalism *à la* Saad-Shenker-Stanford does not generalize in any straightforward way. Yet, the understanding of the contributions of arbitrary topologies is important for many physical applications. On the one hand, these contributions are responsible for the so-called factorization problem, which makes it impossible to obtain a standard holographic interpretation in terms of a dual quantum theory. On the other hand, they seem to be essential for guaranteeing the full consistency of the quantum theory, in particular, in relation with the black hole information problem.

In any case, our work indicates that the real-time quantization of JT gravity needs to be revisited. The UV-complete Euclidean theory we have developed in the present paper from the discretized point of view can also be formulated directly in the continuum by using the

conformal gauge [16,28]. This approach strongly suggests that the correct degree of freedom is the boundary conformal factor and the action to start with when quantizing has to include a “Liouville-Hilbert” term that is not present in the classical JT gravity action. It also suggests that, if space-time has the topology of a strip $I \times \mathbb{R}$, for a finite-size spatial interval I , boundary time reparametrization invariance remains a good unbroken gauge symmetry. This is the counterpart of the fact that, in the microscopic Euclidean theory, reparametrizations of the boundary coordinate are perfectly good unbroken gauge symmetries. This implies that there cannot be a nontrivial boundary Hamiltonian. In this sense, the model with boundaries is not so different than a model for a closed universe, with space-time topology $S^1 \times \mathbb{R}$. Both the bulk time and the boundary time seem to be emergent concepts in this framework. The boundary time may emerge as an effective parameter in the near-AdS limit, possibly in a way similar to the emergence of the smooth boundary length parameter in the near-hyperbolic limit of the Euclidean theory.

Finally, we have explained that the positive curvature JT quantum gravity theory is a perfectly consistent model for a positive cosmological constant, even though it has been very little studied up to now. In view of its relation with cosmological models, it may turn out to be the most interesting JT model for physics, especially if the real-time quantization can be studied, potentially shedding light on quantum gravity in de Sitter space and de Sitter holography.

ACKNOWLEDGMENTS

This work has a very long and tortuous history, tightly intertwined with the sad period of the COVID pandemic. Some results, including the role played by self-overlapping curves in JT gravity, were sketched as early as 2021 during a seminar at the Institute for Advanced Study in Dublin, Ireland and at the workshop on Quantum Geometry, Field Theory and Gravity in Corfu, Greece. I am particularly grateful to the organizers of the Quantum Gravity, Random Geometry and Holography program at the Institut Henri Poincaré in Paris, France (J. Barrett, D. Benedetti, J. Ben Geloun, and R. Loll), of the workshop on Random Geometry in Math and Physics in Nijmegen, The Netherlands (T. Budd), and to the APCTP in Pohang, South Korea (J. Yoon), for providing me with an exceptional scientific and working environment, which has enabled the work presented here to mature. I also thank Soumyadeep Chaudhuri for a stimulating collaboration and useful discussions. This work was supported in part by the International Solvay Institutes and the Asia Pacific Center for Theoretical Physics (APCTP) for participating in the APCTP Focus Program, “Entanglement, Large N, and Black Hole 2024.” It was completed during my stay at the APCTP in Pohang, South Korea, and at the Aspen Center for Physics, Aspen, Colorado, USA.

APPENDIX A: CLASSICAL LIMITS OF LIOUVILLE AND JT GRAVITY ON THE DISK

The continuum generating function (also called partition function in this context) $W(\ell, \Lambda)$ for Liouville quantum gravity on the disk, discussed in Sec. IV A 2, has a formal path integral representation

$$W(\ell, \Lambda) = \frac{1}{\text{Vol}(\text{Diff}(\mathcal{D}))} \times \int Dg e^{-\frac{\Lambda}{16\pi} A[g]} Z_{\text{CFT}}[g] \delta\left(\ell - \oint ds\right). \quad (\text{A1})$$

The integral is over all the metrics on the disk, with the constraint, imposed by the δ function, that the disk boundary length is fixed and equal to ℓ . The factor $Z_{\text{CFT}}[g]$ represents the partition function of an arbitrary CFT coupled to gravity. The cosmological constant Λ couples to the disk area $A[g]$. In conformal gauge, the

metrics take the form

$$g = e^{2\sigma} \delta \quad (\text{A2})$$

for an arbitrary background disk metric δ . Standard manipulations then yield⁷¹

$$W(\ell, \Lambda) = \frac{Z_{\text{CFT}}[\delta]}{\text{Vol}(\text{PSL}(2, \mathbb{R}))} \times \int D\sigma e^{\frac{c-26}{24\pi} S_L[\delta, \sigma] - \frac{\Lambda}{16\pi} A[g]} \delta\left(\ell - \oint ds\right). \quad (\text{A3})$$

The partition function $Z_{\text{CFT}}[\delta]$ is an overall irrelevant, parameter-independent, constant. One must still formally divide by the (infinite) volume $\text{Vol}(\text{PSL}(2, \mathbb{R}))$ because there is a $\text{PSL}(2, \mathbb{R})$ subgroup of $\text{Diff}(\mathcal{D})$ that remains unbroken in conformal gauge. The measure $D\sigma$ over the Liouville field σ is a formal background-independent nonlinear integration measure. The action S_L is the Liouville action

$$S_L[\delta, \sigma] = \int_{\mathcal{D}} d^2x \sqrt{\delta} (\delta^{ab} \partial_a \sigma \partial_b \sigma + R[\delta] \sigma) + 2 \oint_{\partial \mathcal{D}} ds s_\delta k[\delta] \sigma. \quad (\text{A4})$$

The quantities $R[\delta]$, s_δ , and $k[\delta]$ refer to the Ricci scalar, arc-length coordinate, and extrinsic curvature of the boundary computed with the background metric δ , respectively. The Liouville action captures the metric dependence of the CFT partition function $Z_{\text{CFT}}[g]$ and of the Fadeev-Popov ghosts determinants that appear when the gauge (A2) is imposed.

Our goal is to discuss the semiclassical limit of the theory, defined in Eq. (4.36), to leading $|c| \rightarrow \infty$ order. To do this, the precise definition of the measure $D\sigma$ or the treatment of the unbroken $\text{PSL}(2, \mathbb{R})$ gauge symmetry are irrelevant. All we need is to solve the classical equations of motion associated with the classical action

$$S_{\text{cl}} = \frac{|c|}{24\pi} (S_L[\delta, \sigma] + \mu A[g]), \quad (\text{A5})$$

where $\mu = \frac{3}{2|c|} \Lambda$ is the rescaled cosmological constant defined in (4.36), taking into account the fact that the boundary length is fixed:

$$\oint ds = \oint ds_\delta e^\sigma = \ell. \quad (\text{A6})$$

The classical action evaluated on the classical solution, $S_{\text{cl}} = S_{\text{cl}}^*$, then yields the leading large- $|c|$ behavior of the generating function according to

⁷¹Basic references for Liouville gravity include [6–8,13].

$$\lim_{c \rightarrow -\infty} \frac{1}{|c|} \ln W(\ell, \Lambda) = -S_{cl}^*. \quad (A7)$$

Dealing with the constraint (A6) with the help of a Lagrange multiplier ζ , it is straightforward to show that the equations of motion derived from (A5) are equivalent to

$$R = -2\tilde{\Lambda}, \quad k = \zeta, \quad (A8)$$

where R and k are the Ricci scalar and extrinsic curvature of the disk boundary computed with the metric $g = e^{2\sigma}\delta$, respectively. The two equations in (A8) come from the vanishing of the bulk and boundary terms in the variation of the action (A5).

The fact that the Ricci curvature is a negative constant and the extrinsic curvature of the boundary is constant as well implies that the solution is a disk embedded in hyperbolic space. Choosing the background metric δ to be the flat Euclidean metric for a disk of area π , the conformal factor of the solution in polar coordinates is given by

$$e^\sigma = \frac{2}{\sqrt{\mu}} \frac{r_0}{1 - r_0^2 \rho^2}, \quad 0 \leq 1 \leq \rho. \quad (A9)$$

The constraint (A6) yields

$$r_0 = \frac{2\pi}{\sqrt{\mu}\ell} \left[-1 + \sqrt{1 + \frac{\mu\ell^2}{4\pi^2}} \right]. \quad (A10)$$

On this solution, the area and the Liouville action evaluate, respectively, to

$$A^* = \frac{4\pi}{\mu} \frac{r_0^2}{1 - r_0^2} = \frac{\ell r_0}{\sqrt{\mu}}, \quad S_L^* = \frac{4\pi r_0^2}{1 - r_0^2} + 4\pi \ln \frac{2r_0}{\sqrt{\mu}}. \quad (A11)$$

Introducing the variable

$$z = \frac{\ell\sqrt{\mu}}{2\pi} \quad (A12)$$

and using (A10), we get

$$\frac{r_0^2}{1 - r_0^2} = \frac{1}{2}(-1 + \sqrt{1 + z^2}), \quad \frac{\sqrt{\mu}}{2r_0} = \frac{\pi}{\ell}(1 + \sqrt{1 + z^2}). \quad (A13)$$

Combining (A5), (A7), and (A11) with the above formulas, we finally obtain

$$\lim_{c \rightarrow -\infty} \frac{1}{|c|} \ln W(\ell, \Lambda) = -\frac{1}{6} \ln \frac{\ell}{2\pi} - \frac{1}{6} (\sqrt{1 + z^2} - 1) + \frac{1}{6} \ln \frac{1 + \sqrt{1 + z^2}}{2}. \quad (A14)$$

This result matches with the predictions of Eq. (4.37), (4.38), and (4.43) in the main text.

In the case of JT gravity in negative curvature, the constraint $R = -2$ is imposed by definition of the model. The semiclassical limit corresponds to $\Lambda \rightarrow -\infty$ at fixed ℓ and was discussed in Sec. II E 4. Since $\Lambda < 0$ in this case, the classical solution, minimizing the action, corresponds to a disk of maximal area, at fixed boundary length ℓ . As is well known in the theory of extremal surfaces, this implies that the extrinsic curvature of the boundary must be constant. One, thus, gets exactly the same classical equations (A8) as in the case of Liouville gravity, with $\mu = 1$. Using the value for the on-shell area given by (A11)–(A13), we obtain the leading behavior of the partition function:

$$\lim_{\Lambda \rightarrow -\infty} \frac{1}{|\Lambda|} \ln Z(\ell, \Lambda) = \frac{A^*}{16\pi} = \frac{1}{8} \sqrt{1 + \left(\frac{\ell}{2\pi}\right)^2} - \frac{1}{8} = \frac{\ell}{16\pi} \sqrt{1 + \left(\frac{2\pi}{\ell}\right)^2} - \frac{1}{8}. \quad (A15)$$

This matches with Eqs. (2.87) and (2.88) modulo the usual counterterms, as discussed below Eq. (2.88).

APPENDIX B: ASYMPTOTIC ANALYSIS OF THE SAP GENERATING FUNCTION

The aim of this appendix is to derive the asymptotic formulas (4.95) for the SAP generating function (4.94) in the limit

$$\Lambda\beta^{3/2} \rightarrow -\infty \quad (B1)$$

of a large negative cosmological constant.

We are going to use a nice integral formula [79] for the coefficients φ_r of the asymptotic expansion of the logarithmic derivative of the Airy function [Eq. (4.90)]:

$$\varphi_r = \frac{3}{2\pi^2} (-1)^{r+1} \int_0^\infty t^{3(r-1)/2} A(t) dt \quad \text{for any } r \geq 1, \quad (B2)$$

where

$$A(t) = \frac{1}{\text{Ai}^2(t) + \text{Bi}^2(t)}. \quad (B3)$$

Let us introduce the functions ψ and $\tilde{\psi}$:

$$\psi(x) = -\sum_{r=1}^{\infty} \frac{\varphi_r}{\Gamma((3r-1)/2)} x^r = \frac{3x}{2\pi^2} \tilde{\psi}(x), \quad (\text{B4})$$

in terms of which the exact formula (4.94) for the generating function reads

$$W^{\text{SAP}}(\beta, \Lambda) = 2\sqrt{\pi}a_0 \frac{e^{-\beta\lambda}}{\beta^{5/2}} \left(\frac{1}{2\sqrt{\pi}} + \psi(x) \right), \quad \text{for } x = \frac{\Lambda\beta^{3/2}}{32\pi^{5/2}a_0}. \quad (\text{B5})$$

Note that the asymptotic behavior of φ_r can be straightforwardly obtained from (B2) by using the saddle-point method and the known large- t behavior of the Airy functions. One gets

$$\varphi_r \underset{r \rightarrow \infty}{\sim} (-1)^{r+1} \sqrt{\frac{2}{\pi}} \left(\frac{3}{4}\right)^r r^{r-\frac{1}{2}} e^{-r}. \quad (\text{B6})$$

This implies, in particular, that the series defining ψ or, equivalently, W^{SAP} have an infinite radius of convergence.

Using (B2) and permuting the infinite sum and integral signs, we can perform explicitly the infinite sum. We get in this way

$$\tilde{\psi}(x) = \int_0^{\infty} A(t) f(t^{3/2}x) dt \quad (\text{B7})$$

for

$$f(u) = -\frac{1}{3} \left(e^{|u|^{2/3}} + 2 \cos\left(\frac{\sqrt{3}}{2}|u|^{2/3}\right) e^{-\frac{1}{2}|u|^{2/3}} \right) + \frac{1}{2\sqrt{\pi}u} (1 - F(u^2/27)), \quad (\text{B8})$$

where the function F is a generalized hypergeometric function:

$$F(z) = {}_1F_3(\{1\}, \{-1/6, 1/6, 1/2\}; z). \quad (\text{B9})$$

When $u \rightarrow -\infty$, which is the relevant regime to study the limit (B1), f has an asymptotic expansion of the form

$$f(u) = -\frac{2}{3} e^{|u|^{2/3}} + \frac{1}{2\sqrt{\pi}u} + O(1/u^3). \quad (\text{B10})$$

This asymptotic form becomes valid when u is of the order of one and larger. In the integral (B7), this is the region where t is of the order of $1/|x|^{2/3}$ or larger. Introducing an arbitrary $M > 0$, we thus write (B7) as a sum of two terms:

$$\tilde{\psi}(x) = \int_0^{M/|x|^{2/3}} A(t) f(t^{3/2}x) dt + \int_{M/|x|^{2/3}}^{\infty} A(t) f(t^{3/2}x) dt. \quad (\text{B11})$$

The first term on the right-hand side of this equation can be evaluated when $|x| \rightarrow \infty$ by making the change of variable $t' = |x|^{2/3}t$ and then expanding the function A near zero. This yields contributions of the order of $1/|x|^{2/3}$, with expansion parameter $1/|x|^{2/3} \propto 1/(|\Lambda|^{2/3}\beta)$. Using (B10), the second integral can approximated as

$$\int_{M/|x|^{2/3}}^{\infty} f(t^{3/2}x) A(t) dt \simeq -\frac{2}{3} \int_{M/|x|^{2/3}}^{\infty} A(t) e^{|x|^{2/3}t} dt \quad (\text{B12})$$

up to exponentially subleading terms. The integral (B12) can be evaluated systematically by using the saddle-point method. The saddle point t_s satisfies

$$\frac{A'(t_s)}{A(t_s)} = -|x|^{2/3}. \quad (\text{B13})$$

Using the asymptotic expansion

$$A(t) \underset{t \rightarrow +\infty}{=} \pi e^{-\frac{4}{3}t^{3/2}} \sqrt{t} \left(1 - \frac{5}{24t^{3/2}} + O(1/t^3) \right) + O(e^{-\frac{8}{3}t^{3/2}}), \quad (\text{B14})$$

we find that $t_s = \frac{1}{4}|x|^{4/3}$. It is, thus, natural to set $t = \frac{1}{4}|x|^{4/3}\tau$. Up to exponentially small terms in the limit $|x| \rightarrow \infty$, the integral (B12) is of the form

$$-\frac{|x|^{4/3}}{6} \int_0^{\infty} e^{-|x|^2 \mathcal{S}(\tau)} d\tau \quad (\text{B15})$$

for

$$\begin{aligned} \mathcal{S}(\tau) &= -\frac{1}{4}|x|^2\tau - \ln A\left(\frac{1}{4}|x|^{4/3}\tau\right) \\ &= \frac{1}{6}|x|^2\left(\tau^{3/2} - \frac{3}{2}\tau\right) - \ln \frac{\pi|x|^{2/3}}{2} - \frac{1}{2}\ln \tau + O(1/|x|^2). \end{aligned} \quad (\text{B16})$$

On this form, a systematic expansion in powers of $1/|x|^2$ around the saddle point $\tau = 1$ can be obtained straightforwardly, to any desired order, by setting $\tau = 1 + \tau'/|x|$ and expanding. One finds in this way

$$\tilde{\psi}(x) \underset{x \rightarrow -\infty}{=} -\frac{\pi^{3/2}}{3} |x| e^{x^2/12} (1 + O(1/x^2)). \quad (\text{B17})$$

Using (B5), this yields (4.96).

- [1] W. T. Tutte, A census of planar triangulations, *Can. J. Math.* **14**, 21 (1962); A census of planar maps, *Can. J. Math.* **15**, 249 (1963); On the enumeration of planar maps, *Bull. Am. Math. Soc.* **74**, 64 (1968).
- [2] G. 't Hooft, A planar diagram theory for strong interactions, *Nucl. Phys.* **B72**, 461 (1974).
- [3] É. Brézin, C. Itzykson, G. Parisi, and J.-B. Zuber, Planar diagrams, *Commun. Math. Phys.* **59**, 35 (1978).
- [4] V. A. Kazakov, The appearance of matter fields from quantum fluctuations of 2D gravity, *Mod. Phys. Lett. A* **04**, 2125 (1989).
- [5] É. Brézin and V. A. Kazakov, Exactly solvable field theories of closed strings, *Phys. Lett. B* **236**, 144 (1990); M. R. Douglas and S. Shenker, Strings in less than one-dimension, *Nucl. Phys.* **B335**, 635 (1990); D. J. Gross and A. A. Migdal, Nonperturbative two-dimensional quantum gravity, *Phys. Rev. Lett.* **64**, 127 (1990).
- [6] A. M. Polyakov, Quantum geometry of bosonic strings, *Phys. Lett.* **103B**, 207 (1981).
- [7] V. G. Knizhnik, A. M. Polyakov, and A. B. Zamolodchikov, Fractal structure of 2D quantum gravity, *Mod. Phys. Lett. A* **03**, 819 (1988); F. David, Conformal field theories coupled to 2D gravity in the conformal gauge, *Mod. Phys. Lett. A* **03**, 1651 (1988); J. Distler and H. Kawai, Conformal field theory and 2D quantum gravity, *Nucl. Phys.* **B321**, 509 (1989).
- [8] H. Dorn and H. J. Otto, On correlation functions for non-critical strings with $c \leq 1$ but $d \geq 1$, *Phys. Lett. B* **291**, 39 (1992); Two and three point functions in Liouville theory, *Nucl. Phys.* **B429**, 375 (1994); A. B. Zamolodchikov and A. B. Zamolodchikov, Structure constants and conformal bootstrap in Liouville field theory, *Nucl. Phys.* **B477**, 577 (1996); A. Kupiainen, R. Rhodes, and V. Vargas, The DOZZ formula from the path integral, *J. High Energy Phys.* **05** (2018) 094.
- [9] J.-F. Marckert and A. Mokkadem, Limit of normalized random quadrangulations: The Brownian map, *Ann. Probab.* **34**, 2144 (2006); J.-F. Le Gall, The topological structure of scaling limits of large planar maps, *Inventiones Mathematicae* **169**, 621 (2007); B. Duplantier and S. Sheffield, Duality and KPZ in Liouville quantum gravity, *Phys. Rev. Lett.* **102**, 150603 (2009); Schramm Loewner evolution and Liouville quantum gravity, *Phys. Rev. Lett.* **107**, 131305 (2011); B. Duplantier, J. Miller, and S. Sheffield, Liouville quantum gravity as a mating of trees, [arXiv:1409.7055](https://arxiv.org/abs/1409.7055); F. David, A. Kupiainen, R. Rhodes, and V. Vargas, Liouville quantum gravity on the Riemann sphere, *Commun. Math. Phys.* **342**, 869 (2016); Y. Huang, R. Rhodes, and V. Vargas, Liouville quantum gravity on the unit disk, *Ann. Inst. Henri Poincaré, Probab. Stat.* **54**, 1694 (2018); R. Rhodes and V. Vargas, Lecture notes on Gaussian multiplicative chaos and Liouville quantum gravity, [arXiv:1602.07323](https://arxiv.org/abs/1602.07323); A. Kupiainen, R. Rhodes, and V. Vargas, Integrability of Liouville theory: Proof of the DOZZ formula, *Ann. Math.* **191**, 81 (2020); The DOZZ formula from the path integral, *J. High Energy Phys.* **05** (2018) 094; J. Miller and S. Sheffield, An axiomatic characterization of the Brownian map, *J. l'Ecole Polytech. Math.* **8**, 609 (2021); Liouville quantum gravity and the Brownian map I: The QLE(8/3,0) metric, *Inventiones Mathematicae* **219**, 75 (2020); Liouville quantum gravity and the Brownian map II: Geodesics and continuity of the embedding, *Ann. Probab.* **49**, 2732 (2021); Liouville quantum gravity and the Brownian map III: The conformal structure is determined, *Probab. Theory Relat. Fields* **179**, 1183 (2021).
- [10] M. Mirzakhani, Simple geodesics and Weil-Petersson volumes of moduli spaces of bordered Riemann surfaces, *Inventiones Mathematicae* **167**, 179 (2007); Weil-Petersson volumes and intersection theory on the moduli space of curves, *J. Am. Math. Soc.* **20**, 1 (2007).
- [11] E. Witten, On the structure of the topological phase of two-dimensional gravity, *Nucl. Phys.* **B340**, 281 (1990); Two-dimensional gravity and intersection theory on moduli space, *Surveys Diff. Geom.* **1**, 243 (1991); M. Kontsevich, Intersection theory on the moduli space of curves and the matrix Airy function, *Commun. Math. Phys.* **147**, 1 (1992).
- [12] B. Eynard, Topological expansion for the 1-Hermitian matrix model correlation functions, *J. High Energy Phys.* **11** (2004) 031; B. Eynard and N. Orantin, Invariants of algebraic curves and topological expansion, *Commun. Num. Theor. Phys.* **1**, 347 (2007); Weil-Petersson volume of moduli spaces, Mirzakhani's recursion and matrix models, [arXiv:0705.3600](https://arxiv.org/abs/0705.3600).
- [13] P. H. Ginsparg and G. W. Moore, Lectures on 2-D gravity and 2-D string theory, [arXiv:hep-th/9304011](https://arxiv.org/abs/hep-th/9304011); P. Di Francesco, P. H. Ginsparg, and J. Zinn-Justin, 2-D Gravity and random matrices, *Phys. Rep.* **254**, 1 (1995); J. Teschner, Liouville theory revisited, *Classical Quantum Gravity* **18**, R153 (2001); J. Ambjorn, B. Durhuus, and T. Jonsson, *Quantum Geometry: A Statistical Field Theory Approach*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2005); B. Eynard, *Counting Surfaces*, Progress in Mathematical Physics, Vol. 70 (Birkhäuser, Basel, 2016); R. Dijkgraaf and E. Witten, Developments in topological gravity, *Int. J. Mod. Phys. A* **33**, 30 (2018).
- [14] C. Teitelboim, Gravitation and Hamiltonian structure in two space-time dimensions, *Phys. Lett. B* **126**, 41 (1983); R. Jackiw, Lower-dimensional gravity, *Nucl. Phys.* **B252**, 343 (1985).
- [15] V. Fateev, A. B. Zamolodchikov, and A. B. Zamolodchikov, Boundary Liouville field theory. 1. Boundary state and boundary two point function, [arXiv:hep-th/0001012](https://arxiv.org/abs/hep-th/0001012); J. Teschner, Remarks on Liouville theory with boundary, *Proc. Sci. tmr2000* (2000) 041 [[arXiv:hep-th/0009138](https://arxiv.org/abs/hep-th/0009138)].
- [16] F. Ferrari, Jackiw-Teitelboim gravity, random disks of constant curvature, self-overlapping curves and Liouville CFT₁, *Phys. Rev. D* **111**, L061901 (2025).
- [17] D. Stanford and E. Witten, Fermionic localization of the Schwarzian theory, *J. High Energy Phys.* **10** (2017) 008.
- [18] A. Almheiri and J. Polchinski, Models of AdS₂ backreaction and holography, *J. High Energy Phys.* **11** (2015) 014.
- [19] K. Jensen, Chaos in AdS₂ holography, *Phys. Rev. Lett.* **117**, 111601 (2016).
- [20] J. Maldacena, D. Stanford, and Z. Yang, Conformal symmetry and its breaking in two dimensional nearly anti-de-Sitter space, *Prog. Theor. Exp. Phys.* **12**, 12C104 (2016).

- [21] J. Engelsöy, T. G. Mertens, and H. Verlinde, An investigation of AdS_2 backreaction and holography, *J. High Energy Phys.* **07** (2016) 139.
- [22] A. Kitaev, A simple model of quantum holography, KITP Program Entanglement in Strongly-Correlated Quantum Matter, unpublished, see <http://online.kitp.ucsb.edu/online/entangled15/>; A. Kitaev and S. J. Suh, The soft mode in the Sachdev-Ye-Kitaev model and its gravity dual, *J. High Energy Phys.* **05** (2018) 183; S. Sachdev, Holographic metals and the fractionalized Fermi liquid, *Phys. Rev. Lett.* **105**, 151602 (2010); S. Sachdev and J. Ye, Gapless spin-fluid ground state in a random quantum Heisenberg magnet, *Phys. Rev. Lett.* **70**, 3339 (1993).
- [23] R. Gurau, The $1/N$ expansion of colored tensor models, *Ann. Henri Poincaré* **12**, 829 (2011); E. Witten, An SYK-like model without disorder, *J. Phys. A* **52**, 474002 (2019); I. R. Klebanov and G. Tarnopolsky, Uncolored random tensors, melon diagrams, and the Sachdev-Ye-Kitaev models, *Phys. Rev. D* **95**, 046004 (2017); F. Ferrari, The large D limit of planar diagrams, *Ann. Inst. Henri Poincaré D* **6**, 427 (2019); F. Ferrari, V. Rivasseau, and G. Valette, A new large N expansion for general matrix-tensor models, *Commun. Math. Phys.* **370**, 403 (2019); S. Carrozza, F. Ferrari, A. Tanasa, and G. Valette, On the large D expansion of Hermitian multi-matrix models, *J. Math. Phys. (N.Y.)* **61**, 073501 (2020).
- [24] T. Azeyanagi, F. Ferrari, and F. I. Schaposnik Massolo, Phase diagram of planar matrix quantum mechanics, tensor, and Sachdev-Ye-Kitaev models, *Phys. Rev. Lett.* **120**, 061602 (2018); F. Ferrari and F. I. Schaposnik Massolo, Phases of melonic quantum mechanics, *Phys. Rev. D* **100**, 026007 (2019).
- [25] J. Maldacena and X. L. Qi, Eternal traversable wormhole, [arXiv:1804.00491](https://arxiv.org/abs/1804.00491).
- [26] G. Penington, Entanglement wedge reconstruction and the information paradox, *J. High Energy Phys.* **09** (2020) 002; A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole, *J. High Energy Phys.* **12** (2019) 063; A. Almheiri, R. Mahajan, J. Maldacena, and Y. Zhao, The page curve of Hawking radiation from semiclassical geometry, *J. High Energy Phys.* **03** (2020) 149; A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, Replica wormholes and the entropy of Hawking radiation, *J. High Energy Phys.* **05** (2020) 013; G. Penington, S. H. Shenker, D. Stanford, and Z. Yang, Replica wormholes and the black hole interior, *J. High Energy Phys.* **03** (2022) 205; A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, The entropy of Hawking radiation, *Rev. Mod. Phys.* **93**, 035002 (2021).
- [27] P. Saad, S. H. Shenker, and D. Stanford, JT gravity as a matrix integral, [arXiv:1903.11115](https://arxiv.org/abs/1903.11115).
- [28] F. Ferrari, Random disks of constant curvature: The conformal gauge story (to be published).
- [29] S. Chaudhuri and F. Ferrari, Finite cut-off JT and Liouville quantum gravities on the disk at one loop, *J. High Energy Phys.* **01** (2025) 160; Dirichlet scalar determinants on two-dimensional constant curvature disks, *Ann. Henri Poincaré* (2025).
- [30] M. Henneaux, Quantum gravity in two dimensions: Exact solution of the Jackiw model, *Phys. Rev. Lett.* **54**, 959 (1985); A. Strominger, AdS_2 quantum gravity and string theory, *J. High Energy Phys.* **01** (1999) 007; J. M. Maldacena, J. Michelson, and A. Strominger, Anti-de Sitter fragmentation, *J. High Energy Phys.* **02** (1999) 011; A. Sen, State operator correspondence and entanglement in $\text{AdS}_2/\text{CFT}_1$, *Entropy* **13**, 1305 (2011).
- [31] D. Harlow and D. Jafferis, The factorization problem in Jackiw-Teitelboim gravity, *J. High Energy Phys.* **02** (2020) 177.
- [32] A. Kitaev and S. J. Suh, Statistical mechanics of a two-dimensional black hole, *J. High Energy Phys.* **05** (2019) 198.
- [33] S. J. Suh, Dynamics of black holes in Jackiw-Teitelboim gravity, *J. High Energy Phys.* **03** (2020) 093; Probabilistic deconstruction of a theory of gravity, Part I: Flat space, *SciPost Phys.* **15**, 174 (2023); Probabilistic deconstruction of a theory of gravity, Part II: Curved space, *SciPost Phys.* **16**, 012 (2024); G. Penington and E. Witten, Algebras and states in JT gravity, [arXiv:2301.07257](https://arxiv.org/abs/2301.07257).
- [34] D. Stanford and Z. Yang, Finite-cutoff JT gravity and self-avoiding loops, [arXiv:2004.08005](https://arxiv.org/abs/2004.08005).
- [35] L. V. Iliesiu, J. Kruthoff, G. J. Turiaci, and H. Verlinde, JT gravity at finite cutoff, *SciPost Phys.* **9**, 023 (2020).
- [36] L. McGough, M. Mezei, and H. Verlinde, Moving the CFT into the bulk with $T\bar{T}$, *J. High Energy Phys.* **04** (2018) 010; D. J. Gross, J. Kruthoff, A. Rolph, and E. Shaghoulian, $T\bar{T}$ in AdS_2 and quantum mechanics, *Phys. Rev. D* **101**, 026011 (2020); Hamiltonian deformations in quantum mechanics, $T\bar{T}$, and the SYK model, *Phys. Rev. D* **102**, 046019 (2020).
- [37] A. J. Guttmann and I. G. Enting, The size and number of rings on the square lattice, *J. Phys. A* **21**, L165 (1988); A. J. Guttmann and A. R. Conway, Square lattice self-avoiding walks and polygons, *Ann. Comb.* **5**, 319 (2001).
- [38] D. Bagrets, A. Altland, and A. Kamenev, Sachdev-Ye-Kitaev model as Liouville quantum mechanics, *Nucl. Phys. B* **911**, 191 (2016); T. G. Mertens, G. J. Turiaci, and H. L. Verlinde, Solving the Schwarzian via the conformal bootstrap, *J. High Energy Phys.* **08** (2017) 136; V. V. Belokurov and E. T. Shavgulidze, Schwarzian functional integrals calculus, *J. Phys. A* **53**, 485201 (2020).
- [39] D. Stanford and E. Witten, JT gravity and the ensembles of random matrix theory, *Adv. Theor. Math. Phys.* **24**, 1475 (2020).
- [40] F. Ferrari, Notes on the Kitaev-Suh model for JT gravity in negative, zero and positive curvature (to be published).
- [41] F. Ferrari, Gauge theory formulation of hyperbolic gravity, *J. High Energy Phys.* **03** (2021) 046.
- [42] O. Coussaert, M. Henneaux, and P. van Driel, The asymptotic dynamics of three-dimensional Einstein gravity with a negative cosmological constant, *Classical Quantum Gravity* **12**, 2961 (1995); M. Rooman and P. Spindel, Holonomies, anomalies and the Fefferman-Graham ambiguity in $\text{AdS}(3)$ gravity, *Nucl. Phys. B* **594**, 329 (2001); S. Carlip, Dynamics of asymptotic diffeomorphisms in $(2+1)$ -dimensional gravity, *Classical Quantum Gravity* **22**, 3055 (2005).
- [43] E. Joung, P. Narayan, and J. Yoon, Gravitational edge mode in asymptotically AdS_2 : JT gravity revisited, *J. High Energy Phys.* **05** (2024) 244.

- [44] L. Griguolo, R. Panerai, J. Papalini, and D. Seminara, Nonperturbative effects and resurgence in Jackiw-Teitelboim gravity at finite cutoff, *Phys. Rev. D* **105**, 046015 (2022).
- [45] J. L. Cardy and A. J. Guttmann, Universal amplitude combinations for self-avoiding walks, polygons and trails, *J. Phys. A* **26**, 2485 (1993); P. Mörters and Y. Peres, *Brownian Motion* (Cambridge University Press, Cambridge, England, 2010); J. Cardy, SLE for theoretical physicists, *Ann. Phys. (Amsterdam)* **318**, 81 (2005); G. F. Lawler, *Schramm-Loewner Evolution*, Lecture notes from a course given at Park City/IAS Institute, Park City, Utah, 2007; R. Bauerschmidt, H. Duminil-Copin, J. Goodman, and G. Slade, *Lectures on Self-Avoiding Walks*, Clay Mathematics Proceedings Vol. 15 (2012).
- [46] A. J. Guttmann, ed., *Polygons, Polyominoes and Polycubes*, Lecture Notes in Physics (Springer, New York, 2009).
- [47] C. Titus, The combinatorial topology of analytic functions of the boundary of a disk, *Acta Math.* **106**, 45 (1961); S. J. Blank, Extending immersions of the circle, Ph.D. dissertation, Brandeis University, 1967.
- [48] V. Poénaru, Extension des immersions en codimension 1, *Astérisque* **10**, 473 (1966).
- [49] P. W. Shor and C. J. Van Wyk, Detecting and decomposing self-overlapping curves, *Comput. Geom.* **2**, 31 (1992).
- [50] U. Mukherjee, Self-overlapping curves: Analysis and applications, *Comput.-Aided Des. Appl.* **46**, 227 (2014).
- [51] G. Brinkmann, O. Delgado-Friedrichs, and U. von Nathusius, *Number of Faces and Boundary Encoding of Patches*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science Vol. 69 (2005).
- [52] J. E. Graver and G. T. Cargo, When does a curve bound a distorted disk?, *SIAM J. Discrete Math.* **25**, 280 (2011); P. Evans and C. Wenk, Combinatorial properties of self-overlapping curves and interior boundaries, *Discrete Comput. Geom.* **69**, 91 (2023).
- [53] P. Lévy, Calcul des probabilité—sur l'aire comprise entre un arc de la courbe du mouvement Brownien plan et sa corde, *C. R. Acad. Sci., Paris* **230**, 430 (1950); Wiener's random function, and other Laplacian random functions, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (1951), p. 171.
- [54] M. Marx, The branch point structure of extensions of interior boundaries, *Trans. Am. Math. Soc.* **131**, 79 (1968).
- [55] X. Guo, P. Hansen, and M. Zheng, Boundary uniqueness of fusenes, *Discrete Appl. Math.* **118**, 209 (2002).
- [56] D. Eppstein and E. Mumford, Self-overlapping curves revisited, in *Proceedings of the 20th ACM-SIAM Symposium on Discrete Algorithms* (New York, 2009), p. 160, [10.1137/1.9781611973068.19](https://doi.org/10.1137/1.9781611973068.19).
- [57] R. C. Brower, C. V. Cofburn, A. L. Fitzpatrick, D. Howarth, and C. I. Tan, Lattice setup for quantum field theory in AdS_2 , *Phys. Rev. D* **103**, 094507 (2021); P. M. Lenggenhager, A. Stegmaier, L. K. Upreti, T. Hofmann, T. Helbig, A. Vollhardt, M. Greiter, C. H. Lee, S. Imhof, H. Brand *et al.*, Simulating hyperbolic space on a circuit board, *Nat. Commun.* **13**, 4373 (2022).
- [58] E. Swierczak and A. J. Guttmann, Self-avoiding walks and polygons on non-Euclidean lattices, *J. Phys. A* **29**, 7485 (1996).
- [59] N. Madras and C. C. Wu, Self-avoiding walks on hyperbolic graphs, *Comb. Probab. Comput.* **14**, 523 (2005); C. Panagiotis, Self-avoiding walks and polygons on hyperbolic graphs, [arXiv:1908.00127](https://arxiv.org/abs/1908.00127); I. Benjamini and C. Panagiotis, Hyperbolic self avoiding walk, *Electron. Commun. Probab.* **26**, 1 (2021).
- [60] G. Moore, N. Seiberg, and M. Staudacher, From loops to states in two-dimensional quantum gravity, *Nucl. Phys.* **B362**, 665 (1991); T. Banks, M. R. Douglas, N. Seiberg, and S. H. Shenker, *Microscopic and Macroscopic Loops in Nonperturbative Two-Dimensional Gravity*, Quantum Mechanics of Fundamental Systems 3, edited by Claudio Teitelboim and Jorge Zanelli (Plenum Press, New York, 1992); I. K. Kostov, B. Ponsot, and D. Serban, Boundary Liouville theory and 2D quantum gravity, *Nucl. Phys.* **B683**, 309 (2004); J. M. Maldacena, G. W. Moore, N. Seiberg, and D. Shih, Exact vs semiclassical target space of the minimal string, *J. High Energy Phys.* **10** (2004) 020; D. Kutasov, K. Okuyama, J. w. Park, N. Seiberg, and D. Shih, Annulus amplitudes and ZZ branes in minimal string theory, *J. High Energy Phys.* **08** (2004) 026.
- [61] D. C. Khandekar and F. W. Wiegel, Distribution of the area enclosed by a plane random walk, *J. Phys. A* **21**, L563 (1988); B. Duplantier, Areas of planar Brownian curves, *J. Phys. A* **22**, 3033 (1989); A. Comtet, J. Desbois, and S. Ouvry, Winding of planar Brownian curves, *J. Phys. A* **23**, 3563 (1990); J. Rudnick and G. Gaspari, The shapes and sizes of closed pressurized random walks, *Science* **252**, 422 (1991); G. Gaspari, J. Rudnick, and A. Beldjenna, The shapes and sizes of two-dimensional pressurized, self-intersecting rings, as models for two-dimensional vesicles, *J. Phys. A* **26**, 1 (1993); E. Haleva and H. Diamant, Smoothing transition of a two-dimensional pressurized polymer ring, *Eur. Phys. J. E* **19**, 461 (2006); M. K. Mitra, G. I. Menon, and R. Rajesh, Phase transitions in pressurised semiflexible polymer rings, *Phys. Rev. E* **77**, 041802 (2008).
- [62] É. Brézin, V. A. Kazakov, and A. B. Zamolodchikov, Scaling violation in a field theory of closed strings in one physical dimension, *Nucl. Phys.* **B338**, 673 (1990); P. H. Ginsparg and J. Zinn-Justin, 2D gravity + 1D matter, *Phys. Lett. B* **240**, 333 (1990); D. J. Gross and I. R. Klebanov, One-dimensional string theory on a circle, *Nucl. Phys.* **B344**, 475 (1990); D. J. Gross and N. Miljkovic, A nonperturbative solution of $D = 1$ string theory, *Phys. Lett. B* **238**, 217 (1990); J. Polchinski, Critical behavior of random surfaces in one-dimension, *Nucl. Phys.* **B346**, 253 (1990); S. S. Gubser and I. R. Klebanov, A modified $c = 1$ matrix model with new critical behavior, *Phys. Lett. B* **340**, 35 (1994).
- [63] V. A. Kazakov, Ising model on a dynamical planar random lattice: Exact solution, *Phys. Lett. A* **119**, 140 (1986); D. V. Boulatov and V. A. Kazakov, The ising model on random planar lattice: The structure of phase transition and the exact critical exponents, *Phys. Lett. B* **186**, 379 (1987); É. Brézin, M. R. Douglas, V. A. Kazakov, and S. H. Shenker, The ising model coupled to 2-D gravity: A nonperturbative analysis, *Phys. Lett. B* **237**, 43 (1990).
- [64] A. B. Zamolodchikov, On the entropy of random surfaces, *Phys. Lett.* **117B**, 87 (1982).

- [65] H. Duminil-Copin and S. Smirnov, The connective constant of the honeycomb lattice equals $(2 + \sqrt{2})^{1/2}$, *Ann. Math.* **175**, 1653 (2012).
- [66] S. Leibler, R. R. P. Singh, and M. E. Fisher, Thermodynamic behavior of two-dimensional vesicles, *Phys. Rev. Lett.* **59**, 1989 (1987).
- [67] K. Schmüdgen, Ten lectures on the moment problem, [arXiv:2008.12698](#).
- [68] C. Garban and J. A. Trujillo Ferreras, The expected area of the filled planar Brownian loop is $\pi/5$, *Commun. Math. Phys.* **264**, 797 (2006).
- [69] C. Richard, A. J. Guttmann, and I. Jensen, Scaling function and universal amplitude combinations for self-avoiding polygons, *J. Phys. A* **34**, L495 (2001).
- [70] A. J. Guttmann and I. G. Enting, The number of convex polygons on the square and honeycomb lattices, *J. Phys. A* **21**, L467 (1988); K. Y. Lin, Exact solution of the convex polygon perimeter and area generating function, *J. Phys. A* **24**, 2411 (1991).
- [71] T. Budd, Discrete flat disks: rigid quadrangulations, [arXiv:2509.24785](#).
- [72] B. B. Mandelbrot, *The Fractal Geometry of Nature* (W.H. Freeman and Company Ed., San Francisco, 1983).
- [73] G. Lawler, O. Schramm, and W. Werner, Values of Brownian intersection exponents I: Half-plane exponents, *Acta Math.* **187**, 237 (2001); Values of Brownian intersection exponents II: Plane exponents, *Acta Math.* **187**, 275 (2001); G. Lawler, O. Schramm, and W. Werner, Analyticity of intersection exponents for planar Brownian motion, *Acta Math.* **189**, 179 (2002).
- [74] P. Di Francesco and C. Itzykson, A generating function for fatgraphs, *Ann. Inst. Henri Poincaré, A* **59**, 117 (1993).
- [75] V. A. Kazakov, M. Staudacher, and T. Wynter, Character expansion methods for matrix models of dually weighted graphs, *Commun. Math. Phys.* **177**, 451 (1996); Almost flat planar diagrams, *Commun. Math. Phys.* **179**, 235 (1996); Exact solution of discrete two-dimensional R^2 gravity, *Nucl. Phys. B* **471**, 309 (1996); Advances in large N group theory and the solution of two-dimensional R^2 gravity, [arXiv:hep-th/9601153](#); A. Guionnet and M. Maida, Character expansion method for the first order asymptotics of a matrix integral, *Probab. Theory Relat. Fields* **132**, 539 (2005).
- [76] E. Duchi, V. Guerrini, S. Rinaldi, and G. Schaeffer, Fighting fish, *J. Phys. A* **50**, 024002 (2017).
- [77] B. Zonneveld, A tree bijection for rigid quadrangulations, [arXiv:2509.24787](#).
- [78] M. E. Gertsenshtein and V. B. Vasiliev, Waveguides with random inhomogeneities and Brownian motion in the Lobachevsky plane, *Theory Probab. Appl.* **3**, 391 (1959); L. Lao and E. Orsingher, Hyperbolic and fractional Brownian motion, *Stochastics: Int. J. Probab. Stochastics Processes* **79**, 505 (2007).
- [79] M. J. Kearney and R. J. Martin, Airy asymptotics: The logarithmic derivative and its reciprocal, *J. Phys. A* **42**, 425201 (2009).