

# No Ensemble Averaging Below the Black Hole Threshold

---

Jean-Marc Schlenker<sup>1</sup> and Edward Witten<sup>2</sup>

<sup>1</sup> *Department of Mathematics, University of Luxembourg*

<sup>2</sup> *School of Natural Sciences, Institute for Advanced Study,  
1 Einstein Drive, Princeton, NJ 08540 USA*

ABSTRACT: In the AdS/CFT correspondence, amplitudes associated to connected bulk manifolds with disconnected boundaries have presented a longstanding mystery. A possible interpretation is that they reflect the effects of averaging over an ensemble of boundary theories. But in examples in dimension  $D \geq 3$ , an appropriate ensemble of boundary theories does not exist. Here we sharpen the puzzle by identifying a class of “fixed energy” or “sub-threshold” observables that we claim do *not* show effects of ensemble averaging. These are amplitudes that involve states that are above the ground state by only a fixed amount in the large  $N$  limit, and in particular are far from being black hole states. To support our claim, we explore the example of  $D = 3$ , and show that connected solutions of Einstein’s equations with disconnected boundary never contribute to these observables. To demonstrate this requires some novel results about the renormalized volume of a hyperbolic three-manifold, which we prove using modern methods in hyperbolic geometry. Why then do any observables show apparent ensemble averaging? We propose that this reflects the chaotic nature of black hole physics and the fact that the Hilbert space describing a black hole does not have a large  $N$  limit.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Volumes and Sub-Threshold Amplitudes</b>	<b>5</b>
2.1	Preliminaries	5
2.2	Review Of The Renormalized Volume	8
2.3	Implications of The Conjecture For Hyperbolic Geometry	13
<b>3</b>	<b>Overview of Mathematical Arguments</b>	<b>17</b>
<b>4</b>	<b>Why Do Some Observables Show Ensemble Averaging?</b>	<b>20</b>
<b>A</b>	<b>Mathematical details</b>	<b>25</b>
A.1	Convex co-compact hyperbolic manifolds	26
A.2	The complex structure at infinity	27
A.3	Measured laminations on surfaces	28
A.4	The convex core and the geometry of its boundary	29
A.5	Relations between the boundary of the convex core and the boundary at infinity	30
A.6	The maximal set of contractible curves in the boundary	31
A.7	A variational formula for the renormalized volume	33
A.8	The renormalized volume and the volume of the convex core	34
A.9	A bound on the Weil-Petersson gradient	35
A.10	Convergence of convex cores when pinching compressible curves	36
A.11	Limit of the renormalized volume when pinching a compressible curve	38
A.12	Geodesically convex subsets of a hyperbolic manifold	38
A.13	The Thurston metric at infinity	40
A.14	The renormalized volume associated to the Thurston metric	40
A.15	Asymptotic behavior of the renormalized volume when pinching a compressible curve	45

## 1 Introduction

Since early days of the AdS/CFT correspondence [1], there has been a puzzle of how to interpret Euclidean amplitudes computed on a connected bulk manifold  $X$  whose conformal boundary  $M$  is not connected [2, 3]. If  $M$  is connected, the sum over all choices of  $X$  is interpreted in the AdS/CFT correspondence as computing what we will call  $Z(M)$ , the conformal field theory (CFT) partition function on  $M$ . What if  $M$  has, say, two connected components  $M_1, M_2$ ? The sum over all choices of  $X$  whose conformal boundary is the disjoint union  $M = M_1 \sqcup M_2$  appears to compute, in some sense, a connected correlation function  $\langle Z(M_1)Z(M_2) \rangle_c = \langle Z(M_1)Z(M_2) \rangle - \langle Z(M_1) \rangle \langle Z(M_2) \rangle$  between the two CFT partition functions  $Z(M_1)$  and  $Z(M_2)$ . We will refer to such connected correlation functions

between different boundaries (possibly with operator insertions on the boundaries) as “connected amplitudes with disconnected boundaries” or CADB amplitudes for short. CADB amplitudes are not a standard concept in CFT or indeed in any Euclidean quantum field theory, so the fact that AdS/CFT duality seems to give a way to calculate them has been puzzling.

A new perspective came from the discovery that a simple, soluble model, namely JT gravity in two dimensions, computes an ensemble average in a random matrix theory [4]. A two-dimensional gravitational theory is expected to have a dual description by an ordinary quantum theory on the 1-dimensional boundary of a 2-dimensional world. A compact connected 1-manifold is a circle  $S$ , say of circumference  $\beta$ , and the partition function is then  $Z(\beta) = \text{Tr} \exp(-\beta H)$ , where  $H$  is the Hamiltonian of the boundary theory. However, it turns out that the theory dual to JT gravity does not have a unique Hamiltonian  $H$ ; rather,  $H$  is drawn from a random matrix ensemble. This provides a rationale for the existence of connected correlation functions between observables associated to different boundary circles. Such amplitudes can be generated in the boundary description by averaging over the random matrix  $H$ . This discovery revived interest in much older ideas about gravitational wormholes and ensemble averages [5–7].

But the interpretation of CADB amplitudes in terms of ensemble averages raises an immediate paradox. In many basic examples of AdS/CFT duality, it is believed that the parameters on which the boundary theory depends are all known and all have a known interpretation in terms of the bulk theory. The duality seems to say that a specific boundary theory, with specific values of the parameters, is dual to a specific bulk theory with the same parameters. For instance, two of the original examples of AdS/CFT duality are the maximally supersymmetric models based on  $\text{AdS}_4 \times \text{S}^7$  and  $\text{AdS}_7 \times \text{S}^4$ . In these examples, the only parameter that the boundary CFT depends on is a single positive integer  $N$ . The bulk theory also depends on  $N$ ; in fact, Newton’s constant  $G$  varies as a negative power of  $N$ , and  $N$  can be measured as the integral of a certain seven-form or four-form on  $\text{S}^7$  or  $\text{S}^4$ . The duality claims an equivalence between bulk and boundary theories for each choice of  $N$ ; and for given  $N$ , there appears to be nothing else one could be averaging over in an ensemble. So how can CADB amplitudes in these theories be interpreted in terms of averaging over an ensemble?

We will try to shed some light on this question by arguing that when the Anti de Sitter dimension  $D$  is at least 3, and therefore the boundary dimension  $d = D - 1$  is at least 2, certain important observables, which we will call fixed energy observables, are *not* affected by ensemble averaging; if one considers only these observables, one will see no sign of ensemble averaging. These are states that can be defined just in terms of the energy and couplings of states whose energy remains above the ground state by only a finite amount as  $N$  becomes large. In particular these observables involve states that are far below the black hole threshold, which in the AdS/CFT correspondence is the energy of the Hawking-Page transition [8] from a thermal gas in Anti de Sitter space to a black hole; in gauge theory examples, this can also be interpreted as a deconfining transition [9]. The fixed energy observables are described more precisely in section 2.1. Another formulation of our proposal involves integrability. Some important examples of AdS/CFT duality are

integrable in the large  $N$  limit, and apparently also in an asymptotic expansion around that limit; for an extensive review, see [10]. Our proposal is that precisely the observables that are accessible via integrability are not affected by ensemble averaging. This is consistent with the fact that calculations based on integrability do appear to describe a definite CFT, not an ensemble average.

In the case  $D = 3$ ,  $d = 2$ , our proposal has implications for hyperbolic geometry that are explained in section 2.3, so we can test the proposal by verifying those predictions. Thus, we consider examples of AdS/CFT duality in which the bulk spacetime is asymptotic to  $\text{AdS}_3 \times B$  for some compact manifold  $B$ . The choice of  $B$  will play no role in our discussion; various examples have been much studied, including  $\text{S}^3 \times \text{T}^4$ ,  $\text{S}^3 \times \text{K3}$ , and  $\text{S}^3 \times \text{S}^3 \times \text{S}^1$  ( $\text{S}^n$  is an  $n$ -sphere,  $\text{T}^n$  is an  $n$ -torus, and K3 is the complex surface of that name). Let  $M_1, M_2, \dots, M_n$  be a collection of Riemann surfaces. If  $X$  is a hyperbolic three-manifold whose conformal boundary is a disjoint union  $M_1 \sqcup M_2 \sqcup \dots \sqcup M_n$ , then a path integral on  $X \times B$  contributes to a connected amplitude  $\langle Z(M_1)Z(M_2) \dots Z(M_n) \rangle_c$ . For small  $G$  or equivalently large  $N$ , the contribution is proportional asymptotically to  $\exp(-V_R(X)/4\pi G\ell^2)$ , where  $V_R(X)$  is the renormalized volume of  $X$ , and  $\ell$ , which we assume much larger than  $G$ , is the AdS radius of curvature. We will deduce the statement that fixed energy amplitudes do not receive contributions from manifolds with disconnected boundary from properties of renormalized volumes. Specifically, we will show that a hyperbolic three-manifold  $X$  contributes to a fixed energy amplitude on a Riemann surface  $M$  only if the boundary of  $X$  is connected and consists only of  $M$ . An overview of the mathematical arguments is given in section 3; details appear in appendix A.

Even if  $X$  has connected boundary  $M$ , it is not necessarily true that  $X$  contributes to fixed energy amplitudes on  $M$ . To contribute to such amplitudes,  $X$  must be a “handlebody” or Schottky manifold. This means that for some embedding of  $M$  in  $\mathbb{R}^3$ ,  $X$  is topologically equivalent to the interior of  $M$ . In other words, we will find that only the simplest hyperbolic three-manifolds with boundary  $M$  contribute to fixed energy amplitudes on  $M$ .

We hope that these statements about hyperbolic three-manifolds are illustrating a general lesson that fixed energy observables are not subject to ensemble averaging, but a number of caveats are necessary. First, we consider only the case  $D = 3$ ,  $d = 2$ . Though we suspect that a similar picture holds for larger values of  $D$ , to show this will require more work with both the physics and the classical geometry. Second, our main arguments concern the case that  $X$  is a classical solution of Einstein’s equations, that is, a hyperbolic three-manifold. However, it is believed that in some cases, non-solutions must be considered.<sup>1</sup> Since little is known about what non-solutions are relevant in general, we cannot make a systematic analysis. However, we will consider the few examples that have been analyzed in the literature, namely  $\mathbb{R} \times \text{T}^2$ , studied in [12], and Seifert fibered manifolds, studied in

---

<sup>1</sup>In at least some cases, these non-solutions can be interpreted as critical points of the action at infinity (see section 4.2 of [11] for discussion), as opposed to classical solutions, which are ordinary critical points. They can also possibly be interpreted, at least in some cases, as complex critical points (critical points of the analytic continuation of the Einstein action to a holomorphic function of a complex-valued metric tensor on  $X$ ).

[13]. The known and conjectured results are consistent with the claim that fixed energy observables do not show apparent ensemble averaging. Finally, even if one restricts to classical solutions, there is no need to consider only spacetimes of the product form  $X \times B$ ; one could consider a more general ten-manifold with the same asymptotic behavior. Little is known about possible solutions of this more general form, and we are not in a position to prove that they do not contribute to fixed energy CADB amplitudes.

Assuming that fixed energy observables indeed do not have CADB contributions, this is a strong indication that in AdS/CFT duality, there is no ensemble average over unknown parameters. It would be very hard for such an average not to affect fixed energy observables. Given this, why do amplitudes that involve black hole states have CADB contributions? This question will be discussed in section 4. We propose that this phenomenon reflects the following two points: (1) black hole physics is highly chaotic [14]; (2) the Hamiltonian and Hilbert space that describe a single black hole apparently do not have a large  $N$  or small  $G$  limit. The first of these assertions is well-known in the present context and was a large part of the motivation for the work that eventually led to the interpretation of JT gravity in terms of an ensemble average [4]. The second assertion, which has not been considered in the present context, is a negative one; since the entropy of a black hole grows as a power of  $N$  (as  $N^2$  in the case of  $\mathcal{N} = 4$  super Yang-Mills in four dimensions), it is hard to see in what sense the Hilbert space describing a black hole could have a large  $N$  limit, and the literature certainly does not contain any proposal for such a limit. See [15] for more discussion. (By contrast, the thermofield double, which is dual to an entangled pair of black holes [16], does have a large  $N$  limit.) Point (1) means that the Hamiltonian  $H_N$  that describes black holes of a given energy for a given value of  $N$  can be viewed as a very large pseudorandom matrix, which will look like a random matrix in any standard calculation. Point (2) suggests that for  $N \neq N'$ , even if  $|N' - N| \ll N$ ,  $H_N$  and  $H_{N'}$  can be viewed to a good approximation as independent draws from a random matrix ensemble. If so, then quantities that depend on  $H_N$ , such as the partition function, will be smooth functions of  $N$  only to the extent that they are self-averaging in random matrix theory. (In random matrix theory, a quantity is called self-averaging if it has almost the same value for almost any draw from a given random matrix ensemble.) Quantities that are not self-averaging in random matrix theory will depend erratically on  $N$  or  $G$ . With present techniques, what we know how to calculate from the gravitational path integral are smooth functions of  $N$  or  $G$ , and given the facts just stated, above the Hawking-Page transition, we can only calculate quantities that are self-averaging in random matrix theory.<sup>2</sup> If it is possible to compute erratically varying quantities from a gravitational path integral, this will involve an unfamiliar type of path integral that is not dominated by a simple sum over critical points, not even critical points at infinity. In random matrix theory, CADB amplitudes make sense and are sometimes self-averaging; when that is the case, there can be a simple way to calculate them from the gravitational path integral. A shorthand way to summarize the proposal made here is that the ensemble averaging that is seen in gravitational path

---

<sup>2</sup>If a quantity is not self-averaging but does have a nonzero average in random matrix theory, it may be possible to compute this average. Examples are discussed in [17].

integral calculations is simply an averaging over nearby values of  $N$  to eliminate erratic fluctuations. This makes sense as a general proposal because in all known examples of AdS/CFT duality for  $D \geq 3$ ,  $G$  varies inversely with one or more integers  $N$ .

We should stress again that we unfortunately do not know for sure how to extrapolate from the specific results we will prove about hyperbolic three-manifolds to a general lesson about AdS/CFT duality. As noted previously, in some important examples of AdS/CFT duality, states whose energy above the ground state is fixed for  $N \rightarrow \infty$  are described by an integrable system.<sup>3</sup> Given this, and since integrability is the antithesis of chaos, and given also the close relation of apparent ensemble averaging to chaos, the safest conjecture in the context of the present article is that energies and couplings of fixed energy states are not affected by apparent ensemble averaging. Thus a conservative title for this article would be “No Ensemble Averaging at Fixed Energy Above the Ground State.” However, in practice, in a theory that can be well approximated by pure gravity up to the black hole threshold, the detailed results we obtain in section 3 are valid for any state that is below that threshold; therefore, especially in section 3, when discussing classical solutions of pure gravity, we use the language of sub-threshold states, rather than fixed energy states. We should note, though, that in a theory that can be approximated by pure gravity up to the black hole threshold, the sub-threshold states are all Virasoro descendants of the identity, which explains why their couplings behave similarly to those of the fixed energy states.

After v1 of this paper was submitted to the arXiv, there appeared a very interesting article [18] on couplings in 3d gravity of states whose energy, in the large  $N$  limit, is above the ground state by a fixed fraction – positive but less than 1 – of the energy required to make a black hole. (This regime might be similar to the regime mentioned in footnote 3 with  $k \sim N$ .) Such states could be solitons, for example. Couplings of such states do show apparent ensemble averaging. A possible interpretation, in the spirit of section 4, is that couplings of these states are described in the semiclassical limit by the nonlinear gravity or supergravity theory, which (if it has the assumed states) is not integrable and is likely to lead to classical and quantum chaos.

## 2 Volumes and Sub-Threshold Amplitudes

### 2.1 Preliminaries

As was remarked in the introduction, known examples of AdS/CFT duality in  $D \geq 3$  always depend on one or more integers with an inverse relation to Newton’s constant  $G$ . For example, four-dimensional maximally supersymmetric Yang-Mills theory with gauge group  $SU(N)$  has a dual description in  $AdS_5 \times S^5$  with  $G \sim 1/N^2$ . In examples with  $D = 3$ ,

---

<sup>3</sup> As pointed out to us by J. Maldacena, even when there is no integrable system, the spectrum of fixed energy states is never truly chaotic, since for large  $N$  the dimension of a multi-trace operator  $\text{Tr } \mathcal{O}_1 \text{ Tr } \mathcal{O}_2 \cdots \text{Tr } \mathcal{O}_k$  is simply the sum of the dimensions of the individual factors. This assumes that the number  $k$  of factors is kept fixed for  $N \rightarrow \infty$ . For  $k \sim N$ , one enters a different regime in which nonlinear interactions are important in the large  $N$  limit, potentially leading to classical and quantum chaos (though in general probably not the maximal chaos of black hole physics).

$d = 2$ , which will be our main focus in this article, one has<sup>4</sup>

$$G = \frac{3\ell}{2c}, \quad (2.1)$$

where  $c$  is the central charge of the boundary CFT, and  $\ell$  is the radius of curvature of the  $\text{AdS}_3$  space. For  $D = 3$ ,  $d = 2$ , the “large  $N$  limit” is a limit in which  $c$  is large, and therefore  $G \ll \ell$ ; that last condition enables gravity to be treated semiclassically, by summing over classical solutions, as we will assume in this article. For instance, in one much-studied family of models,  $c = 6Q_1Q_5$ , where  $Q_1$  and  $Q_5$  are integer-valued one-brane and five-brane charges. In that example, by “large  $N$  limit,” we mean the limit in which  $Q_1$  and  $Q_5$  are taken to be large, with a fixed ratio, ensuring, in particular, that  $c$  is large.

In a  $d$ -dimensional CFT, a local operator  $\mathcal{O}$  is inserted in a correlation function at a point  $p$  in a  $d$ -manifold  $M$ . A local operator  $\mathcal{O}$  has a dimension  $\Delta$ , which determines how it behaves under a conformal transformation that rescales the tangent space at  $p$ , and it transforms in an irreducible representation  $J$  of the group  $\text{SO}(d)$  of rotations around the point  $p$ , or (in a theory with fermions) its double cover  $\text{Spin}(d)$ .  $J$  is called the spin of the operator. In our main example of  $d = 2$ , the group  $\text{Spin}(2)$  is abelian and  $J$  can be viewed as an integer or half-integer, an element of  $\frac{1}{2}\mathbb{Z}$ . In a CFT that participates in AdS/CFT duality, the dimensions  $\Delta_i$  and representations  $J_i$  of local operators  $\mathcal{O}_i$  have a large  $N$  limit. This is a basic prediction of the duality, and in gauge theory examples in  $d = 4$  it is a consequence of the planar diagram expansion [20]. There is precisely one local operator of dimension 0, namely the identity operator 1, which transforms in a trivial 1-dimensional representation of  $\text{Spin}(d)$ . The other local operators  $\mathcal{O}_i$ ,  $i = 1, 2, \dots$  have positive dimensions and can be labeled in order of increasing dimension  $0 < \Delta_1 \leq \Delta_2 \leq \dots$ .

By the operator-state correspondence of CFT, local operators correspond to Hilbert space states if a CFT is quantized on a spatial manifold  $\mathbb{S}^{d-1}$  (with a round metric). For odd  $d$ , the identity operator corresponds to a state of energy 0, but for even  $d$  the identity operator corresponds to a state with an energy that is determined by the anomaly in a conformal mapping from  $\mathbb{R}^d$  with a point removed to  $\mathbb{R} \times \mathbb{S}^{d-1}$ . We will be mainly interested in  $d = 2$ , in which case the identity operator corresponds to a state of energy

$$E_0 = -\frac{c}{12}. \quad (2.2)$$

Importantly, this value is negative and, in the large  $N$  limit, it is large. That will lead to a prediction that certain renormalized volumes of three-manifolds should go to  $-\infty$  in appropriate limits. Any other operator  $\mathcal{O}_i$  corresponds to a state (more precisely a collection of states transforming in an irreducible representation  $J_i$  of  $\text{Spin}(d)$ ) of energy

$$E_i = E_0 + \Delta_i. \quad (2.3)$$

Thus the dimension  $\Delta_i$  of an operator is the same as the excitation energy of the corresponding state above the ground state.  $\Delta_i$  is by definition the energy of the  $i^{\text{th}}$  excited state, and in the context of AdS/CFT duality, it has a limit for  $c \rightarrow \infty$ . So in our terminology, the  $i^{\text{th}}$  excited state for each  $i$  is a fixed energy state.

---

<sup>4</sup>This formula was actually discovered before the general understanding of AdS/CFT duality [19].

$\Delta_i$  and  $J_i$  are the first basic examples of fixed energy observables that we propose are not subject to ensemble averaging. The other such observables are essentially the trilinear correlation functions of the  $\mathcal{O}_i$ . The  $\mathcal{O}_i$  can be normalized to put their two-point functions in a standard form (for spinless operators, the standard form is  $\langle \mathcal{O}_i(x)\mathcal{O}_j(y) \rangle = \delta_{ij}/|x-y|^{2\Delta_i}$ ) and then the trilinear or three-point correlation functions<sup>5</sup>

$$\lambda_{ijk} = \langle \mathcal{O}_i(x)\mathcal{O}_j(y)\mathcal{O}_k(z) \rangle \quad (2.4)$$

are important observables of a CFT. We propose that also the  $\lambda_{ijk}$  are not affected by ensemble averaging.

In  $d = 2$ , all observables of a CFT are completely determined, in principle, in terms of the  $\Delta_i$ ,  $J_i$ , and  $\lambda_{ijk}$ . This can be proved by using the fact that any two-manifold without boundary can be built by gluing together three-holed spheres, a fact that we will exploit in section 2.3. As a result, in  $d = 2$ , we expect that all observables not subject to ensemble averaging are actually determined by  $\Delta_i$ ,  $J_i$ , and  $\lambda_{ijk}$ . In  $d = 2$ , a complete set of conditions on  $\Delta_i$ ,  $J_i$ , and  $\lambda_{ijk}$  so that they are CFT data is known in principle (but often hard to use in practice). Above  $d = 2$ , none of these statements have equally simple analogs, and in particular we do not know whether to expect that the  $\Delta_i$ ,  $J_i$ , and  $\lambda_{ijk}$  are a complete set of observables that are free of ensemble averaging.

If  $\Delta_i$ ,  $J_i$ , and  $\lambda_{ijk}$  are a complete set of CFT observables in  $d = 2$ , and are not subject to ensemble averaging, then why in  $d = 2$  do any observables appear to be subject to ensemble averaging? The answer to this question depends on the fact that the observables that appear to be subject to ensemble averaging are the ones that receive contributions from black hole states. We will refer to a spacetime asymptotic to  $\text{AdS}_3$  at spatial infinity as an  $\text{AAdS}_3$  spacetime. The Einstein equations have an  $\text{AAdS}_3$  black hole solution, namely the BTZ black hole [21]. It was understood in the original paper on the BTZ black hole that if energy is defined by the usual ADM recipe of general relativity, then  $\text{AdS}_3$  itself has negative energy. Later it was understood [22] that this negative energy can be understood in terms of the central charge of the BTZ black hole. Thus  $\text{AdS}_3$  itself corresponds to the ground state of the CFT, with energy

$$E_0 = -\frac{c}{12}, \quad (2.5)$$

with  $c = 3\ell/2G$  [19]. Small perturbations of  $\text{AdS}_3$  give the states that have energy  $-c/12 + \Delta_i$ , with fixed  $\Delta_i$  and large  $c$ . We get to the black hole regime if we take  $c$  large with  $\Delta \gg c/12$ , meaning that the total energy  $E_0 + \Delta$  is large and positive. In that regime, the density of states per unit energy is exponentially large; it is  $e^{S(E)}$ , where  $S(E)$ , which is the Bekenstein-Hawking entropy of the black hole at energy  $E$ , is of order  $c$  for large  $c$  and fixed  $E/c$ .

For any given value of  $c$  or  $N$ , there is no useful notion of whether a given state is a black hole or not. The black hole region is defined only in terms of a limiting process: if we

---

<sup>5</sup>In the case of spinless conformal primary fields, these three-point functions depend only on the chosen points  $x, y, z$ . More generally, one has to pick local parameters at  $x, y, z$ . The details are not important in the present article.

go to large  $c$  with fixed  $\Delta$ , we get an ordinary state, and if we go to large  $c$  with  $E/c$  fixed and positive, we get a black hole. The distinction is only sharp in the limit of large  $c$ , but because the gravitational calculations that we know how to perform involve an asymptotic expansion at large  $c$ , the distinction is quite sharp in computations that we can actually perform.

Concretely, the Hawking-Page phase transition occurs as follows. In AdS/CFT duality, the partition function on a boundary manifold  $M$  is computed by summing over bulk manifolds with conformal boundary  $M$ . In stating the following, we will assume that  $d = 2$  and that we can assume the bulk manifold is a hyperbolic three-manifold  $X$  with conformal boundary  $M$ . (As explained in the introduction, in general there are other possibilities.) The contribution of a given  $X$  is, for small  $G$ , asymptotic to  $\exp(-V_R(X)/4\pi G\ell^2)$ . That simple exponential is multiplied by an asymptotic series of quantum corrections; this series depends only on powers of  $G$ , not an exponential of  $1/G$ . The choice of  $X$  depends on the complex structure of  $M$ , and therefore so does the volume  $V_R(X)$ . For example, if we are trying to compute  $\text{Tr } e^{-\beta H}$ , we take  $M$  to be a torus with a complex structure that depends on  $\beta$  (see section 2.2 for more detail). Then the renormalized volume  $V_R(X)$  depends on  $\beta$  and we can write it more explicitly as  $V_R(X, \beta)$ . Since we have to sum over the choices of  $X$  to compute  $\text{Tr } e^{-\beta H}$ , we get

$$\text{Tr } e^{-\beta H} \sim \sum_{\alpha} \exp(-V_R(X_{\alpha}, \beta)/4G) \cdot F_{\alpha}(\beta, G), \quad (2.6)$$

where the sum runs over the possible choices of three-manifold  $X_{\alpha}$ , and for each  $\alpha$ ,  $F_{\alpha}(\beta, G)$  is the corresponding series of quantum corrections. For any given value of  $c$  or  $G$ , the sum in eqn. (2.6) just produces an analytic function of  $\beta$ . However, the asymptotic behavior for  $c \rightarrow \infty$  or  $G \rightarrow 0$  is dominated by the term in the sum with the smallest possible  $V_R(X_{\alpha}, \beta)$ . As  $\beta$  is varied, there can be a ‘‘crossover’’ with a jump in the choice of  $X$  that minimizes  $V_R(X, \beta)$  and therefore a discontinuous change in the asymptotic behavior of the partition function for small  $G$ . That jump is the Hawking-Page transition. What has just been explained (or its analog in higher dimensions) was actually the original explanation of the Hawking-Page transition [8], long before AdS/CFT duality was understood.

## 2.2 Review Of The Renormalized Volume

The relation of the renormalized volume of a hyperbolic three-manifold to the negative ground state energy of a two-dimensional CFT will be important in what follows, so we will review it in detail. First we recall the relation between the Einstein action and the renormalized volume.

With negative cosmological constant, the Einstein action in a three-dimensional space-time  $X$  of Euclidean signature, with boundary  $\partial X$ , is

$$I = -\frac{1}{16\pi G} \int d^3x \sqrt{g} \left( R + \frac{2}{\ell^2} \right) - \frac{1}{8\pi G} \int_{\partial X} d^2x \sqrt{\gamma} K, \quad (2.7)$$

where  $R$  is the Ricci scalar of  $X$ , and  $\gamma$  and  $K$  are the induced metric and the trace of the second fundamental form of  $\partial X$  ( $R_{\gamma}$  will denote the scalar curvature of  $\gamma$ ). The last

term is the Gibbons-Hawking-York (GHY) boundary term. Using Einstein's equations  $R_{ij} - \frac{1}{2}g_{ij}R - \frac{1}{\ell^2}g_{ij} = 0$ , one can rewrite the bulk part of the action as a multiple of the volume of  $X$ :

$$I = \frac{1}{4\pi G\ell^2} \int_X d^3x \sqrt{g} - \frac{1}{8\pi G} \int_{\partial X} d^2x \sqrt{h} K, \quad (2.8)$$

To this action, one can add ‘‘counterterms,’’ which are simply local integrals over  $\partial X$  of invariant functions of the induced metric. For our purposes there are two relevant terms:

$$\int_{\partial X} d^2x \sqrt{\gamma} \left( \frac{\alpha}{\ell^2} + \lambda R_\gamma \right), \quad \alpha, \lambda \in \mathbb{R}. \quad (2.9)$$

One adjusts  $\alpha$  and  $\lambda$  to cancel the divergent part of the action (2.7) and to make the action conformally invariant, that is, invariant under Weyl transformations of the boundary metric  $\gamma$ , apart from the usual  $c$ -number Weyl anomaly of two-dimensional CFT. The bulk term in eqn. (2.8) is formally  $V/4\pi G$ , where  $V$  is the volume of  $X$ . For a hyperbolic three-manifold with non-empty conformal boundary, this volume is divergent. After renormalization it will be replaced with a renormalized volume  $V_R$ . On the other hand, it turns out that in  $D = 3$ , the GHY boundary term is entirely canceled by renormalization. So the renormalized action will be just  $I_R = V_R/4\pi G$ .

A convenient reference on the necessary computation is [23], which we will follow here (with minor changes of notation). To evaluate the action on an AAdS<sub>3</sub> spacetime, one puts the metric of  $X$  in the form

$$ds^2 = \frac{\ell^2}{4\rho^2} d\rho^2 + \frac{1}{\rho} \sum_{i=1,2} \bar{g}_{ij}(x, \rho) dx^i dx^j \quad (2.10)$$

near the conformal boundary of  $X$ , which in these coordinates is at  $\rho = 0$ . In two dimensions,  $\bar{g}(x, \rho)$  has an expansion

$$\bar{g}(x, \rho) = \bar{g}_{(0)}(x) + \rho \bar{g}_{(2)}(x) + \rho \log \rho h_{(2)}(x) + \mathcal{O}(\rho^2 \log \rho). \quad (2.11)$$

The only facts we need to know from the Einstein equations (eqn. (7) of [23]) are that

$$\begin{aligned} \bar{g}_{(0)}^{kl} \bar{g}_{(2)kl} &= \frac{\ell^2}{2} R_{(0)} \\ \bar{g}_{(0)}^{kl} h_{(2)kl} &= 0, \end{aligned} \quad (2.12)$$

where  $R_{(0)}$  is the Ricci scalar of the metric  $g_{(0)}$ .

The first step in defining the renormalized volume is to ‘‘cut off’’  $X$  by restricting to the region  $X_\epsilon$  with  $\rho \geq \epsilon$ ; to define  $V_R$ , one computes the volume of  $X_\epsilon$  and then takes the limit  $\epsilon \rightarrow 0$  after adjusting the counterterms of eqn. (2.9) to cancel divergences. Using the form (2.10) of the metric with the expansion (2.11) together with (2.12), one finds the divergent parts in the volume  $V(X_\epsilon)$ :

$$\begin{aligned} V(X_\epsilon) &= \int_\epsilon d\rho \int_{\partial X_\epsilon} d^2x \sqrt{\det g_{(0)}} \left( \frac{\ell}{\rho^2} + \frac{\ell^3}{4\rho} R_{(0)} + \mathcal{O}(\log \rho) \right) \\ &= \frac{\ell}{\epsilon} \int_{\partial X_\epsilon} d^2x \sqrt{\det g_{(0)}} + \frac{\ell^3 \log(1/\epsilon)}{4} \int_{\partial X_\epsilon} d^2x \sqrt{\det g_{(0)}} R_{(0)} + \text{finite}. \end{aligned} \quad (2.13)$$

Subtracting the divergent part, we arrive at the definition of the renormalized volume of  $X$ :<sup>6</sup>

$$V_R(X) = \lim_{\epsilon \rightarrow 0} \left( V(X_\epsilon) - \frac{\ell}{\epsilon} \int_{\partial X_\epsilon} d^2x \sqrt{\det g_{(0)}} - \frac{\ell^3 \log(1/\epsilon)}{4} \int_{\partial X_\epsilon} d^2x \sqrt{\det g_{(0)}} R_{(0)} \right). \quad (2.14)$$

The GHY boundary term can be analyzed similarly. One finds with the help of eqn. (2.12) that the GHY boundary term in the action of  $X_\epsilon$  is a linear combination of the two counterterms in eqn. (2.9) (plus a remainder that vanishes for  $\epsilon \rightarrow 0$ ). Hence the GHY boundary term does not contribute to the renormalized Einstein action, which is just  $I_R = V_R/4\pi G\ell^2$ .

Now we can understand in terms of the renormalized volume the fact that  $\text{Tr} e^{-\beta H}$  diverges for large  $\beta$  and large  $c$  as  $e^{\beta c/12} = e^{\beta \ell/8G}$ . To compute  $\text{Tr} e^{-\beta H}$ , we take the conformal boundary to be a two-torus  $M$  parametrized by  $\phi$  and  $t$  with metric  $ds^2 = d\phi^2 + dt^2$  and periodicities<sup>7</sup>  $\phi \cong \phi + 2\pi$  and  $t \cong t + \beta$ . To compute  $\text{Tr} e^{-\beta H}$ , we have to sum over hyperbolic three-manifolds with boundary  $M$ . The dominant one for large  $\beta$  is just  $\text{AdS}_3$  itself with a periodic identification  $t \cong t + \beta$ , making what we might call thermal  $\text{AdS}_3$ . The metric is<sup>8</sup>

$$ds^2 = \left( \frac{r^2}{\ell^2} + 1 \right) \ell^2 dt^2 + \frac{dr^2}{\frac{r^2}{\ell^2} + 1} + r^2 d\phi^2. \quad (2.15)$$

To put this in the desired form of eqn. (2.10), we have to solve

$$\frac{\ell}{2} \frac{d\rho}{\rho} = - \frac{dr}{\left( \frac{r^2}{\ell^2} + 1 \right)^{1/2}}, \quad (2.16)$$

leading to  $\rho = \frac{1}{r^2} - \frac{\ell^2}{2r^4} + \dots$  or

$$r^2 = \frac{1}{\rho} - \ell^2/2 + \mathcal{O}(\rho). \quad (2.17)$$

The cutoff at  $\rho \geq \epsilon$  therefore corresponds to  $r \leq r_m = 1/\epsilon - \ell^2/2$ , so the volume of  $X_\epsilon$  is

$$V(X_\epsilon) = \int_0^{r_m} dr \int_0^{2\pi} d\phi \int_0^\beta dt \sqrt{\det g} = \ell \int_0^{r_m} dr \int_0^{2\pi} d\phi \int_0^\beta dt r = \pi\beta \left( \frac{\ell}{\epsilon} - \frac{\ell^3}{2} \right). \quad (2.18)$$

Subtracting the divergent part, we are left with  $V_R(X) = -\pi\beta\ell^3/2$ , so  $\exp(-V_R/4\pi G\ell^2) = \exp(\beta\ell/8G)$ , as expected.

<sup>6</sup>The conformal anomaly arises from the logarithmically divergent term. If we repeat the calculation after making a Weyl rescaling of the boundary metric  $g_{(0)}$ , the renormalized volume  $V_R(X)$  is shifted by the conformal anomaly.

<sup>7</sup> $\phi$  should have period  $2\pi$  because the statement that the ground state energy of a CFT is  $-c/12$  assumes that the CFT is quantized on a circle of circumference  $2\pi$ , here parametrized by  $\phi$ . To compute  $\text{Tr} e^{-\beta H}$ , we propagate the  $\phi$  circle through imaginary time  $\beta$ , so we need  $t \cong t + \beta$ .

<sup>8</sup>This metric is often written in terms of  $\tilde{t} = \ell t$ . Our normalization ensures that the metric on the torus at infinity is conformal to  $d\tilde{t}^2 + d\phi^2$ .

The moral of the story is that the large negative CFT ground state energy  $-c/12$  for large  $c$  corresponds to the fact that for large  $\beta$ ,  $V_R = -\pi\beta\ell^3/2$  becomes very negative. Generalizing this, our hypothesis that there is no ensemble averaging for fixed energy observables leads to predictions about precisely when  $V_R(X)$ , for a hyperbolic three-manifold  $X$ , can go to  $-\infty$ . This will be explained in section 2.3. But first, we will discuss other contributions to  $\text{Tr} e^{-\beta H}$  to illustrate the fact that in most cases,  $V_R(X)$  does not go to  $-\infty$  when  $\beta$  becomes large.

Before being general, we will describe the special case that is actually important in understanding the Hawking-Page phase transition. In a CFT, we are free to rescale the metric of the torus  $M$ , so instead of saying that  $\phi$  and  $t$  have periods  $2\pi$  and  $\beta$ , we could rescale the metric and say that the periods are  $4\pi^2/\beta$  and  $2\pi$ . Now we can write down the same metric as before but with  $t$  and  $\phi$  exchanged:

$$ds^2 = \left( \frac{r^2}{\ell^2} + 1 \right) \ell^2 d\phi^2 + \frac{dr^2}{\frac{r^2}{\ell^2} + 1} + r^2 dt^2. \quad (2.19)$$

Assuming that  $t$  is regarded as the Euclidean time direction, this is the Euclidean version of the BTZ black hole; we get the actual BTZ black hole if we continue to Lorentz signature<sup>9</sup> by  $t \rightarrow it$ . Obviously, the renormalized volume can be evaluated just as before, but with  $\beta$  replaced by  $4\pi^2/\beta$ . So the leading black hole contribution to  $\text{Tr} e^{-\beta H}$  is  $\exp(\pi^2\ell/2\beta G)$ , obtained from the previous  $\exp(\beta\ell/8G)$  by  $\beta \rightarrow 4\pi^2/\beta$ . Clearly, for  $G$  asymptotically small,  $\exp(\beta\ell/8G)$  and  $\exp(\pi^2\ell/2\beta G)$  exchange dominance at  $\beta = 2\pi$ , and this is the Hawking-Page phase transition.

We can generalize this slightly to describe all hyperbolic three-manifolds whose conformal boundary is the torus  $M$ . First, going back to  $\text{AdS}_3$ , notice that we can slightly generalize the equivalence relation on  $\phi$  and  $t$  that we used before so that the torus at infinity is defined by

$$\begin{aligned} (\phi, t) &\cong (\phi + 2\pi, t) \\ (\phi, t) &\cong (\phi + \alpha, t + \beta). \end{aligned} \quad (2.20)$$

A shift in imaginary time by  $\beta$  is now accompanied by a rotation of the circle parametrized by  $\phi$  by an angle  $\alpha$ , so a CFT path integral on this torus computes  $\text{Tr} e^{-\beta H + i\alpha P}$ , where  $P$  is the operator that generates a rotation of the circle. The eigenvalues of  $H$  and  $P$  on a state that corresponds to an operator of dimension  $\Delta$  and spin  $J$  are  $-\frac{c}{12} + \Delta$  and  $J$ , so

$$\text{Tr} e^{-\beta H + i\alpha P} = e^{\beta c/12} \sum_{i=0}^{\infty} e^{-\beta\Delta_i + i\alpha J_i}. \quad (2.21)$$

---

<sup>9</sup>If we set  $r^2 + \ell^2 = R^2$  and substitute  $t \rightarrow it$ , the line element (2.19) becomes  $R^2 d\phi^2 + \ell^2 \frac{dR^2}{R^2 - \ell^2} - (R^2 - \ell^2) dt^2$ . This is a commonly written form of the BTZ black hole metric (with zero angular momentum) up to constant rescalings of  $R$ ,  $\phi$ , and  $t$ . Note that in this way of writing the BTZ metric, the black hole mass is encoded entirely in the period of the  $\phi$  variable. The black hole horizon is at  $R = \ell$ , where the coefficient of  $dt^2$  vanishes. As usual the vanishing of this coefficient represents only a breakdown of the coordinate system, and the Lorentz signature geometry has a real analytic continuation beyond this horizon.

The computation of  $V_R$  is not affected by the angle  $\alpha$ , so  $V_R = -\pi\beta\ell^3/2$ , independent of  $\alpha$ . This agrees with the fact that the exponentially growing factor  $e^{\beta c/12}$  in the CFT partition function (2.21) does not depend on  $\alpha$ , since the spin of the ground state is  $J_0 = 0$ .

If we set  $z = (\phi + it)/2\pi$ , we see that the two-torus that is defined in eqn. (2.20) is the quotient of the complex  $z$ -plane by the lattice generated by 1 and  $(i\beta + \alpha)/2\pi$ . So it can be viewed as a complex elliptic curve with the modular parameter

$$\tau = \frac{i\beta + \alpha}{2\pi}. \quad (2.22)$$

In this language, the formula for the renormalized volume becomes

$$V_R(\tau) = -\pi^2\ell^3\text{Im}(\tau). \quad (2.23)$$

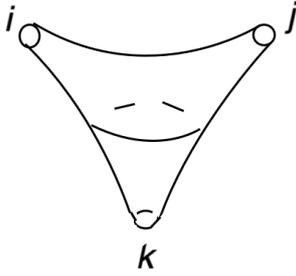
Let  $u$  be the image in the boundary torus  $M$  of a straight line from  $z = 0$  to  $z = 1$ , and let  $v$  be the image of a straight line from  $z = 0$  to  $z = \tau$ . Thus  $u$  and  $v$  are circles in  $M$ . In the thermal  $\text{AdS}_3$  space of eqn. (2.15),  $u$  is the boundary of a disc and  $v$  is not.<sup>10</sup> In the Euclidean black hole, since it is defined by exchanging  $t$  and  $\phi$ , the roles are reversed:  $v$  is the boundary of a disc and  $u$  is not. More generally, if  $c, d$  are any relatively prime integers, there exists a hyperbolic three-manifold  $X_{c,d}$  with boundary  $M$  such that  $cu + dv$  is the boundary of a disc in  $X_{c,d}$  and other linear combinations are not. In this notation,  $X_{0,1}$  is thermal  $\text{AdS}_3$  and  $X_{1,0}$  is the Euclidean black hole. Generalizing the way we introduced the Euclidean black hole,  $X_{c,d}$  is defined by replacing the angles  $2\pi t/\beta$  and  $\phi$  with integer linear combinations of themselves. In other words,  $X_{c,d}$  is actually the same manifold as the original thermal  $\text{AdS}_3$  space, but with the boundary parametrized differently. The  $X_{c,d}$  are actually the complete set of hyperbolic three-manifolds with boundary  $M$ . The renormalized volume of  $X_{c,d}$  is

$$V_R(X_{c,d}) = -\pi^2\ell^3 \frac{\text{Im}(\tau)}{|c\tau + d|^2}. \quad (2.24)$$

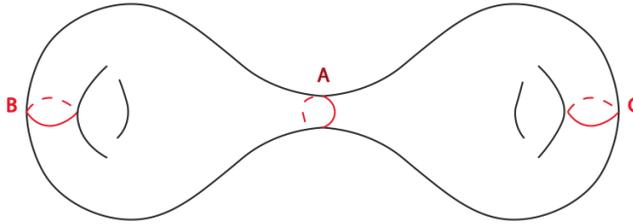
For our purposes, what is notable about this formula is that except in the original case  $c = 0, d = 1$  of thermal  $\text{AdS}_3$ , it never happens that  $V_R \rightarrow -\infty$  for  $\beta \rightarrow \infty$ . On the contrary, whenever  $c \neq 0$ , eqn. (2.14) implies that  $V_R \rightarrow 0$  for  $\beta \rightarrow \infty$ . Therefore, the thermal  $\text{AdS}_3$  space  $X_{0,1}$  is the only one of these manifolds that contributes to the partition function of the fixed energy states, or indeed to any states below the black hole threshold.

None of the observations in this section are in any way new. The computation of the renormalized volume of thermal  $\text{AdS}_3$  and the BTZ black hole is equivalent to the analysis of ADM masses in the original BTZ paper [21]. As noted earlier, the CFT interpretation of the result of this computation goes back to [22], and in turn to the construction of a boundary stress tensor [19]. Summation over the manifolds  $X_{c,d}$  has been considered in several previous papers – in [24, 25] to count certain supersymmetric black hole states in  $\text{AdS}_3$ , and in [26] in an attempt to construct a partition function of three-dimensional pure gravity. The special role of  $X_{0,1}$  was part of those analyses.

<sup>10</sup>The coordinates  $r, \phi$  can be viewed as polar coordinates for a copy of  $\mathbb{R}^2$ . Then  $u$  is a large circle in this  $\mathbb{R}^2$  and therefore is the boundary of a disc.



**Figure 1.** A three-holed sphere, with its boundaries labeled by fixed energy states  $i, j, k$ .



**Figure 2.** A genus two surface  $M$  with three nonintersecting and homotopically independent one-cycles labeled A, B, and C.  $M$  is the conformal boundary of a hyperbolic three-manifold  $X$ . If  $X$  is such that  $V_R(X) \rightarrow -\infty$  when A is pinched, then the contribution of  $X$  to  $Z(M)$  has a part that describes propagation through A of a fixed energy state. In general, however, black hole states are propagating through B and C and this amplitude is subject to ensemble averaging. If we stipulate that  $V_R(X) \rightarrow -\infty$  when any of A, B, or C is pinched, we get an amplitude that has a contribution that describes fixed energy states propagating through each of A, B, and C, and interacting via two three-holed spheres, one to the left of A and one to the right. A contribution of this type, according to our conjecture, should not be subject to ensemble averaging. So this behavior of  $V_R(X)$  should be possible only if the conformal boundary of  $X$  is connected.

### 2.3 Implications of The Conjecture For Hyperbolic Geometry

We have proposed that certain observables are not affected by ensemble averaging: the excitation energies  $\Delta_i$  and spins  $J_i$  of the fixed energy states, and their trilinear couplings  $\lambda_{ijk}$ . Those couplings can be computed as three-point functions  $\langle \mathcal{O}_I(x) \mathcal{O}_j(y) \mathcal{O}_k(z) \rangle$  on  $S^2$ . Another description will be more useful in what follows: because of the operator-state correspondence of CFT, the  $\lambda_{ijk}$  can also be computed by a path integral on a three-holed sphere, with the fixed energy states  $i, j, k$  inserted on its boundaries (fig. 1).

In order to test our proposal using properties of hyperbolic three-manifolds, we want to identify observables in genus  $g \geq 2$  that can be determined in terms of  $\Delta_i$ ,  $J_i$ , and  $\lambda_{ijk}$ . Hopefully, the hyperbolic geometry will work out in such a way that amplitudes that can be determined in terms of  $\Delta_i$ ,  $J_i$ , and  $\lambda_{ijk}$  receive contributions only from hyperbolic manifolds with connected boundary.

To illustrate the idea, drawn in fig. 2 is a genus 2 surface  $M$ , along with three nonintersecting and homotopically independent one-cycles A, B, and C. From a conformal point

of view, it is equivalent to say that a one-cycle such as  $A$  is becoming short, or is being “pinched,” or that the tube through  $A$  is becoming long. (We used this equivalence in section 2.2 when we said that it is equivalent conformally to consider the circle parametrized by  $t$  to be long, of circumference  $\beta \gg 1$ , or to consider the circle parametrized by  $\phi$  to be short, of circumference  $4\pi^2/\beta \ll 1$ .) When  $A$  is pinched or the tube is long, the CFT partition function  $Z(M)$  grows exponentially, because of the negative CFT ground state energy.

Now let  $X$  be a hyperbolic three-manifold that has  $M$  in its conformal boundary. If the conformal boundary of  $X$  consists only of  $M$ , then the path integral  $Z_X$  contributes to  $Z(M)$ . If the conformal boundary is  $M \sqcup M'$  for some  $M'$  (which may or may not be connected), then  $Z_X$  contributes to a connected correlation function  $\langle Z(M)Z(M') \rangle_c$ . In either case, we ask what happens when a one-cycle such as  $A$  or  $B$  or  $C$  is pinched. Does  $Z_X$  grow exponentially, reflecting in terms of the boundary CFT a sub-threshold state propagating through the cycle that is being pinched?<sup>11</sup>

Since  $Z_X$  is asymptotic for small  $G$  to  $\exp(-V_R/4\pi G)$ , a necessary condition for  $Z_X$  to show exponential growth in the pinching limit is that  $V_R(X)$  must go to  $-\infty$  in that limit. In section 3 and appendix A we determine the condition on  $X$  such that  $V_R(X)$  goes to  $-\infty$  when a given boundary cycle  $A$  is pinched. The answer is that this occurs if and only if  $A$  is the boundary of a disc in  $X$ . This generalizes the previously known facts for genus 1 that were summarized in section 2.2: of the manifolds  $X_{c,d}$ , the only one with the property that  $\lim_{\beta \rightarrow \infty} V_R(X_{c,d}) = -\infty$  is the thermal AdS<sub>3</sub> space  $X_{0,1}$ , and this is also the only one in which the circle parametrized by  $\phi$  (which is the one that is pinched for  $\beta \rightarrow \infty$ ) is the boundary of a disc. Mathematically, if a circle  $A$  in a component  $M$  of the conformal boundary of  $X$  is the boundary of a disc in  $X$  (but not in  $M$ ), then  $A$  is said to be “compressible” in  $X$ . If  $M$  contains a compressible circle, then  $M$  itself is said to be compressible.

In most cases, a given boundary circle  $A$  is not compressible, so  $Z_X$  does not contribute to an amplitude in which a sub-threshold state propagates through  $A$ . For many choices of  $X$ , there is no compressible circle at all in a given boundary component  $M$ . (An example is the Fuchsian manifold that we discuss later.) But even if  $A$  is compressible in  $X$ ,  $X$  typically does not contribute to an amplitude on  $M$  that we expect to be free of ensemble averaging, because even if a sub-threshold state is propagating through  $A$ , there may be black hole states propagating in the rest of the Riemann surface. For example, in fig. 2, even if a sub-threshold state is propagating through  $A$ , the states propagating on the genus one Riemann surfaces to the left and right of  $A$  may be black hole states. To identify something that we expect to be free of ensemble averaging, we reason as follows. If we cut the genus two surface  $M$  on the three nonintersecting and homotopically independent one-cycles  $A$ ,  $B$ , and  $C$ , it decomposes into the union of two three-holed spheres. A contribution to  $Z(M)$  in which specified sub-threshold states  $i, j, k$  are propagating through  $A$ ,  $B$ , and  $C$  can be

<sup>11</sup>In order for a state of fixed energy above the ground state to propagate through the cycle that is being pinched, we need a stronger condition that  $Z_X$  grows as  $e^{\beta c/12}$ . It turns out, however, that the interesting constraints on  $X$  arise if one merely asks for exponential growth of  $Z_X$ , without specifying the rate, and therefore it will not be important to distinguish fixed energy states from sub-threshold states.

evaluated in terms of the dimensions and spins of the three states and the path integrals on the two three-holed spheres, The boundaries of the two three-holed spheres are labeled by the particular sub-threshold states that propagate through A, B, and C (in the example sketched in the figure, if states  $i, j, k$  propagates respectively through A, B, and C, then the labels are  $ijj$  for the three-holed sphere on the left and  $ikk$  for the three-holed sphere on the right). The path integral on such a labeled three-holed sphere computes a trilinear coupling of sub-threshold states. Since, according to our conjecture, the dimensions and spins of the sub-threshold states and their trilinear couplings are all unaffected by ensemble averaging, we expect that a contribution to  $Z(M)$  with sub-threshold states propagating through A, B, and C is not subject to ensemble averaging. On the other hand, a necessary condition for  $X$  to contribute to an amplitude with sub-threshold states propagating through each of A, B, and C is that  $V_R(X)$  must go to  $-\infty$  when any of A, B, or C is pinched. Equivalently, in view of what has already been said, A, B, and C must all be compressible in  $X$ . Conversely, if A, B, and C are all compressible in  $X$ , then  $V_R(X)$  does go to  $-\infty$  when any of those cycles is pinched. In that case, we expect that the conformal boundary of  $X$  is connected and consists only of  $M$ , ensuring that  $X$  does not contribute to a connected amplitude between disconnected boundaries.

This discussion has a straightforward generalization to higher genus. If a surface  $M$  of genus  $g \geq 2$  is cut along  $3g - 3$  nonintersecting and homotopically independent circles  $A_\sigma$ ,  $\sigma = 1, \dots, 3g - 3$ , then it decomposes into a union of  $2g - 2$  three-holed spheres. A contribution to  $Z(M)$  in which a sub-threshold state is propagating through each of the  $A_\sigma$ , should, according to our conjecture, not be subject to ensemble averaging. For a hyperbolic manifold  $X$  whose conformal boundary contains  $M$  to contribute to such an amplitude, the  $A_\sigma$  must all be compressible in  $X$ . So we expect that if  $M$  contains  $3g - 3$  nonintersecting and homotopically independent circles  $A_\sigma$  that are all compressible in  $X$ , then the conformal boundary of  $X$  consists only of  $M$ .

As we explain in section 3 and appendix A, this is true and in fact more is true. If at least  $3g - 5$  homotopically independent and non-intersecting one-cycles in  $M$  are compressible in  $X$  (or  $3g - 6$  of them for  $g \geq 3$ ), then the conformal boundary of  $X$  is connected and consists only of  $M$ , and moreover  $X$  is a Schottky manifold. A Schottky manifold is the simplest type of hyperbolic three-manifold. A Schottky manifold with conformal boundary  $M$  is, topologically, the “interior” of  $M$  for some embedding of  $M$  in  $\mathbb{R}^3$ . For example, the picture drawn in fig. 2 suggests an embedding of a genus two surface in  $\mathbb{R}^3$ . The interior of  $M$  for this embedding is a three-manifold in which A, B, and C are all compressible. Topologically, this interior is called a handlebody. A handlebody with boundary  $M$  admits a hyperbolic metric with  $M$  as the conformal boundary (for any choice of the conformal structure on  $M$ ). Endowed with such a metric the handlebody is called a Schottky manifold.

From what has just been explained, we learn, modulo the arguments in section 3 and appendix A, that three-dimensional hyperbolic geometry, at least for the questions that we have asked, is consistent with our hypothesis that certain AdS/CFT observables are not affected by ensemble averaging.

Some of the points can be illustrated with the simple example of a Fuchsian manifold.

This is a three-manifold that is topologically of the form  $X = M \times \mathbb{R}$ , where  $M$  is a Riemann surface of genus  $g \geq 2$ .  $X$  carries a hyperbolic metric of the form<sup>12</sup>

$$ds^2 = \ell^2 (dt^2 + \cosh^2 t d\Omega^2), \quad (2.25)$$

where  $d\Omega^2$  is a metric on  $M$  of constant scalar curvature  $R = -2$ . The conformal boundary of  $X$  consists of two copies of  $M$ , at  $t = -\infty$  and  $t = +\infty$ , respectively; let us call these  $M_1$  and  $M_2$ .  $M_1$  and  $M_2$  have the same complex structure since they have the same metric  $d\Omega^2$ . The path integral on  $X$  contributes to a connected amplitude  $\langle Z(M_1)Z(M_2) \rangle_c$ . However, a simple calculation shows that  $V_R(X)$  is a topological invariant, independent of the complex structure of  $M_1$  and  $M_2$ .  $V_R$  never goes to  $-\infty$  regardless of how we vary the complex structure of the boundary, so the path integral on  $X$  never shows the exponential growth characteristic of a sub-threshold state.

The drawback of this simple example is that since  $M_1$  and  $M_2$  have the same metric, we cannot vary their complex structures independently. What happens, for example, if we pinch a cycle in  $M_1$  without changing  $M_2$ ? In fact, there is a more general family of hyperbolic metrics on  $X$  such that the complex structures on  $M_1$  and  $M_2$  do vary independently. These metrics are called quasi-Fuchsian, and unfortunately in a situation that involves pinching on only one side, they are only known by existence proofs, not explicit formulas. At any rate, no one-cycle in either  $M_1$  or  $M_2$  is compressible, so the results of section 3 and appendix A imply that  $V_R(X)$  never goes to  $-\infty$ . Thus, although  $X$  does contribute to a connected amplitude  $\langle Z(M_1)Z(M_2) \rangle_c$ , its contribution only involves black hole states, with no sub-threshold states appearing in any channel.

So far in this section, we have only considered the case of a surface of genus  $\geq 2$ , and in fact the case of genus 1 is exceptional. A hyperbolic three-manifold whose conformal boundary contains a component of genus 1 is always one of the manifolds  $X_{c,d}$  that were discussed in section 2.2. In particular, at the level of hyperbolic geometry there are no disconnected amplitudes involving a torus  $\mathbb{T}^2$ . This is, however, not the whole story. There is good reason to suspect [17] the existence in AdS/CFT models of a connected correlation function  $\langle Z(M)Z(M') \rangle_c$  where  $M$  and  $M'$  are tori, even though there are no classical hyperbolic manifolds that can generate such a contribution. It has been argued in [12] that a path integral on  $\mathbb{R} \times \mathbb{T}^2$  contributes to  $\langle Z(M)Z(M') \rangle_c$ , even though no classical solution is available on this manifold. That paper actually contains in eqn. (3.56) an interesting formula for the contribution of  $\mathbb{R} \times \mathbb{T}^2$  to  $\langle Z(M)Z(M') \rangle_c$ , with independent complex structures on  $M$  and  $M'$ . We do not know if this formula is precisely correct. However, the formula is independent of  $G$  and in particular shows no exponential growth for  $G \rightarrow 0$ . So if it is even qualitatively correct, the contribution of  $\mathbb{R} \times \mathbb{T}^2$  to  $\langle Z(M)Z(M') \rangle_c$  is completely consistent with our conjectures.

---

<sup>12</sup> The submanifold  $X_C$  defined by  $t = 0$  is totally geodesic and moreover is geodesically convex (any geodesic in  $X$  between two points in  $X_C$  is actually contained in  $X_C$ ).  $X_C$  is called the “convex core” of  $X$ , a notion that will be important in section 3. This example is exceptional, because  $X_C$  has volume 0. Apart from a Fuchsian manifold, or a solid torus, such as the BTZ black hole, the convex core of any other hyperbolic three-manifold, including the quasi-Fuchsian ones that are discussed momentarily, has positive volume. (The convex core is not defined for  $\text{AdS}_3$  itself.)

There is one other case known of manifolds that do not admit classical hyperbolic metrics but nonetheless make relatively well understood contributions to the path integral of three-dimensional gravity with  $\Lambda < 0$ . These are Seifert fibered manifolds, whose contributions were analyzed in [13] in the Kaluza-Klein limit, that is, the limit that the fiber is small. As the remaining moduli are varied, these path integrals do not show contributions of sub-threshold states.

### 3 Overview of Mathematical Arguments

In the following,  $M$  is an oriented two-manifold that is a component of the conformal boundary of a hyperbolic three-manifold  $X$ . It is known by elementary arguments that if  $M$  has genus 0, then  $X$  must be  $\text{AdS}_3$  itself, and if  $M$  has genus 1,  $X$  is one of the manifolds  $X_{c,d}$  that were introduced in section 2.2. Therefore, we are primarily interested in the case that  $M$  has genus at least 2. Such an  $M$  admits a hyperbolic metric of constant scalar curvature  $R = -2$ ; the space of such metrics, up to diffeomorphisms of  $M$  that are isotopic to the identity, is the Teichmüller space  $\mathcal{T}(M)$ . As one approaches the boundary of  $\mathcal{T}(M)$ , it is possible for the length of a simple closed geodesic  $\gamma$  to go to 0; we will say that in that limit,  $\gamma$  is “pinched.” We say that a collection of embedded circles in  $M$  are “independent” if they are non-intersecting and homotopically independent. If  $M$  has genus  $g \geq 2$ , a maximal set of independent circles in  $M$  consists of  $3g - 3$  circles, as in fig. 3. If  $g = 1$ , a maximal set consists of just one circle. An embedded circle in  $M$  is compressible in  $X$  if it is homotopically nontrivial in  $M$  but bounds a disc in  $X$ . As explained in more detail in section 2.3, a Schottky manifold is topologically the “interior” of  $M$ , for some embedding of  $M$  in  $\mathbb{R}^3$ . Topologically the  $X_{c,d}$  are solid tori, the genus 1 analogs of a Schottky manifold.

The arguments of section 2 relied on two mathematical assertions:

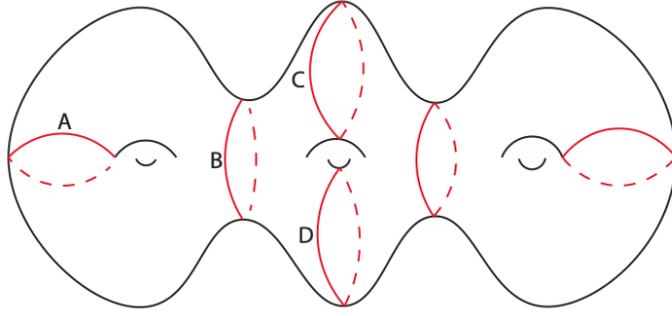
**Proposition 3.1.** *If a component  $M$  of the conformal boundary of  $X$  contains a collection of at least  $3g - 5$  independent circles that are compressible in  $X$ , then the boundary of  $X$  is connected and consists only of  $M$ , and moreover  $X$  is a Schottky manifold.*

**Theorem 3.2.** *In the limit that a simple closed geodesic  $\gamma \subset M$  is pinched, the renormalized volume  $V_R(X)$  remains finite if  $\gamma$  is not compressible in  $X$ , and approaches  $-\infty$  if  $\gamma$  is compressible in  $X$ .*

Here we will give a very rough sketch of the proofs.<sup>13</sup> Further detail is explained in appendix A.

One important tool in the arguments is that, by varying the hyperbolic metric of  $X$ , the conformal structure of  $M$  can be varied arbitrarily. Indeed, by a classic result, the moduli space of hyperbolic metrics on  $X$  is the space of all complex structures on the

<sup>13</sup>The bound  $3g - 5$  in Proposition 3.1 is stronger than we actually needed in section 2, where a bound of  $3g - 3$  would have sufficed. We do not know if the stronger bound is significant in the AdS/CFT context. In appendix A.6, we prove a slightly sharper bound for  $g \geq 3$  (Theorem A.4). Likewise Theorem 3.2 is sharpened in Theorem A.15, as remarked at the end of this section.



**Figure 3.** A surface  $M$  of genus  $g = 3$ , with  $3g - 3 = 6$  independent circles marked. Surgery on  $A$  produces a surface  $M_1$  of genus  $g_1 = g - 1 = 2$ . In passing to  $M_1$ ,  $A$  disappears,  $B$  becomes nullhomotopic, and  $C$  and  $D$  become homotopic to each other, leaving three independent circles on  $M_1$ . Surgery on  $B$  leaves a surface  $M'_1$  of genus  $g'_1 = 1$  and a surface  $M''_1$  of genus  $g''_1 = 2$ . Of the original 6 independent circles on  $M$ ,  $B$  disappears in the surgery,  $A$  remains on  $M'_1$ , and  $C$  and  $D$  become homotopic. We remain with  $3g''_1 - 3 = 3$  independent circles on  $M'_2$ .

conformal boundary of  $X$  (whether that conformal boundary is connected or not) modulo diffeomorphisms of the conformal boundary that extend over  $X$ .

Therefore, in discussing Proposition 3.1, we can pick the complex structure of  $M$  to be such that the  $3g - 5$  independent compressible circles correspond to disjoint simple closed geodesics that are all very short. Let  $\gamma$  be one of these. As  $\gamma$  is compressible, it is the boundary of a disc  $P \subset X$ . In appendix A, we show that one can choose  $P$  to be a totally geodesic plane (thus, a copy of  $\text{AdS}_2$ ) embedded in  $X$ . This makes it possible to perform a simple “surgery” on  $X$ . We cut  $X$  along  $P$ . Along each of the resulting boundaries, we glue in a copy of half of  $\text{AdS}_3$ . (Concretely, one can cut  $\text{AdS}_3$  in half along a geodesic plane  $\text{AdS}_2 \subset \text{AdS}_3$ , and then glue in the resulting pieces along the cuts in  $X$ .) This gives a new hyperbolic manifold  $X_1$ , which may or may not be connected. The effect of the surgery on  $M$  is to cut  $M$  along  $\gamma$  and glue in a disc on each side, giving a new oriented two-manifold  $M_1$ . Even if  $X_1$  is connected,  $M_1$  may not be. Components of the conformal boundary of  $X$  other than  $M$ , if there are any, are not affected by the surgery.

If  $M_1$  is connected, it has genus  $g_1 = g - 1$ . Of the  $3g - 5$  independent compressible curves that we started with on  $M$ , one, namely  $\gamma$ , disappeared in the surgery. The other  $3g - 4$  independent compressible circles in  $M$  are still compressible in  $X_1$ . At most 2 of them are no longer independent in  $M_1$  (they are nullhomotopic or homotopic to each other; see fig. 3). Therefore, if  $M$  has  $3g - 5$  independent compressible circles, then  $M_1$  has at least  $3g - 8 = 3g_1 - 5$  such circles. Now let us assume inductively that Proposition 3.1 is true for a conformal boundary component of genus  $g - 1$ . By this inductive hypothesis applied to  $M_1$ , we learn that  $X_1$  is a Schottky manifold. Given this, an elementary geometric argument shows that  $X$  is also a Schottky manifold, and in particular its conformal boundary is connected.

If  $M_1$  is not connected, it is the union of components  $M'_1, M''_1$  of genera  $g'_1, g''_1 \geq 1$ , with  $g'_1 + g''_1 = g$ . First consider the case  $g'_1, g''_1 \geq 2$ . Let  $n'_1, n''_1$  be the maximum number of

independent compressible circles on  $M'_1, M''_1$ . We have  $n'_1 + n''_1 \geq 3g - 8$  (we started with at least  $3g - 5$  such circles on  $M$ ;  $\gamma$  was lost and at most one compressible circle on  $M'_1$  and one on  $M''_1$  is no longer independent after the surgery, leaving at least  $3g - 8$  of them). On the other hand,  $n'_1 \leq 3g'_1 - 3$ ,  $n''_1 \leq 3g''_1 - 3$ , since a Riemann surface of genus  $h$  supports at most  $3h - 3$  independent circles. So  $n'_1 \geq 3g'_1 - 5$ ,  $n''_1 \geq 3g''_1 - 5$ . Hence by the inductive hypothesis applied to  $M'_1$  and  $M''_1$ ,  $X_1$  must consist of disjoint components  $X'_1$  and  $X''_1$ , where  $X'_1$  is a Schottky manifold of conformal boundary  $M'_1$ , and  $X''_1$  is a Schottky manifold of conformal boundary  $M''_1$ . Again it follows by an elementary geometric argument that  $X$  is also a Schottky manifold, in particular with connected conformal boundary.

The same conclusion applies if  $M'_1$  and/or  $M''_1$  has genus 1. For example, if  $M'_1$  has genus 1, then  $M''_1$  has genus  $g''_1 = g - 1$ . In this case, of the original compressible circles on  $M$ , at most 1 was originally on  $M'_1$  and (assuming  $g''_1 \geq 2$ ) at most 1 which was originally on  $M''_1$  is no longer independent after the surgery. So there are at least  $3g - 8 = 3g''_1 - 5$  independent compressible circles on  $M''_1$ , and the inductive hypothesis applies to  $M''_1$  as before. As for  $M'_1$ , since it has genus 1, we invoke the fact that a hyperbolic three-manifold whose conformal boundary has a genus 1 component is one of the  $X_{c,d}$ , topologically a solid torus. So  $X$  is the disjoint union of two components, one a solid torus with conformal boundary  $M'_1$  and one a Schottky manifold with conformal boundary  $M''_1$ . Again it follows that  $X$  is also a Schottky manifold. A similar argument applies in the special case  $g = 2$ ,  $g'_1 = g''_1 = 1$ . This completes the proof of Proposition 3.1.

The proof of Theorem 3.2 requires more sophisticated tools. We must prove two statements: (i) if  $\gamma$  is not compressible in  $X$ , then when  $\gamma$  is pinched,  $V_R(X)$  remains bounded; (ii) if  $\gamma$  is compressible, then when  $\gamma$  is pinched,  $V_R(X) \rightarrow -\infty$ .

To prove the first statement, we use the fact that the pinching locus is at finite distance in the Weil-Petersson metric on Teichmüller space. Therefore, to show that  $V_R(X)$  remains bounded as one approaches the pinching locus, it suffices to know that the gradient of  $V_R(X)$  in the Weil-Petersson metric remains bounded. In fact, it is known that the Weil-Petersson gradient of  $V_R(X)$  remains bounded as long as no curve in  $M$  that is compressible in  $X$  becomes short. Specifically, if  $\ell$  is the length of the shortest non-trivial closed curve in  $M$  that is compressible in  $X$ , and  $\chi(M)$  is the Euler characteristic of  $M$ , then [51, 53] the gradient satisfies the bound

$$\|dV_R\|_{\text{WP}} \leq \frac{3\sqrt{\pi|\chi(\partial_\infty X)|}}{\sqrt{2}\tanh^2(\ell/4)}. \quad (3.1)$$

For completeness, we provide a proof in appendix A.

For the second statement, we need an upper bound on  $V_R(X)$ . There is a useful upper bound in terms of the volume of the convex core of  $X$ , which we will denote as  $V_C(X)$ , and the length of the measured bending lamination of the convex core, which we will denote as  $L_m(l)$ . The meaning of these terms is described in appendix A. In terms of these quantities, one has a bound on  $V_R(X)$ :

$$V_R(X) \leq V_C(X) - \frac{1}{4}L_m(l) + \frac{\pi \log 2}{2}|\chi(M)|. \quad (3.2)$$

(The coefficient of the last term on the right of this equation depends on a choice of normalization in the definition of  $V_R$ .) This inequality can be found in [39, Theorem 1.1] for quasi-Fuchsian manifolds, but the proof extends without change for general  $X$ ; see [51, Section 3].

In appendix A, we show that  $V_C(X)$  remains bounded when one or more compressible curves is pinched. One also has a result of Bridgeman and Canary (see [67, Theorem 2], and also [51, Theorem 4.2]) showing that  $L_m(l) \rightarrow \infty$  when a compressible curve is pinched. Specifically, there are constants  $P, Q$  (one can take  $P = 74$  and  $Q = 36$ ) such that if  $M$  contains a compressible closed geodesic of length  $r < 1$ , then

$$L_m(l) \geq \frac{P}{r} - Q. \quad (3.3)$$

The second part of Theorem 3.2, asserting that  $V_R(X) \rightarrow -\infty$  when a compressible curve is pinched, follows from the bounds (3.2) and (3.3) along with the fact that  $V_C(X)$  remains bounded in this limit.

Physically, one would expect a more precise result than we have stated so far. One would expect that when a compressible cycle is pinched, the divergence of  $V_R$  would precisely reflect the CFT ground state propagating through the cycle in question. With some more detailed arguments, we establish this in Theorem A.15.

#### 4 Why Do Some Observables Show Ensemble Averaging?

As explained in the introduction, connected amplitudes with disconnected boundary, or CADB amplitudes for short, have been a puzzle since early days of the AdS/CFT correspondence. A possible explanation has been that actually, the dual of a specific bulk theory is the average of an ensemble of boundary theories, rather than a specific boundary theory. Averaging over an ensemble can readily generate CADB amplitudes. There is a standard objection to this proposal: in many examples of AdS/CFT duality, it is believed that all of the parameters that the CFT can depend upon (consistent with its general properties such as the supersymmetry algebra it satisfies) are known, and the bulk theory depends on all of the same parameters. So what ensemble could one possibly be averaging over to generate CADB amplitudes?

In this article, we have attempted to sharpen this puzzle by arguing that a certain important class of observables, namely the ones that can be defined purely in terms of energies and couplings of states that are below the black hole threshold, does not receive any contributions with disconnected boundaries and thus is not affected by ensemble averaging.

If there is no ensemble to average over, and if states below the black hole threshold do not show any sign of ensemble averaging, why is it that when we compute observables involving black holes states, the gravitational path integral appears to give ensemble-averaged answers? Clearly the answer must involve some essential difference between fixed energy states and black hole states.

Here we will propose a simple answer to this question, based on two assertions:

- Black hole physics is highly chaotic.

- The Hamiltonian  $H_N$  describing black hole states does not have a large  $N$  limit, and likewise other CFT observables involving black hole states, such as the trilinear couplings  $\langle \mathcal{O}_i \mathcal{O}_j \mathcal{O}_k \rangle$ , do not have a large  $N$  limit, even in a rather general sense, as will be further discussed presently.

The first statement is generally accepted, based on a reinterpretation [14] of older calculations [27] of the behavior of perturbations in the field of a black hole. This statement involves a contrast between black holes and fixed energy states, because in a number of important examples, the spectrum of fixed energy states is believed to be described by an integrable model [10], not a system with chaotic behavior. The second statement also involves a contrast between black holes and fixed energy states. Fixed energy states are the states that we see if we take  $N \rightarrow \infty$  keeping fixed the excitation energy above the ground state. AdS/CFT duality implies that the energies and couplings of such states have a large  $N$  limit; when the boundary theory is a gauge theory, this can also be seen via a classic analysis of Feynman diagrams [20]. To reach the black hole region, we take  $N \rightarrow \infty$  with an excitation energy of order  $N^2$  if the boundary theory is a gauge theory (and a different positive power of  $N$  in other cases). The literature does not contain any proposal concerning a sense in which the Hamiltonian and other observables of black hole states have a large  $N$  limit. Since the entropy (for black hole states of a fixed temperature) is also growing as a power of  $N$ , the dimension of the black hole Hilbert space (at a fixed temperature) increases by a vast factor from one large value of  $N$  to the next. For example, in the case that the boundary theory is a gauge theory, since the entropy is asymptotically  $S = bN^2$ , with  $b$  of order 1, when one changes  $N$  from  $10^6$  to  $10^6 + 1$ , the dimension  $e^S$  of the Hilbert space increases by a vast factor  $e^{2b \times 10^6}$ . This makes it unclear in what sense one might hope that the black hole Hilbert space and other observables would have a large  $N$  limit. In the somewhat analogous problem of quantum statistical mechanics with the volume  $V$  playing the role of  $N$ , the standard answer is that the Hilbert space and Hamiltonian do not have a large  $V$  limit. By contrast, the thermofield double state of a pair of entangled systems does have both a large  $V$  [28] and large  $N$  [16] limit. See [15] for more discussion.

Most likely, the black hole Hamiltonian and couplings do not have a large  $N$  limit, in the sense that, in general, energies and couplings of black hole states do not have any regularity for large  $N$  beyond what follows from the fact that thermodynamic functions and other averages over the spectrum depend smoothly on  $N$  and that, similarly, certain asymptotic averages of functions of couplings are also smooth functions of  $N$ . Asymptotic formulas for averages of functions of couplings were introduced in [29] and have been studied in a number of more recent papers.

Our proposal is that these differences between black hole states and fixed energy states are the reason that apparent ensemble averaging affects black hole states and not fixed energy states. Let  $H_N$  be the CFT Hamiltonian at given  $N$ , on a sphere  $S^{d-1}$  with round metric.  $H_N$  commutes with a symmetry group  $G$  consisting of rotations of  $S^{d-1}$  and possible additional symmetries of the CFT, and so is block diagonal with blocks labeled by representations of  $G$ . Chaos in black hole physics means that if we restrict to states

in a band of energies that is above the black hole threshold, then  $H_N$  in each block is an enormous pseudorandom matrix. A pseudorandom matrix is a matrix that cannot be distinguished from a truly random matrix by any simple measurement. If it is true that  $H_N$  does not have a large  $N$  limit above the black hole threshold, this suggests that in each block, the  $H_N$  for neighboring values of  $N$  can be viewed as independent pseudorandom draws from a random matrix ensemble. (As we explain later, it seems that this statement is actually subject to corrections that are exponentially small in  $N$ , but it can serve as a first approximation.) The random matrix ensemble is characterized by specifying the entropy  $S$  as a function of the energy  $E$  and other conserved charges, so it depends smoothly on  $N$ .

Let us consider CFT observables that can be constructed just in terms of  $H_N$ . The most important such observables are the twisted partition functions  $Z_{N,R}(\beta) = \text{Tr} e^{-\beta H_N} R$ , where  $\text{Re} \beta$  is positive and is small enough that the trace is dominated by black hole states, and  $R \in G$ . But the following explanation may be clearer if we think first about an arbitrary observable  $\mathcal{W}_N$  that depends only on the pseudorandom matrix  $H_N$ .  $\mathcal{W}_N$  may be a “self-averaging” function in random matrix theory, meaning that it has almost the same value for almost any draw from the random matrix ensemble. In that case  $\langle \mathcal{W}_N \rangle$  will be a smooth function of  $N$ , modulo exponentially small corrections that reflect the fact that even self-averaging functions of a random matrix differ slightly from draw to draw. (These corrections are exponentially small because the size  $e^S$  of the random matrix is exponentially large, as observed in [17].) The corrections to self-averaging behavior will depend erratically on  $N$ , since they depend on a pseudorandom draw which is different for each  $N$ . If  $\mathcal{W}$  is not self-averaging, it will be an erratic function of  $N$ . With presently known methods, the gravitational path integral always produces a smooth function of  $N$ , typically by summing over contributions of saddle points corresponding to classical solutions. Even when classical solutions are not available, calculations that we know how to perform lead to smooth functions of  $N$ , as in [12].

Based on this, what might be calculable with presently available methods? If  $\mathcal{W}_N$  is self-averaging, we can hope to calculate  $\langle \mathcal{W}_N \rangle$  modulo exponentially small terms that depend erratically on  $N$  and depend on a particular draw from the random matrix ensemble. If  $\mathcal{W}_N$  is not self-averaging, we will not be able to compute any approximation to  $\langle \mathcal{W}_N \rangle$  with presently known methods. However, an observable that is not self-averaging might still have a nonzero average value in a random matrix ensemble (see [17] for examples), and it might be possible to compute this from the gravitational path integral. In that case, the expression for  $\langle \mathcal{W}_N \rangle$  that the gravitational path integral would compute would really be an average value, averaged over nearby values of  $N$ .

Now consider several observables  $\mathcal{W}_{N,k}$ ,  $k = 1, \dots, s$  that are all functions of the pseudorandom matrix  $H_N$ . Whether or not individually they have nonzero averages, the connected correlation function  $\langle \mathcal{W}_{N,1} \mathcal{W}_{N,2} \dots \mathcal{W}_{N,k} \rangle_c$  may have a nonzero average in random matrix theory, in which case we may be able to compute this average from the gravitational path integral. Let us focus on the special case  $\mathcal{W}_{N,k} = \text{Tr} e^{-\beta_k H_N} R_k$ , for some  $\beta_k, R_k$ . In this special case,  $\mathcal{W}_{N,k}$  is a partition function  $Z_N(\beta_k, R_k)$  on a manifold  $S^1 \times S^{d-1}$ . Here if  $\beta_k$  is real, it is the circumference of  $S^1$ ; it is also interesting to analytically continue these

observables to complex  $\beta$ , as in [17]. The group element  $R_k \in G$  determines a holonomy around the  $S^1$  factor; this holonomy consists of a rotation of  $S^{d-1}$  and/or an internal symmetry. From what we have just said, the gravitational path integral with known methods may be able to calculate an averaged value of the connected correlation function

$$\langle Z_N(\beta_1, R_1) Z_N(\beta_2, R_2) \cdots Z_N(\beta_k, R_k) \rangle_c. \quad (4.1)$$

How the gravitational path integral would calculate this function, or more precisely an approximation to it with a smooth dependence on  $N$ , is not immediately clear just from the hypothesis that the  $H_N$  are independent pseudorandom matrices. But using everything we know about path integrals and quantum gravity, the obvious hypothesis is that  $\langle Z_N(\beta_1, R_1) Z_N(\beta_2, R_2) \cdots Z_N(\beta_k, R_k) \rangle_c$  should be computed from a path integral with a connected bulk and a boundary that is the disjoint union of  $k$  copies of  $S^1 \times S^{d-1}$ .

This is a plausible interpretation of CADB amplitudes for the special case that the boundary is a union of copies of<sup>14</sup>  $S^1 \times S^{d-1}$ . If it is correct, then presumably something similar must be true for CADB amplitudes with the  $k$  copies of  $S^1 \times S^{d-1}$  replaced by more general  $d$ -manifolds. The rough idea must be that the CFT at a specific large value of  $N$ , though actually it is a definite CFT (dependent in some cases on a few known parameters), looks, if one only has access to asymptotic expansions near  $N = \infty$ , like a pseudorandom solution of the axioms<sup>15</sup> of CFT. Then one would repeat everything we have said so far with the assertion that the  $H_N$  for different  $N$  are independent pseudorandom matrices replaced by the statement that the CFT's for different  $N$  are independent pseudorandom draws from a family of asymptotic solutions of CFT axioms.

Since it is not believed that an ensemble of CFT's with the appropriate properties actually exists, the idea here is really that the ensemble of random solutions of CFT axioms from which a given large  $N$  CFT appears to be drawn only exists in an asymptotic sense, for large  $N$ . A rough analogy is that in low energy effective field theory, the  $S$ -matrix of a relativistic quantum field theory appears to be a special case of a family of unitary, relativistic  $S$ -matrices that can be obtained by giving arbitrary coefficients to all possible parameters in the low energy effective action. It is generally believed that the generic element of this family exists only as an asymptotic expansion at low energies.

Thus, our proposal can be stated as follows. The CFT's that govern black hole states for different large values of  $N$  look, in simple measurements, like (nearly) independent pseudorandom draws from a “swampland” of effective CFT's that are defined asymptotically for large  $N$  and cannot be completed to true theories at integer values of  $N$ . This

---

<sup>14</sup>Considering this example first made possible a description in terms of  $H_N$  only, which was helpful, because random matrix theory is on a much clearer footing than random CFT, which we require in a more general case. But unfortunately, this example is actually inconvenient from a different point of view, because  $S^1 \times S^{d-1}$  has positive Ricci scalar. In any dimension, the boundary of an asymptotically AdS solution of Einstein's equations, if not connected, does not contain any component of positive Ricci scalar [2]. So a bulk computation of the observables in eqn. (4.1) has to rely on contributions that are less well understood, perhaps somewhat along the lines of [12].

<sup>15</sup>For example, an important axiom that contains much of the content of CFT is a quadratic “crossing” equation satisfied by the trilinear couplings  $\langle \mathcal{O}_i \mathcal{O}_j \mathcal{O}_k \rangle$ . This relation is found by comparing different ways to analyze a four-point function  $\langle \mathcal{O}_i \mathcal{O}_j \mathcal{O}_k \mathcal{O}_l \rangle$ .

CFT “swampland” would be analogous to the usual “swampland” of low energy effective field theories that are believed not to have ultraviolet completions [30]. The gravitational path integral, with known methods, calculates averages over the pseudorandom CFT’s with neighboring values of  $N$ .

Finally, we should point out that in the context of AdS/CFT duality, it is not true that the CFT’s for different  $N$  are truly independent above the black hole threshold. That is because (in known examples) the theories with different values of  $N$  are unified in string/M-theory and are connected by domain walls. We will illustrate this point with a simple example that generalizes the Fuchsian manifold that was introduced in eqn. (2.25). Let  $M$  be a compact hyperbolic  $d$ -manifold with metric  $d\Omega^2$  and set  $X = M \times \mathbb{R}$ . On  $X$  there is a complete hyperbolic metric

$$ds^2 = dt^2 + \cosh^2 t d\Omega^2. \quad (4.2)$$

The conformal boundary of  $X$  consists of two copies of  $M$ , at  $t = \pm\infty$ . Let  $U$  be the submanifold of  $X$  defined by  $t = 0$ . Then  $U$  is a minimal submanifold,<sup>16</sup> so there is a classical solution in which a brane is placed on  $U$ . If this brane is of the appropriate type, the integer that characterizes the CFT (or one of those integers in the case of a CFT that depends on multiple integers) will jump from  $N$  to  $N + 1$  in crossing  $U$ . Thus a path integral on  $X$  in the presence of this brane generates a connected correlation function between partition functions with different values of  $N$  on the same manifold  $M$ :

$$\langle Z_N(M) Z_{N+1}(M) \rangle_c \neq 0. \quad (4.3)$$

So the pseudorandom matrices or CFT’s for different values of  $N$  are not truly independent. However, they are nearly independent, in the sense that

$$|\langle Z_N(M) Z_{N+1}(M) \rangle_c| \ll \sqrt{|\langle Z_N(M) Z_N(M) \rangle_c| |\langle Z_{N+1}(M) Z_{N+1}(M) \rangle_c|}, \quad (4.4)$$

because the brane action contributes to the left hand side and not to the right hand side. Hopefully this is enough to justify the explanation of CADB amplitudes based on pseudorandomness. Still, the existence of correlations between the theories for different values of  $N$  seems to mean that the Hamiltonians and CFT’s of different  $N$  are not truly independent pseudorandom objects. Perhaps corrections involving branes lead to exponentially small departures from what one would expect based on independent draws from a random ensemble.

One may summarize what we have said as a proposal that ensemble averaging in gravity is averaging over nearby values of  $N$  to produce smooth approximations that can be computed by the gravitational path integral with known methods. That obviously leaves the question of what kind of path integral or what new method is needed, at least in principle, to describe the non-smooth contributions. There have been several papers aiming to find simple models of how this can work [31, 32].

---

<sup>16</sup>In  $D = 3$ ,  $U$  is the convex core of  $X$ ; see footnote 12.

*Acknowledgements* Research of JMS supported in part by FNR Grant O20/14766753. Research of EW supported in part by NSF Grant PHY-1911298. JMS thanks Ian Agol, Martin Bridgeman and Ken Bromberg for useful remarks and references. EW thanks L. Takhtajan and Jinsung Park for explanations about the renormalized volume, and K. Krasnov for helpful advice.

## A Mathematical details

This appendix contains detailed proofs of Proposition 3.1 and of Theorem 3.2, which were already explained heuristically in Section 3, as well as Theorem A.4, which slightly improves on Proposition 3.1. In addition, we provide two results which help better understand the properties of the renormalized volume under pinching of compressible curves.<sup>17</sup>

- Theorem A.10, which shows that the renormalized volume associated to the hyperbolic metric at infinity, denoted by  $V_R$  here, is within a bounded constant, depending only on the topology of the boundary, from the renormalized volume associated to the Thurston metric at infinity, denoted here by  $V'_R$ . Note that  $V'_R$  is equal to the volume of the convex core minus one fourth of the length of the measured bending lamination on its boundary.
- Theorem A.15, which gives the first term in the asymptotic development of the renormalized volume when a compressible curve is pinched.

The arguments are quite elementary but based on recent developments in the study of the renormalized volume of hyperbolic manifolds, which has recently been a focus of some interest among hyperbolic geometers. The renormalized volume was found to have close relations to topics of interest in geometry, and to be a useful or promising tool for well-established mathematical questions. We list here some of those developments.

A first motivation stemmed from the identification in [33, 34] between the renormalized volume of (some) hyperbolic manifolds and the Liouville functional studied for instance in [35, 36].

Another connection was made in [37–39] between the renormalized volume and the volume of the convex core of convex cocompact hyperbolic manifolds. This relationship was then used for instance in [40], to relate the entropy of pseudo-Anosov diffeomorphisms to their hyperbolic volume of their mapping torus, in [41, 42] to study the symplectic structure on moduli spaces of quasi-Fuchsian manifolds, and in [43] to study the metric geometry of moduli space (such as its inradius or systole). In addition, geometric properties of the renormalized volume were investigated, such as its convexity at the critical points [44, 45] and continuity under geometric limits [46, 47].

It was proved in [48] that the renormalized volume of almost-Fuchsian manifolds (quasi-Fuchsian manifolds containing a closed minimal surface with principal curvatures less than 1) is non-negative, a result that was then extended to quasi-Fuchsian hyperbolic manifolds [49] and more generally convex co-compact manifolds with incompressible boundary [50].

---

<sup>17</sup>Those to results were not contained in the first arxiv version.

In contrast, the renormalized volume of hyperbolic manifolds with compressible boundary can be negative – this remark, which plays a key role here, already appeared e.g. in [51, 52].

A particularly active current direction of research concerns the Weil-Petersson gradient flow of the renormalized volume [49, 50, 53], considered as a tool to understand the structure of 3-dimensional hyperbolic manifolds.

The properties of the renormalized volume for Schottky manifolds are considered specifically in [52], in view of the comparison of volumes of quasi-Fuchsian and Schottky manifolds with a given conformal boundary.

Since this section is geared towards more mathematical arguments, we use a slightly different notation than in the previous sections. We will always consider the hyperbolic space  $\mathbb{H}^3$  of constant sectional curvature  $-1$ , which is equivalent to setting  $\ell = 1$ .

### A.1 Convex co-compact hyperbolic manifolds

Before entering the arguments, it is useful to clarify some definitions.

We consider here a complete hyperbolic structure  $g$  on an oriented 3-dimensional manifold  $X$ , which will always be the interior of a compact manifold with boundary. Such a hyperbolic structure is the quotient of the 3-dimensional hyperbolic space  $\mathbb{H}^3$  by  $\rho(\pi_1 X)$ , where  $\rho : \pi_1 X \rightarrow PSL(2, \mathbb{C})$  is the holonomy representation of  $(X, g)$ .

The boundary at infinity of  $\mathbb{H}^3$  can be identified with  $\mathbb{CP}^1$ , and it is tempting to consider  $\rho$  as an action of  $\pi_1 X$  on  $\partial_\infty \mathbb{H}^3 = \mathbb{CP}^1$ . However, this action on  $\mathbb{CP}^1$  is not properly discontinuous, so that one cannot take the quotient. To avoid this issue, one needs to “remove” from  $\mathbb{CP}^1$  the *limit set*  $\Lambda_\rho$  of  $\rho$ , defined as the intersection with  $\partial_\infty \mathbb{H}^3$  of the closure in  $\mathbb{H}^3 \cup \partial_\infty \mathbb{H}^3$  of the orbit  $\rho(\pi_1 X)(x)$  of any point  $x \in \mathbb{H}^3$ . It turns out (see Section A.2) that  $\rho$  acts properly discontinuously on  $\mathbb{CP}^1 \setminus \Lambda_\rho$ .

We say that the subgroup  $\rho(\pi_1 X)$  is *elementary* if its limit set has at most 2 points.

A hyperbolic manifold is *convex co-compact* if

- its holonomy representation acts co-compactly (i.e. with compact quotient) on a convex domain in  $\mathbb{H}^3$ , and
- the image of its fundamental group in  $PSL(2, \mathbb{C})$  is non-elementary.

In other terms, it is the quotient of  $\mathbb{H}^3$  by a non-elementary subgroup of  $PSL(2, \mathbb{C})$ , which contains a non-empty compact geodesically convex subset.<sup>18</sup>

**Definition A.1.** *Let  $X$  be a hyperbolic manifold. A subset  $K \subset X$  is geodesically convex if any geodesic segment of  $X$  with endpoints in  $K$  is contained in  $K$ .*

Note that geodesic convexity is a strong property, for instance a small ball in a complete hyperbolic manifold with non-trivial fundamental group is not geodesically convex. In fact, if  $K$  is a non-empty geodesically convex subset of  $X$ , then the inclusion of  $K$  in  $X$  is a homotopy equivalence, see Section A.12.

<sup>18</sup>The term “convex co-compact” is perhaps a bit misleading. What can properly be called convex co-compact is rather the holonomy representation  $\rho : \pi_1 X \rightarrow PSL(2, \mathbb{C})$ , since it acts on a convex subset (the convex hull of  $\Lambda_\rho$  in  $\mathbb{H}^3$ ) with compact quotient.

Here we will use the equivalent definition of a convex co-compact manifold, which is more convenient for the proofs.

**Definition A.2.** *A convex co-compact hyperbolic structure on a manifold  $X$  is a complete hyperbolic structure for which  $X$  contains a non-empty, compact, geodesically convex subset  $K$ , and such that  $X$  is not topologically a ball or a solid torus.*

We exclude from the definition the case where  $X$  is a ball or a torus, which correspond to elementary group actions. Therefore, a complete hyperbolic manifold which contains a compact, non-empty, geodesically convex subset can be either  $\mathbb{H}^3$ , a solid torus, or a convex co-compact manifold as defined here.

For a hyperbolic manifold, being convex co-compact is equivalent to being conformally compact, that is, to having a Riemannian metric that can be written as  $g = \bar{g}/\rho^2$ , where  $\bar{g}$  is a Riemannian metric which is smooth on  $\bar{X}$  up to the boundary, while  $\rho : \bar{X} \rightarrow \mathbb{R}_{\geq 0}$  is a smooth function that vanishes on the boundary, with  $\|d\rho\|_{\bar{g}} = 1$  on  $\partial X$ . Indeed:

- If  $(X, g)$  is conformally compact, a direct computation shows that the surfaces

$$S_\epsilon = \{x \in X \mid \rho(x) = \epsilon\}$$

are locally convex for  $\epsilon > 0$  small enough. This implies that the (compact) set

$$X_\epsilon = \{x \in X \mid \rho(x) \geq \epsilon\}$$

is geodesically convex for  $\epsilon > 0$  small enough. Indeed, a geodesic segment with endpoints in  $X_\epsilon$  must stay in  $X_\epsilon$  since otherwise, at the point where  $\rho$  achieves its minimum  $\rho_{min}$ , it would need to be tangent to  $S_{\rho_{min}}$  on the convex side, a contradiction.

- Conversely, if  $X$  is convex co-compact, it contains a geodesically convex subset  $K$  which is compact. Replacing if necessary  $K$  by an  $r$ -neighborhood and smoothing its boundary, we can assume that  $K$  has smooth boundary. If  $r : X \rightarrow \mathbb{R}_{\geq 0}$  is defined as the distance to  $K$ , the function  $\rho = e^{-r}$  is a defining function and  $(X, g)$  is conformally compact.

## A.2 The complex structure at infinity

The set  $\Omega_\rho = \partial_\infty \mathbb{H}^3 \setminus \Lambda_\rho$  is called the *discontinuity domain* of  $\rho$ . Since  $\rho$  acts by hyperbolic isometries on  $\mathbb{H}^3$ , it acts by complex transformations on  $\Omega_\rho$ , and it can be proved that this action is properly discontinuous, see [54, Sections 8.1 and 8.2]. The quotient  $\partial_\infty X = \Omega_\rho/\rho(\pi_1 X)$  is therefore equipped with a complex structures, which will be denoted by  $c$  here.

By a series of results of Ahlfors, Bers, Kra, Marden, Maskit, Sullivan and Thurston, a convex co-compact hyperbolic metric on  $X$  is uniquely determined by  $c$ , considered as a point in the Teichmüller space  $\mathcal{T}_{\partial X}$  of  $\partial X$ . If  $X$  has incompressible boundary, then this map from  $\mathcal{T}_{\partial X}$  to the moduli space of convex co-compact hyperbolic metrics is one-to-one. However, if  $X$  has compressible boundary, two points in  $\mathcal{T}_{\partial X}$  can determine the same

convex co-compact structure on  $X$ . This happens when one is the image of the other by an (isotopy class of) homeomorphism which extends over the manifold – for instance, a homeomorphism corresponding to a Dehn twist along a compressible simple closed curve (a curve in  $\partial X$  which bounds a disk in  $X$ ).

As a consequence, the space of convex co-compact hyperbolic structures on  $X$  is parameterized in  $\mathcal{T}_{\partial X}/\Gamma$ , where  $\Gamma$  is the group of isotopy classes of  $X \cup \partial X$  which are homotopic to the identity, see [55, Section 3].

Note that  $\partial_{\infty}X = \Omega_{\rho}/\rho(\pi_1 X)$  is equipped with more than a complex structure: each point has a neighborhood that can be identified with a domain in  $\mathbb{C}\mathbb{P}^1$ , and this identification is well-defined up to elements of  $PSL(2, \mathbb{C})$ .<sup>19</sup> The existence of those local charts in  $\mathbb{C}\mathbb{P}^1$  will be relevant in Section A.7.

### A.3 Measured laminations on surfaces

Measured laminations play a significant role in the arguments below, so we provide here a brief introduction to their definition and key properties. Measured laminations occur in the next section when describing the geometric structure on the boundary of the convex core of a convex co-compact hyperbolic manifold.

Let  $M$  be a closed surface, equipped with a hyperbolic metric  $h$  – one can consider more generally complete hyperbolic surfaces of finite area (or even, with some adaptations, of infinite volume). A geodesic lamination is then defined as a closed subset of  $M$  which is a disjoint union of complete geodesics. A *measured* geodesic lamination is a geodesic lamination equipped with a transverse measure, that is, each transverse curve is equipped with a measure, and this measure does not change when the curve is moved while keeping the intersection with the lamination transverse, see [56, Section 10].

The simplest case of geodesic laminations is the disjoint union of a finite family  $c_1, \dots, c_n$  of disjoint closed geodesics. A transverse measure is then defined simply as a positive weight  $w_i$  associated to each geodesic  $c_i$ , yielding a *weighted multicurve*. There is a natural topology on the space of weighted multicurves, where two weighted multicurves are close if they have a similar intersection with any transverse closed curve on  $M$ . The space of measured geodesic laminations can be defined as the completion of the space of weighted multicurves for this topology. It follows that weighted multicurves (and in fact even weighted closed geodesics) are dense in the space of measured geodesic laminations.

However, “generic” geodesic laminations can be more complex than weighted multicurves. While their support has Hausdorff dimension equal to 1 [57], a short transverse segment might have an uncountable set of intersections, none of which has an atomic weight.

Given a hyperbolic metric  $m$  on  $M$  and a measured lamination  $l$ , Thurston (see e.g. [58, Section 2]) proved that one can define the *hyperbolic length* of  $l$  for  $m$ , denoted here by  $L_m(l)$ . It is defined as the limit of the hyperbolic lengths of a sequence of weighted multicurves  $((c_i^n, w_i^n)_{i=1, \dots, k_n})$  converging to  $l$ , where the hyperbolic length of a weighted

---

<sup>19</sup>This can be formalized as the existence on  $\partial_{\infty}X$  of a *complex projective structure*, but this point of view will not be necessary here.

multicurve  $(w_i, c_i)_{i=1, \dots, k}$  is defined as

$$L_m((w_i, c_i)_{i=1, \dots, k}) = \sum_{i=1}^k w_i L_m(c_i) .$$

The notion of measured lamination does not in fact require a hyperbolic metric, and can be defined “topologically”. Any measured lamination on a closed surface  $M$  can then be realized uniquely as a geodesic measured lamination for any hyperbolic metric  $m$  on  $M$ , much like any closed curve and be realized uniquely as a geodesic.

#### A.4 The convex core and the geometry of its boundary

The arguments presented here rely heavily on the relations between the renormalized volume and the volume of the *convex core* of convex co-compact hyperbolic manifolds, as well as on the geometry of the boundary of this convex core.

It follows from the definition that, in a hyperbolic manifold, the intersection of two closed, geodesically convex subsets is still geodesically convex. In addition it can be shown (see Section A.12) that, if the manifold has non-trivial topology, then the intersection is non-empty. As a consequence, any convex co-compact hyperbolic manifold contains a smallest non-empty, closed, geodesically convex subset, called its convex core, which is compact. We denote it here by  $C(X)$ .

The convex core has finite volume, denoted here by  $V_C(X)$ . This volume is closely related to the renormalized volume – when the boundary is incompressible, there is a bound on the difference between the two, which only depends on the topology of the boundary, see Section A.8 below.

The geometry of the boundary of the convex core was analysed by Thurston (see [54, Section 8.5]). Since the convex core is a minimal geodesically convex subset, its boundary cannot have extremal points, so it is a locally convex *pleated surface* – it is the union of a finite set of totally geodesic ideal triangles intersecting along their boundary, and as a consequence the induced metric on the boundary is hyperbolic (of constant curvature  $-1$ ). We will denote this hyperbolic metric on  $\partial C(X)$  by  $m$ . (Note that the ideal triangles mentioned here generally “wrap” around the boundary of  $C(X)$ , so they appear more clearly in the universal cover of  $X$ , where their vertices can be identified as points of the limit set of  $X$ .)

However the pleating locus – the set of points which do not have a neighborhood which is totally geodesic – can be quite complicated. It is a geodesic lamination, as seen in Section A.3. Moreover the pleating locus is equipped with a transverse measure, which records the amount of “bending” along each geodesic. In the simplest cases where the bending locus is a disjoint union of closed geodesics, each of the closed geodesics is equipped with a positive weight, which records the exterior angle of the boundary of the convex core along this “edge”.

In generic examples, no geodesic in the pleating locus has an atomic weight. However simple closed geodesics equipped with a positive weight are dense, in a suitable topology (see Section A.3) in the space of measured laminations. Heuristically, it is therefore often

sufficient to think of (arbitrarily long and complicated) weighted simple closed curves, rather than of generic measured laminations.

Bonahon and Otal [59] gave a complete description of the measured laminations that can arise in this manner on the boundary  $\partial X$ . The conditions are that each closed curve has weight at most  $\pi$ , that for any essential annulus  $A$  or Möbius band in  $X$  (with boundary in  $\partial X$ ),  $\partial A$  has positive intersection with  $l$ , and that for any essential disk  $D$  (with boundary in  $\partial X$ ),  $\partial D$  has intersection larger than  $2\pi$  with  $l$ .

### A.5 Relations between the boundary of the convex core and the boundary at infinity

Given a convex co-compact hyperbolic manifold  $X$ , let  $N\partial C(X)$  be the unit normal bundle of  $\partial C(X)$ , that is, the set of outwards-pointing unit vectors normal to a support plane of  $C(X)$  at a boundary point. For each  $n \in N\partial C(X)$ , let  $g_n : \mathbb{R}_{\geq 0} \rightarrow X$  be the geodesic ray determined by the initial vector  $n$ . It can be proved (see Section A.12) that the images of the  $g_n$  are all disjoint (and do not intersect  $C(X)$  except at their origin). As a consequence, the map sending  $n$  to the endpoint at infinity of  $g_n$  defines a homeomorphism between  $N\partial C(X)$  and  $\partial_\infty X$ .

Moreover, when  $X$  is not Fuchsian<sup>20</sup>, this homeomorphism can be deformed, by “smoothing out” the bending locus, to a (non-canonical) homeomorphism between  $\partial C(X)$  and  $\partial_\infty X$ . (If  $X$  is Fuchsian, the convex core is a smooth surface, and no smoothing is necessary.) We will often use this homeomorphism implicitly below, and identify the two surfaces. (If  $X$  is Fuchsian, then  $\partial_\infty X$  is homeomorphic to the disjoint union of two copies of  $C(X)$ , which in this case is a totally geodesic surface, so that  $N\partial C(X)$  is a double cover of  $C(X)$ .)

The boundary at infinity  $\partial_\infty X$  is equipped with a conformal structure, and therefore with a hyperbolic metric provided by the Poincaré-Riemann uniformization theorem – it is called the Poincaré metric, and denoted here by  $h$ . It can be compared to the induced metric  $m$  on the boundary of the convex core.

- A simple closed curve is short for  $h$  if and only if it is short in  $m$ . Specifically, for all  $\epsilon > 0$  small enough, there exists  $\epsilon' > 0$  such that if a simple closed curve  $\gamma$  has length less than  $\epsilon'$  for  $h$ , then it has length less than  $\epsilon$  for  $m$ , and conversely (with a different value of  $\epsilon'$ ), see [60, 61].
- Curves which do not enter the “thin” part of  $\partial X$  for either  $m$  or  $l$  (the subset composed of points where the injectivity radius is smaller than a fixed constant  $\epsilon_0$  – for  $\epsilon_0$  small enough, this thin part is the disjoint union of long and “thin” annuli, each associated to a simple closed geodesic of length less than  $\epsilon_0$ ) have lengths for  $m$  and for  $l$  which are comparable, up to bounded multiplicative constants, see [62].

---

<sup>20</sup>A Fuchsian manifold is the quotient of  $\mathbb{H}^3$  by a surface group acting properly discontinuously and cocompactly on a totally geodesic plane. So the convex core of a Fuchsian manifold is a totally geodesic surface, while for any other convex co-compact manifold it is a 3-dimensional domain with positive volume. See also Section 2.3, and specifically footnote 9.

## A.6 The maximal set of contractible curves in the boundary

The proof of Proposition 3.1, presented in Section 3, is based on a surgery that can be applied on convex co-compact hyperbolic structures when the length of a compressible curve (in the Poincaré metric at infinity) is sufficiently short. Similar ideas will be used again below in Section A.10 when considering the limit of the volume of the convex core when a finite set of disjoint closed compressible curves in the boundary is pinched.

We consider a convex co-compact hyperbolic manifold  $X$ , a connected component  $M$  of  $\partial C(X)$ , and a simple closed curve  $\gamma$  in  $M$  which is compressible, i.e., bounds a disk in  $X$ . We first deform  $X$  so that the geodesic representative of  $\gamma$  is pinched to have length less than  $\epsilon$  (for a value of  $\epsilon$  that will be made more precise below) for the induced metric  $m$  on  $\partial C(X)$ . Then the geodesic representative  $\gamma_0$  of  $\gamma$  in  $(M, m)$  is the “center” of a long collar, of width  $w(\epsilon)$  arbitrary large if  $\epsilon$  is small.

Let  $\Delta$  be a complete geodesic in  $C(X)$  intersecting the disk  $D$  in  $C(X)$  bounded by  $\gamma_0$ , and let  $P$  be a totally geodesic plane in  $X$  orthogonal to  $\Delta$  and intersecting  $\gamma_0$ . (Note that  $P$  is not entirely determined by those conditions, since  $\gamma_0$  is not necessarily contained in a plane orthogonal to  $\Delta$ .)

The following elementary lemma is used in Section 3.

**Lemma A.3.**  *$P$  is embedded in  $X$ , that is, it has no self-intersection.*

*Proof.* Assume the opposite, it would mean that there are two distinct lifts  $P_1$  and  $P_2$  of  $P$  to  $\tilde{X} \simeq \mathbb{H}^3$  which intersect at a point  $x$ . Let  $\gamma_1$  and  $\gamma_2$  be the corresponding lifts of  $\gamma_0$ , and let  $x_1 \in \gamma_1 \cap P_1$ ,  $x_2 \in \gamma_2 \cap P_2$ . Let  $\alpha_1$  be the geodesic segment connecting  $x_1$  to  $x$  (which is in  $P_1$ ) and let  $\alpha_2$  be the geodesic segment connecting  $x_2$  to  $x$ . Finally, let  $\beta$  be the geodesic segment connecting  $x_1$  to  $x_2$ .

Since  $C(X)$  is geodesically convex, the segment  $\beta$  is contained in its universal cover  $\widetilde{C(X)}$ , and is therefore almost orthogonal (for  $\epsilon$  small enough) to both  $P_1$  and  $P_2$ , and also to  $\alpha_1$  and  $\alpha_2$ . So if we call  $\theta_1$  (resp.  $\theta_2$ ) the angle between  $\beta$  and  $\alpha_1$  (resp.  $\alpha_2$ ) then both are close to  $\pi/2$  as  $\epsilon \rightarrow 0$ . Moreover,  $\beta$  has length at least  $2w(\epsilon)$ , which goes to infinity as  $\epsilon \rightarrow 0$ . However, a standard hyperbolic triangle formula, applied to the triangle with edges  $\alpha_1, \alpha_2$  and  $\beta$ , ensures that

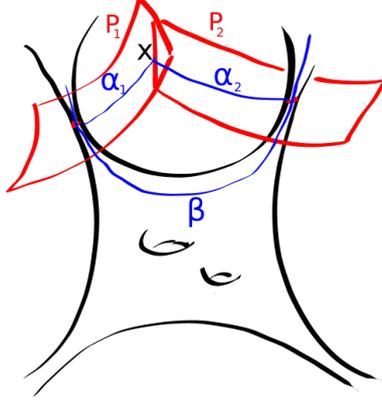
$$\cos(\theta) = -\cos(\theta_1)\cos(\theta_2) + \cosh(L(\beta))\sin(\theta_1)\sin(\theta_2),$$

where  $\theta$  is the angle between  $\alpha_1$  and  $\alpha_2$  at  $x$ .

For  $\epsilon$  small enough, this cannot hold since the two sine terms on the right are close to 1 and  $\cosh(L(\beta))$  is large.  $\square$

Note that the same argument shows that if we consider two totally geodesic planes  $P$  and  $P'$  corresponding to two short contractible curves  $\gamma$  and  $\gamma'$ , then  $P$  and  $P'$  are disjoint – this is used in Section A.10.

We are now equipped to prove the following small improvement of Proposition 3.1. The proof repeats the proof of Proposition 3.1 given in Section 3 with more attention given to surfaces of genus 2.



**Figure 4.** Planes almost orthogonal to thin tubes cannot intersect.

**Theorem A.4.** *Let  $X$  be a convex co-compact hyperbolic manifold, and let  $M$  be a boundary component of  $X$ . If  $M$  has genus 2 and contains at least one compressible curve, or if  $M$  has genus  $g \geq 3$  and contains at least  $3g - 6$  disjoint, non-homotopic compressible curves, then  $X$  is a Schottky manifold and  $\partial X = M$ .*

*Proof.* The first step is to analyse more carefully what happens to a family of independent curves when a surface  $M$  is cut along one of them, say  $c$ . Different cases can be distinguished, see Figure 3.

1.  $c$  separates a genus 1 surface  $M_1$  from a genus  $g'$  surface  $M_{g'}$ ,  $g' \geq 2$ . Then at most one pair of curves on  $M$  can become homotopic on  $M_{g'}$ .
2.  $c$  separates a surface  $M_{g'}$  from a surface  $M_{g''}$ , with  $g' \geq 2, g'' \geq 2$ . Then at most one pair of non-homotopic curves *on each side of  $c$*  can become homotopic.
3.  $c$  is non-separating. Then at most one curve can become homotopically trivial, while at most one pair of non-homotopic curves can be homotopic in the complement of  $M$ .

The proof proceeds by induction on the genus of  $M$ , as for the proof of Proposition 3.1. The statement is already known, from Section 3, when  $M$  has genus 1 or genus 2. We assume that it is true for genus at most  $g - 1$ , and consider a boundary component  $M$  of genus  $g$ . We assume that  $M$  contains  $3g - 6$  independent compressible curves, and choose one of those curves, say  $c$ . We now consider different cases.

- If  $c$  splits  $M$  into two surfaces  $M', M''$  of genus  $g' \geq 3$  and  $g'' \geq 3$ . The argument is then exactly as seen in the proof of Proposition 3.1:  $g' + g'' = g$ , the number of remaining independent compressible curves is at least  $3g - 9 = 3(g' + g'') - 9$ , but  $M'$  can have at most  $3g' - 3$  while  $M''$  can have at most  $3g'' - 3$ . It follows that  $M'$  has at least  $3g' - 6$  independent compressible curves, while  $M''$  has at least  $3g'' - 6$ . So both are boundaries of Schottky manifolds by the induction hypothesis, and the surgery done on the interiors means that  $M$  also bounds a Schottky manifold.

- Similarly, the same argument as in the proof of Proposition 3.1 can be applied if  $c$  is non-separating and  $M$  has genus at least 4.
- Suppose now that  $c$  splits  $M$  into  $M'$  of genus  $g' = 2$  and  $M''$  of genus  $g'' \geq 3$ . If  $M'$  contains at least one compressible curve, then the same argument as before shows that  $M''$  contains at least  $3g'' - 6$  independent compressible curves, so by induction  $M''$  is the boundary of a Schottky manifold, and therefore  $M$  is the boundary of a Schottky manifold.

If  $M'$  contained no compressible curve, then  $M''$  would have at least  $3g - 8$  independent compressible curves, since cutting along  $c$  can only “destroy”  $c$ , plus one compressible curve on the side of  $M''$  (two compressible curves might become homotopic). But  $3g - 8 = 3g'' - 2$ , a contradiction because  $M''$  cannot have more than  $3g'' - 3$  independent compressible curves.

- If  $M$  has genus 4 and at least 6 independent compressible curves, and is cut by  $c$  into two surfaces of genus 2, the same argument shows that there must be at least one compressible curve remaining on each side. Otherwise at most 2 curves would be “lost” ( $c$  plus one pair of compressible curves becoming homotopic) so one side would need to have 4 independent compressible curves, a contradiction.
- Finally suppose that  $M$  has genus 3 and at least 3 compressible curves, and is cut along a non-separating compressible curve  $c$ . One obtains a surface of genus 2 containing at least one compressible curve, because at most one compressible curve could become nul-homotopic. So the surface obtained after cutting along  $c$  is the boundary of a Schottky manifold, and therefore  $X$  is Schottky.

□

## A.7 A variational formula for the renormalized volume

Although convex co-compact hyperbolic manifolds have infinite volume, one can define a *renormalized volume*, see [63]. We will need a variational formula for the renormalized volume, see [35–37] or [39, Corollary 3.11]. To state it, note that  $\partial_\infty X$  is equipped with a holomorphic quadratic differential  $q$ , which can be defined as follows. Let  $\partial_i X$  be a connected component of  $\partial_\infty X$ . Since  $\partial_i X$  is a closed surface of genus at least 2, equipped with a complex structure (the restriction of  $c$  to  $\partial_i X$ ), there is by the Poincaré-Riemann uniformization theorem a holomorphic map  $f_i : \mathbb{D} \rightarrow \partial_i X$ , where  $\mathbb{D}$  is the unit disk in  $\mathbb{C}$ .

Note that  $\partial_\infty X$  can be locally identified with  $\mathbb{CP}^1$  (or  $\mathbb{C}$ , using the holomorphic identification of  $\mathbb{C}$  to  $\mathbb{CP}^1 \setminus \{\infty\}$ ), and this local identification is well-defined up to left composition by an element of  $PSL(2, \mathbb{C})$ . This makes it possible to consider the Schwarzian derivative of  $f_i$ . Recall that given a holomorphic map  $f : \Omega \rightarrow \mathbb{C}$ , where  $\Omega \subset \mathbb{C}$ , its Schwarzian derivative is defined as:

$$\mathcal{S}(f) = \left( \left( \frac{f''}{f'} \right)' - \frac{1}{2} \left( \frac{f''}{f'} \right)^2 \right) dz^2 .$$

The following properties are relevant:

- $\mathcal{S}(f) = 0$  if and only if  $f$  corresponds to an element of  $PSL(2, \mathbb{C})$ ,
- $\mathcal{S}(f \circ g) = \mathcal{S}(f) + f^*\mathcal{S}(g)$ .

It follows from those two properties that the Schwarzian derivative of  $f$  is invariant under composition of  $f$  on the left by an element of  $PSL(2, \mathbb{C})$ . As a consequence, the Schwarzian derivative of  $f_i$  is well-defined – even if the identification of  $\partial_\infty X$  with  $\mathbb{CP}^1$  is only local. We define

$$q = f_{i*}\mathcal{S}(f_i) .$$

By construction,  $q$  is a holomorphic quadratic differential on  $\partial_i X$ . <sup>21</sup>

**Lemma A.5.** *Let  $(c_t)_{t \in [0,1]}$  be a one-parameter family of complex structures in  $\mathcal{T}_{\partial X}/\Gamma$ . Then*

$$\frac{dV_R(c_t)}{dt} = \operatorname{Re} \left( \left\langle q, \frac{dc_t}{dt} \right\rangle \right) ,$$

where  $\langle, \rangle$  denotes the natural duality product between holomorphic quadratic differentials and Beltrami differentials.

### A.8 The renormalized volume and the volume of the convex core

There are close relations between the  $V_C(X)$  and  $V_R(X)$ , given in particular by the following lemma.

**Lemma A.6.** *For any convex co-compact hyperbolic manifold  $X$ ,*

$$V_C(X) - \frac{1}{4}L_m(l) - C(\partial X) \leq V_R(X) \leq V_C(X) - \frac{1}{4}L_m(l) + \frac{\pi \log(2)}{2}|\chi(\partial X)| .$$

where  $C(\partial X)$  is a constant depending only on the topology of  $\partial X$ .

(Note that the additive constant on the right of the equation depends on a choice of normalization in the definition of  $V_R$ .) The inequality on the right can be found as [39, Theorem 1.1] for quasi-Fuchsian manifolds, but the proof extends without change to convex co-compact hyperbolic manifolds, see [51, Section 3]. The inequality on the left is Theorem A.10 below.

There is also a lower bound on renormalized volume, in terms of the volume of the convex core, for convex co-compact manifolds with incompressible boundary, see [39, 51]. For convex co-compact manifolds with compressible boundary, the constant depends on the injectivity radius of the boundary.

---

<sup>21</sup>There is another way to introduce  $q$ , in relation to the second term of the asymptotic development of the metric near infinity when the metric at infinity (corresponding to the first term) has constant curvature. The real part of  $q$  is then minus the traceless part of this second term, see e.g. [37, Lemma 8.3]. The holomorphic quadratic differential  $q$  already appears in this form in [23]. We will not need this different point of view here.

### A.9 A bound on the Weil-Petersson gradient

We recall here a bound on the Weil-Petersson gradient of  $V_R$  when no compressible curve is short. Similar estimates can be found in [50, 64]. We provide a proof for completeness. A more precise analysis of the convergence of the geometric structure when an incompressible curve is pinched can be found in [65].

**Lemma A.7.** *Let  $X$  be a convex co-compact hyperbolic manifold, such that the length for the Poincaré metric of any non-trivial simple closed curve in  $\partial_\infty X$  compressible in  $X$  is at least  $l$ . Then the Weil-Petersson gradient of  $V_R$  on  $\mathcal{T}_{\partial X}/\Gamma$  is bounded by*

$$\|dV_R\|_{WP} \leq \frac{3\sqrt{\pi|\chi(\partial_\infty X)|}}{\sqrt{2}\tanh^2(l/4)} .$$

*Proof.* The proof is based on a classical bound on the Schwarzian derivative at the center of a holomorphic map which is injective on a disk, and on the fact that if the shortest compressible curve in  $(\partial X, h)$  has length at least  $l$ , then every point in  $(\partial\tilde{X}, h)$  is the center of an embedded open disk of radius  $l/2$ . This second point follows from the fact that if  $x \in \partial\tilde{X}$  realizes the minimum of the injectivity radius (the radius  $r$  of the largest embedded disk centered at  $x$ ) then there is a embedded open disk of radius  $r$  centered at  $x$  with a self-tangency, and it follows that  $x$  is on a closed geodesic of length  $2r$  in  $(\partial\tilde{X}, h)$ . This closed geodesic projects to a closed compressible geodesic of length  $2r$  on  $\partial X$ .

We first note that if  $\mathbb{D}_r$  is the disk of radius  $r$  in  $\mathbb{C}$  and  $f : \mathbb{D}_r \rightarrow \mathbb{C}$  is a univalent holomorphic map, then  $\mathcal{S}(f)$  can be written as  $\sigma dz^2$ , with  $|\sigma(0)| \leq 6/r^2$ . Indeed, the function  $\bar{f} : z \mapsto f(rz)$  is then a univalent holomorphic map from  $\mathbb{D}$  to  $\mathbb{C}$ , so that by the Nehari-Kraus estimate [66], its Schwarzian differential can be written as  $\mathcal{S}(\bar{f}) = \bar{\sigma}(z)dz^2$ , with  $|\bar{\sigma}(0)| \leq 6$ . But it follows from the definition of the Schwarzian derivative that

$$\mathcal{S}(\bar{f}) = r^2 \mathcal{S}(f) .$$

Since the hyperbolic metric at 0 is  $4(dx^2 + dy^2)$ , the norm of the real part of  $\mathcal{S}(f)$  with respect to the hyperbolic metric  $h$  on  $\mathbb{D}$  is bounded (pointwise) by

$$\|\operatorname{Re}(\mathcal{S}(f)(0))\|_h \leq \frac{3\sqrt{2}}{2r^2} .$$

Now let  $x \in \partial_\infty \tilde{X}$  be a point where the injectivity radius is at least  $l/2$ . Consider the Riemann uniformization map  $f$  from  $\mathbb{D}$  to the connected component of  $\partial_\infty \tilde{X}$ , chosen so that  $f(0) = x$ . By construction,  $f$  is a local isometry between the hyperbolic metric on  $\mathbb{D}$  and the Poincaré metric on  $\partial_\infty \tilde{X}$ . Moreover, the disk of center  $x$  and radius  $l/2$  (for the Poincaré metric) is embedded in  $\partial_\infty \tilde{X}$ , so that the restriction of  $f$  to a disk of center 0 and hyperbolic radius  $l/2$  is injective. But a disk of hyperbolic radius  $l/2$  and center 0 (for the hyperbolic metric on  $\mathbb{D}$ , that is, the Poincaré disk model of the hyperbolic plane) is a disk of Euclidean radius  $r = \tanh(l/4)$ . So the norm of the real part of the Schwarzian derivative of  $f$  at  $x$  is bounded, in the hyperbolic metric, by

$$\|\operatorname{Re}(\mathcal{S}(f))\|_h \leq \frac{3\sqrt{2}}{2\tanh^2(l/4)} .$$

Integrating over  $\partial_\infty X$ , we obtain that the  $L^2$  norm of  $\text{Re}(q)$  is bounded by

$$\left( \int_S \|\text{Re}(q)\|_h^2 da_h \right)^{1/2} \leq \frac{3\sqrt{\pi|\chi(\partial_\infty X)|}}{\tanh^2(l/4)}.$$

It then follows from Lemma A.5 that

$$\|dV_R\|_{WP} \leq \frac{3\sqrt{\pi|\chi(\partial_\infty X)|}}{\sqrt{2}\tanh^2(l/4)}.$$

□

We notice for future reference that the WP estimate here could be improved, since the pointwise estimate on  $\text{Re}(q)$  is better at each point where the injectivity radius is larger than  $l/2$ .

It follows from this lemma that on any 1-parameter family of boundary complex structures  $(c_t)_{t \in [0,1]}$  in  $\mathcal{T}_{\partial X}/\Gamma$  of finite Weil-Petersson length, ending on a stratum of the Weil-Petersson completion of  $\mathcal{T}_{\partial X}/\Gamma$  corresponding to pinching a closed curve which is not compressible in  $X$ ,  $V_R$  remains bounded (see [46]). This follows from the lemma since, in this 1-parameter family, the lengths of simple, non-trivial closed curves compressible in  $X$  remains bounded from below by a positive constant.

### A.10 Convergence of convex cores when pinching compressible curves

In this section we consider a sequence of conformal structures  $(c_n)_{n \in \mathbb{N}}$  on  $\partial X$ ,  $c_n \in \mathcal{T}_{\partial X}/\Gamma$ , and denote by  $h_n$  the hyperbolic metric in the conformal class  $c_n$ . We assume that  $(c_n)_{n \in \mathbb{N}}$  converges, in the Weil-Petersson metric completion of  $\mathcal{T}_{\partial X}/\Gamma$ , to a boundary point  $c_\infty$ , where a set of disjoint simple closed curves  $\gamma_i$ ,  $1 \leq i \leq k$  is pinched, with each  $\gamma_i$  bounding a disk  $D_i$  in  $X$ . Let  $X_j$ ,  $1 \leq j \leq l$ , be the connected components of  $X \setminus (D_1 \cup \dots \cup D_k)$ .

We denote by  $g_n$  the convex co-compact hyperbolic metric on  $X$  associated to  $c_n$  by the Ahlfors-Bers theorem, and by  $X(n) = (X, g_n)$ . The limit conformal structure  $c_\infty$  determines a conformal structure on  $\partial X_j$ ,  $1 \leq j \leq l$ , marked by  $2k$  points  $\xi_1, \dots, \xi_{2k}$  corresponding to the pinching of the  $\gamma_i$ ,  $1 \leq i \leq k$ , and we denote by  $\xi(j)$  the set composed of the  $\xi_i$  which are in the boundary of  $X_j$ ,  $1 \leq j \leq l$ . (Note that if  $X$  is connected, then each of the  $X_j$  is equipped with at least one of the  $\xi_i$  on its boundary, so  $\xi(j)$  has at least one element.)

Now consider  $\epsilon > 0$  small enough so that for  $n \geq n_0$ , for a  $n_0$  large enough, the  $\gamma_i$  have length less than  $\epsilon$  for  $1 \leq i \leq k$ , while all closed curves not homotopic to a finite cover of one of the  $\gamma_i$  have length larger than  $\epsilon$ . For  $n \geq n_0$ ,  $\partial C(X(n))$  is the union of:

- $k$  tubes  $T_i(n)$ ,  $1 \leq i \leq k$ , composed of points where the injectivity radius is less than  $\epsilon/2$  – each tube having having the geodesic representative of one of the  $\gamma_i$  as its core meridian,
- $l$  connected regions  $C_j(n)$ ,  $1 \leq j \leq l$ , one for each of the  $X_j$ , composed of points where the injectivity radius is at least  $\epsilon/2$ .

The diameter of each of the  $C_j(n)$ ,  $1 \leq j \leq l$ , is uniformly bounded from above (independently of  $n$ ), because  $\partial C(X(n))$  has bounded area by the Gauss-Bonnet theorem, and the  $C_j(n)$  are connected and composed of points where the injectivity radius is at least  $\epsilon/2$ .

We choose in each of the  $C_j(n)$ ,  $1 \leq j \leq l$ , a point  $x_j(n)$ . The  $x_j(n)$  will be used as base-points below, when considering the Gromov-Hausdorff convergence of the  $X_j(n)$ . The precise choice of the  $x_j(n)$  is not important, since they are constrained to be contained in a region of bounded diameter – this will be sufficient to ensure convergence in the Gromov-Hausdorff topology pointed at  $x_j(n)$ , after extracting a sub-sequence.

**Lemma A.8.** *Under the hypothesis above, for each  $j \in \{1, 2, \dots, l\}$  (and after extracting a subsequence):*

1. *The pointed manifold  $(X(n), x_j(n))$  converges in the Gromov-Hausdorff topology on compact subsets to a complete hyperbolic manifold  $(\bar{X}_j, \bar{x}_j)$ , with  $\bar{X}_j$  diffeomorphic to  $X_j$ . ( $\bar{X}_j$  can be either convex co-compact, a solid torus, or a ball.)*
2. *The sequence of pointed convex cores  $(C(X(n)), x_j(n))$  converges to  $CH(\xi(j)) \subset \bar{X}_j$ , the convex hull of  $\xi(j)$  in  $\bar{X}_j$ .*
3.  *$V_C(X(n)) \rightarrow \sum_{j=1}^l V(CH(\xi(j)))$ , which is finite.*

Here by the *convex hull* of  $\xi(j)$  we mean the smallest geodesically convex subset of  $\bar{X}_j$  containing  $\xi(j)$  in its asymptotic boundary. By definition this subset contains the convex core of  $\bar{X}_j$ .

*Proof.* We will use the same argument as in Section A.6, and notice that there exists  $n_0 \in \mathbb{N}$  such that for  $n \geq n_0$  the lengths of all the  $\gamma_i$  is at most  $\epsilon'$ , for a fixed value of  $\epsilon' > 0$ . If  $\epsilon'$  is small enough, we can then consider a totally geodesic plane  $P_i$  as in Section A.6, that is, a plane orthogonal to one of the lines in the thin tube with core meridian  $\gamma_i$ . The planes  $P_i$  and  $P_{i'}$  are then disjoint for  $i \neq i'$ , as seen in Section A.6. We now only consider  $n \geq n_0$ .

We then let  $\Omega_j(n)$  be the connected component of  $X(n) \setminus (P_1 \cup \dots \cup P_k)$  containing  $x_j(n)$ , and let  $K_j(n) = \Omega_j(n) \cap C(X, g_n)$ . As in Section A.6, we glue a half-space to  $\Omega_j(n)$  at each of the  $P_i$  adjacent to  $\Omega_j(n)$ , and obtain in this manner a complete hyperbolic manifold  $X_j(n)$ , which is either convex co-compact, a solid torus, or a ball. Abusing notations a bit, we consider the  $P_i$  adjacent to  $\Omega_j(n)$  as disjoint, totally geodesic planes in  $X_j(n)$ .

To this surgery in the hyperbolic metric corresponds a simple surgery on the conformal structure at infinity of  $X(n)$ : a curve (which is short in say the Poincaré metric on  $\partial_\infty X(n)$ ) is cut and one side replaced by a small disk. As  $n \rightarrow \infty$ , the conformal structure at infinity obtained in this manner converges to the conformal structure  $c_\infty$  on  $\partial X_j$ . So  $X_j(n)$  converges in the Gromov-Hausdorff topology to  $\bar{X}_j$ . This proves the first point.

The closure of the convex subset  $K_j(n)$  is the convex hull in  $X_j(n)$  of the  $P_i \cap \partial\Omega_j(n)$ , which are topological disks with boundaries corresponding to the  $\gamma_i$ . This is clear because  $K_j(n)$  is geodesically convex by definition, and its boundary is a pleated surface outside of the  $P_i \cap \partial K_j(n)$ , so its closure is the minimal closed geodesically convex subset of  $X_j(n)$  containing the  $P_i \cap \partial K_j(n)$ . The boundary of  $P_i \cap \partial\Omega_j(n)$  corresponds to  $\gamma_i$  and its length

goes to 0 as  $n \rightarrow \infty$ . Moreover, for each  $i$ ,  $d(x_j(n), P_i \cap \partial\Omega_j(n)) \rightarrow \infty$  as  $n \rightarrow \infty$ , because the length of the tube  $T_i(n)$  around  $\gamma_i$  goes to infinity. As each of the  $P_i \cap \partial\Omega_j(n)$  converges to one of the points of  $\xi(j)$  as  $n \rightarrow \infty$ , we see that  $K_j(n)$  converges to the convex hull of  $\xi(j)$  in  $\overline{X}_j$ , which proves point (2).

The convergence of  $V_C(X(n))$  to the sum of the volumes of the  $CH(\xi(j))$  follows from the Gromov-Hausdorff convergence of the different components of  $X(n)$ , pointed at the  $x_j(n)$ .

To see that the volume of  $CH(\xi(j))$  is bounded, consider  $\xi \in \xi(j)$ , let  $P$  be a totally geodesic plane in  $\overline{X}_j$  separating  $\xi$  from  $C(\overline{X}_j)$  and from the other elements of  $\xi(j)$ , and let  $H$  be the half-space bounded by  $P$  containing  $\xi$  in its boundary. Then  $H \cap CH(\xi)$  is the convex hull of  $\xi$  and of a compact domain in  $P$  (the intersection of  $CH(\xi(j))$  with  $P$ ). So  $H \cap CH(\xi)$  has finite volume. Since this holds for all the  $\xi \in \xi(j)$ , we see that  $CH(\xi(j))$  can be written as the union of a finite family of subsets of finite volume – one for each element of  $\xi(j)$  – and the remaining part which is compact. So  $CH(\xi(j))$  has finite volume.  $\square$

### A.11 Limit of the renormalized volume when pinching a compressible curve

We now consider the case where  $(c_t)_{t \in [0,1]}$  pinches a curve  $\gamma$  which is contractible in  $X$ . The fact that  $V_R(c_t) \rightarrow -\infty$  as  $t \rightarrow 1$  then follow from Lemma A.6 and Lemma A.8, together with the following result of Bridgeman and Canary, see [67, Theorem 2’], and also [51, Theorem 4.2].

**Theorem A.9** (Bridgeman-Canary). *There exists constants  $P$  and  $Q$  (one can take  $P = 74$  and  $Q = 36$ ) such that if  $X$  is a convex co-compact hyperbolic manifold such  $\partial_\infty \tilde{X}$  contains a closed compressible geodesic of length  $r < 1$  in the Poincaré metric, then the length of the measured bending lamination on the boundary of the convex core is bounded from below by:*

$$L_m(l) \geq \frac{P}{r} - Q . \tag{A.1}$$

Note that the coefficient 74 above is twice that found in [51], since we consider here the length of closed contractible curves in  $\partial X$ , rather than the injectivity radius of  $\partial\tilde{X}$ .

Without getting into the precise value of the constants, we can indicate a heuristic explanation for (A.1). When  $\partial_\infty X$  contains a closed geodesic  $\gamma$  of length  $r$  for the Poincaré metric  $h$ , then it contains a collar of width approximately  $L = |\log(r)|$  around  $\gamma$ . If  $\gamma$  is contractible, the induced metric  $m$  on  $\partial C(X)$  contains a tube of length  $w$  approximately  $\exp(L) = 1/r$  around  $\gamma$ . But since  $\gamma$  is contractible, the intersection with  $\gamma$  of the measured bending lamination  $l$  (that is, the transverse measure of  $l$  evaluated on  $\gamma$ ) is at least  $2\pi$ , and in fact very close to  $2\pi$  as  $r \rightarrow 0$ . Finally, all leaves of  $l$  intersecting  $\gamma$  must cross the whole length of the tubular collar around  $\gamma$ , so must have length at least  $2w$ , so of the order of  $1/r$ . More precise arguments of this type can lead to (A.1).

### A.12 Geodesically convex subsets of a hyperbolic manifold

We first indicate why the intersection of two closed, non-empty, geodesically convex subsets  $K, K'$  in a complete hyperbolic manifold  $X \neq \mathbb{H}^3$  is non-empty. Since  $X$  has non-trivial

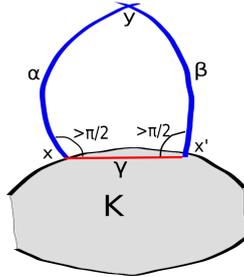
topology and is complete, it contains a closed, oriented geodesic  $\gamma$ . Let  $x \in K$ , and, for  $n \in \mathbb{N}$ , let  $\gamma_n$  be the geodesic segment starting and ending at  $x$ , and homotopic to a path going from  $x$  to a point of  $\gamma$ , doing  $n$  turns around  $\gamma$ , and going back to  $x$ .

Let  $\bar{\gamma}_n$  and  $\bar{\gamma}$  be lifts of  $\gamma_n$  and of  $\gamma$  to the universal cover of  $X$ , chosen so that the distance from the endpoints of  $\bar{\gamma}_n$  to  $\bar{\gamma}$  is equal to the distance of  $x$  to  $\gamma$  in  $X$ . For  $n$  large,  $\bar{\gamma}_n$  is a long geodesic segment with endpoints at bounded distance from  $\bar{\gamma}$ , so there is a sub-segment of  $\bar{\gamma}_n$  of length at least the length of  $\gamma$  which is arbitrarily close to  $\bar{\gamma}$ , say at distance less than  $\epsilon_n$ , for some  $\epsilon_n > 0$  with  $\lim_{n \rightarrow \infty} \epsilon_n = 0$ .

It follows that each point of  $\gamma$  is at distance at most  $\epsilon_n$  from a point of  $\gamma_n$ . Since  $K$  is geodesically convex,  $\gamma_n \subset K$ , and since  $K$  is closed and  $\epsilon_n \rightarrow 0$ ,  $\gamma \subset K$ . The same argument shows that  $\gamma \subset K'$ , and it follows that  $K \cap K' \neq \emptyset$ .

Let now  $K \subset X$  be a geodesically convex, let  $x \in \partial K$ , and let  $n$  be the outward oriented unit normal to a support planes<sup>22</sup> of  $K$  at  $x$ . Let  $\alpha$  be the half-geodesic starting from  $x$  in the direction of  $n$ . Then  $\alpha \cap K = \{x\}$ , since otherwise the whole intersection of  $\alpha$  between  $x$  and its first intersection with  $K$  would be contained in  $K$ .

If  $\beta$  is another such geodesic ray, starting from a point  $x' \in \partial K$  in the direction of a unit normal vector to a support plane of  $K$  at  $x'$ , then  $\alpha$  and  $\beta$  must be disjoint. Suppose indeed that they intersect at a point  $y$ , and let  $\gamma$  be the geodesic segment from  $x$  to  $x'$  homotomic to the union of the segment of  $\alpha$  from  $x$  to  $y$  union the segment of  $\beta$  from  $y$  to  $x'$ . The angle between  $\gamma$  and  $\alpha$  (resp.  $\gamma$  and  $\beta$ ) must be bigger than  $\pi/2$ , because both  $\alpha$  and  $\beta$  are directed by outwards unit normals of support planes of  $K$ , while  $\gamma$  is towards the interior of  $K$ . But having two angles larger than  $\pi/2$  contradicts the Gauss-Bonnet relation for hyperbolic triangles (the sum of the interior angles is equal to  $\pi$  minus the area).



**Figure 5.** A pair of geodesic rays  $\alpha, \beta$  normal to a convex subset  $K$  cannot intersect, because the sum of the angles of triangle  $xx'y$  would be greater than  $\pi$ .

It follows from this remark that, if we denote again by  $N\partial K$  the unit normal bundle

<sup>22</sup>A support plane of  $K$  at  $x$  is a totally geodesic plane containing  $x$ , which locally bounds a closed half-space containing  $K$ .

of the boundary of  $K$ , then the map

$$\begin{aligned} \exp : N\partial K \times \mathbb{R}_{>0} &\rightarrow X \\ (n, t) &\mapsto \exp(nt) \end{aligned}$$

is a diffeomorphism from  $N\partial K \times \mathbb{R}_{>0}$  to  $X \setminus K$ .

It also follows that  $K$  is homotopic to  $X$ . This applies in particular to  $C(X)$ , the convex core of  $X$ .

### A.13 The Thurston metric at infinity

Finally, this section describes briefly some properties of a natural metric at infinity of convex co-compact hyperbolic manifolds, which appears prominently in Sections A.14 and A.15.

The boundary at infinity  $\partial_\infty X$  of a convex co-compact hyperbolic manifold is equipped naturally with a Riemannian metric in the standard conformal class, called the Thurston metric (or projective metric, or sometimes the grafting metric), closely related to the boundary of  $C(X)$ . We denote it by  $h_{Th}$ , and give three descriptions of it.

The first description is simpler when the measured bending lamination  $l$  on  $\partial C(X)$  is along disjoint closed geodesics  $c_1, \dots, c_n$ , with each  $c_i$  equipped with a positive weight  $w_i$ . In this case,  $h_{Th}$  is obtained by cutting  $(M, m)$  along the geodesics  $c_i$  and replacing (or “grafting”)  $c_i$  by a flat strip of width  $w_i$ .

A second description is as the natural metric induced on  $N\partial C(X)$ , the unit normal bundle of the boundary of  $C(X)$ . As seen in Section A.5, the normal exponential map is a homeomorphism between  $N\partial C(X)$  and  $\partial_\infty X$ , so that the metric can then be pushed to  $\partial_\infty X$ .

A third description is as the metric at infinity defined by the equidistant foliation of  $X \setminus C(X)$ . For  $r > 0$ , let  $M_r$  be the set of points at distance  $r$  from  $C(X)$ . The surfaces  $M_r$  can be identified through the flow of the normal directions. Let  $h_r$  be the induced metric on  $M_r$ . As  $r \rightarrow \infty$ ,  $h_r$  “expands” exponentially, but the “normalized” metric  $4e^{-2r}h_r$  converges to  $h_{Th}$ .

A consequence of this last description is that the renormalized volume associated to the metric  $h_{Th}$  at infinity is (up to an additive, topological constant)

$$V_C(X) - \frac{1}{4}L_m(l),$$

where  $V_C(X)$  is the volume of the convex core and  $L_m(l)$  the length of the measured bending lamination on the boundary of the convex core.

### A.14 The renormalized volume associated to the Thurston metric

In this section we show that the renormalized volume  $V'_R$  of  $X$  associated to the Thurston metric  $h_{Th}$  at infinity is within a bounded additive constant from the renormalized volume  $V_R$  associated to the Poincaré metric  $h$  at infinity.

**Theorem A.10.** *There exists a constant  $C(\partial X)$ , depending only on the topology of  $\partial X$ , such that*

$$V'_R \leq V_R + C(\partial X) .$$

Note that the opposite inequality  $V_R \leq V'_R$  holds, up to an additive constant depending on normalization, as already stated in Lemma A.6.

The heuristic idea of the proof of Theorem A.10 is quite simple.

- The difference  $V'_R - V_R$  can be expressed as an integral of a function of the conformal factor between the hyperbolic metric and the Thurston metric, see Definition A.11. Or, more specifically, in terms of the function  $u$  such that  $h_{Th} = e^{2u}h$ .
- The contribution to this integral of the “thick” part of  $\partial X$  – the set of points where the injectivity radius for  $h$  is bounded from below – is uniformly bounded, because the conformal factor and its gradient are bounded in this region, see Lemma A.12. So we can focus on the long thin tubes around closed geodesics which are short for  $h$ .
- On those tubes, the Thurston metric can be approximated by a flat metric on a long cylinder of perimeter  $2\pi$ . We approximate the Thurston metric  $h_{Th}$  by such a flat metric  $h_\gamma$ , and write  $u = u_0 + u_\Delta$ , where  $h_\gamma = e^{2u_0}h$ , and  $h_{Th} = e^{2u_\Delta}h_\gamma$ .
- The integral term corresponding to  $u_0$  can then be explicitly computed (see Lemma A.13) and it is bounded.
- Moreover, one can find sufficient bounds on  $u_\Delta$  to show that the correction coming from  $u_\Delta$  is also bounded (see Lemma A.14).

We now proceed with the proof.

**Definition A.11.** *Let  $S \subset \partial X$ , let  $g$  be a Riemannian metric on  $S$ , and let  $u : S \rightarrow \mathbb{R}$  be a function. Let*

$$W_S(e^{2u}g, g) = -\frac{1}{4} \int_S (\|du\|_g^2 - 2K_g u) da_g .$$

It follows from the “Polyakov formula” for the dependence of the renormalized volume on the metric at infinity that

$$V'_R - V_R = W_{\partial X}(h_{Th}, h) .$$

Moreover  $h_{Th}$  is conformal to the Poincaré metric  $h$ , so we can write

$$h_{Th} = e^{2u}h ,$$

for a function  $u : \partial X \rightarrow \mathbb{R}$ .

We will use the following well-known  $C^1$  bound on  $u$  in the “thick” part of  $x \in (\partial X, h)$ .

**Lemma A.12.** *There exists a constant  $C_1 > 0$  such that for all  $x \in (\partial X, h)$  where the injectivity radius of  $h$  is at least  $\epsilon_0/2$ ,  $u \leq C_1$  and  $\|du\|_h \leq C_1$ .*

*Proof.* The bound on  $u$  can be found in the proof of Theorem 2.17 in [50].

The bound on  $du$  then follows from the bound on  $u$  and the bound on  $\Delta u$  together with the lower bound on the injectivity radius. Although this estimate is well-known to analyst, we include an informal argument for completeness. Let  $x$  be such a point, and let  $r > 0$  be such that the disk  $D(x, r)$  of center  $x$  and radius  $r$  is embedded. We can write the restriction of  $u$  to  $D(x, r)$  as

$$u = v + w ,$$

with

$$\begin{aligned} v|_{\partial D(x, r)} &= u|_{\partial D(x, r)} , \quad \Delta_h v = 0 , \\ w|_{\partial D(x, r)} &= 0 , \quad \Delta_h w = \Delta_h u . \end{aligned}$$

Then  $w$  can be written as an integral over  $D(x, r)$  of Green functions for  $\Delta_h$  on  $D(x, r)$ , multiplied by  $\Delta_h u$ , which is uniformly bounded. The uniform bound on  $dw$  at  $x$  follows. In addition,  $dv$  is uniformly bounded at  $x$  because  $u$  (and therefore also  $v$ ) is uniformly bounded on  $\partial D(x, r)$ . The uniform bound on  $du$  at  $x$  follows.  $\square$

We denote by  $T_\gamma$  the ‘‘Margulis tube’’ associated to  $\gamma$  in the hyperbolic metric  $h$ . That is, we fix a constant  $\epsilon_0 > 0$ , and let  $T_\gamma$  be the set of points at distance at most  $r$  (from  $h$ ) from the geodesic representative of  $\gamma$ , with  $r$  chosen so that the boundary of  $T_\gamma$  is the disjoint union of two closed curves of length  $\epsilon_0$  – this is possible if  $\epsilon_0$  is small enough. This tube  $T_\gamma$  can also be defined as the connected component of the geodesic representative of  $\gamma$  in the set of points in  $(\partial X, h)$  where the injectivity radius is at most  $\epsilon_0/2$ .

This tube is also equipped with a standard Euclidean metric  $h_\gamma$  conformal to the restriction of  $h$  to  $T_\gamma$ . We choose this metric  $h_\gamma$  to be isometric to  $S^1 \times [-L_\gamma, L_\gamma]$ , where  $L_\gamma$  will be determined below.

**Lemma A.13.** *There exists a constant  $C_0$  such that  $|W_{T_\gamma}(h_\gamma, h)| \leq C_0$ .*

*Proof.* By definition,  $h_\gamma$  is conformal to  $h$ , so there exists a function  $u_0 : T_\gamma \rightarrow \mathbb{R}$  such that

$$h_\gamma = e^{2u_0} h$$

on  $T_\gamma$ , with  $u_0$  constant on  $\partial T_\gamma$ .

This function  $u_0$  is clearly invariant by rotation, and only depends on the distance  $z$  to the core curve of  $T_\gamma$  (the geodesic representative of  $\gamma$  for  $h$ ). Since the curve composed of points at (oriented) distance  $z$  from the core curve has length  $l_\gamma \cosh(z)$  for  $h$ , while it has length  $2\pi$  for  $h_\gamma$ . Since each boundary component of  $T_\gamma$  has length  $2\pi$  for  $h$ ,

$$e^{u_0(z)} = \frac{2\pi}{l_\gamma \cosh(z)} ,$$

and therefore

$$|u'_0(z)| = |-\tanh(z)| \leq 1 .$$

As a consequence,

$$\begin{aligned}
|W_{T_\gamma}(h_\gamma, h)| &= \frac{1}{4} \left| \int_{T_\gamma} \|du_0\|_h^2 + 2u_0 \right| da_h \\
&\leq \frac{1}{4} \text{Area}(T_\gamma, h) + \frac{1}{2} \int_{T_\gamma} u_0 da_h \\
&\leq \frac{1}{4} \text{Area}(T_\gamma, h) + \int_0^{L_\gamma} l_\gamma \cosh(z) \log\left(\frac{2\pi}{l_\gamma \cosh(z)}\right) dz ,
\end{aligned}$$

where  $L_\gamma$  is the half-length of  $T_\gamma$  for  $h$ , that is, such that

$$l_\gamma \cosh(L_\gamma) = \epsilon_0 .$$

A direct computation shows that

$$\begin{aligned}
&\int_0^{L_\gamma} l_\gamma \cosh(z) \log(l_\gamma \cosh(z)) dz = \\
&= l_\gamma \sinh(L_\gamma) \log(l_\gamma \cosh(L_\gamma)) + 2l_\gamma(\pi/4 - \arctan(e^{-L_\gamma})) - l_\gamma \sinh(L_\gamma) ,
\end{aligned}$$

and since  $l_\gamma \cosh(L_\gamma) = \epsilon_0$ ,

$$\left| \int_0^{L_\gamma} l_\gamma \cosh(z) \log(l_\gamma \cosh(z)) dz \right| \leq \epsilon_0 \log(\epsilon_0) + \epsilon_0 + l_\gamma \frac{\pi}{2} .$$

Finally,

$$\int_0^{L_\gamma} l_\gamma \cosh(z) dz = l_\gamma \sinh(L_\gamma) \leq \epsilon_0 .$$

Adding the terms in the upper bound on  $|W_{T_\gamma}(h_\gamma, h)|$  yields the result.  $\square$

We can also compare the flat metric  $h_\gamma$  on  $T_\gamma$  to the Thurston metric  $h_{Th}$ . Since  $h_{Th}$  is conformal to  $h$ , it is also conformal to  $h_\gamma$ , so we can write

$$h_{Th} = e^{2u_\Delta} h_\gamma ,$$

for a function  $u_\Delta : T_\gamma \rightarrow \mathbb{R}$ . By definition,  $u = u_0 + u_\Delta$  on  $T_\gamma$ . The next lemma states a bound on  $u_\Delta$ , with fixed constants, over  $T_\gamma$ .

**Lemma A.14.** *There exists a constant  $C_2 > 0$  such that, on  $T_\gamma$ ,*

$$|u_\Delta| \leq C_2$$

and

$$\int_{T_\gamma} \|du_\Delta\|_{h_{Th}}^2 da_{h_{Th}} \leq C_2 .$$

*Proof.* Since  $h_\gamma = e^{-2u_\Delta} h_{T_h}$  and  $h_\gamma$  is flat, the curvature of  $h_{T_h}$  satisfies

$$\Delta_{T_h} u_\Delta = K_{T_h} .$$

Since  $K_{T_h} \leq 0$ ,  $\Delta_{T_h} u_\Delta \leq 0$ , so  $u_\Delta \geq C_1$  by the maximum principle.

In addition,  $\Delta_{T_h} u_\Delta \geq -1$ . It then follows from standard arguments (using the fact that  $K_{T_h} \in [-1, 0]$ ) that there exists  $c > 0$  such that if  $u_\Delta(x_0) \geq c$ , then  $u_\Delta \geq 1$  on the disk of center  $x_0$  and radius  $2\pi$  in  $h_{T_h}$ . Since  $(T_\gamma, h_{T_h})$  is approximated, outside a neighborhood of its boundary, by a tube of perimeter  $2\pi$ , there exists a closed curve  $\gamma'$  homotopic to  $\gamma$  going through  $x_0$ , of length less than  $3\pi$  for  $h_{T_h}$ . Then  $u_\Delta \geq 1$  on  $\gamma'$ . It would then follow that the length of  $\gamma'$  for  $h_\gamma$  is at most  $3\pi e^{-2} < 2\pi$ , a contradiction since  $(T_\gamma, h_\gamma)$  is isometric to a cylinder of perimeter  $2\pi$  and  $\gamma'$  is homotopic to  $\gamma$ . So  $u_\Delta \leq c$  on  $T_\gamma$ . We can already conclude that, for a certain  $C_2 > 0$ ,  $|u_\Delta| \leq C_2$  on  $T_\gamma$ .

Notice that

$$\begin{aligned} \int_{T_\gamma} \|du_\Delta\|_{h_{T_h}}^2 da_{h_{T_h}} &= \int_{T_\gamma} u_\Delta \Delta_{T_h} u_\Delta da_{h_{T_h}} + \int_{\partial T_\gamma} u_\Delta du_\Delta(n) ds \\ &= \int_{T_\gamma} u_\Delta K_{T_h} da_{h_{T_h}} + \int_{\partial T_\gamma} u_\Delta du_\Delta(n) ds , \end{aligned}$$

so

$$\begin{aligned} \left| \int_{T_\gamma} \|du_\Delta\|_{h_{T_h}}^2 da_{h_{T_h}} \right| &\leq C_2 \left| \int_{T_\gamma} K_{T_h} da_{h_{T_h}} \right| + \int_{\partial T_\gamma} |u_\Delta du_\Delta(n)| ds \\ &\leq C_2 \text{Area}(T_\gamma, m) + 2C_1^2 \epsilon_0 . \end{aligned}$$

The uniform bound on the integral follows.  $\square$

*Proof of Theorem A.10.* Let  $\gamma_1, \dots, \gamma_n$  be the closed geodesics of length less than  $\epsilon_0$  in  $(\partial X, h)$ . If  $\epsilon_0$  is small enough, those short closed geodesics are disjoint, and they are the core curves of disjoint long thin tubes in  $(\partial X, h)$ , denoted here by  $T_{\gamma_1}, \dots, T_{\gamma_n}$ . Since the  $\gamma_i$  are disjoint,  $n \leq (3/2)|\chi(\partial X)|$ .

For each  $i \in \{1, \dots, n\}$ , we have seen that

$$\left| W_{T_{\gamma_i}}(h_{T_h}, h) \right| = \left| \frac{1}{4} \int_{T_{\gamma_i}} \|d(u_0 + u_\Delta)\|_h^2 + 2(u_0 + u_\Delta) da_h \right| .$$

Since  $u = u_0 + u_\Delta > 0$ ,

$$\begin{aligned} \left| W_{T_{\gamma_i}}(h_{T_h}, h) \right| &\leq \left| \frac{1}{2} \int_{T_{\gamma_i}} \|_h du_0\|_h^2 + \|u_\Delta\|_h^2 + u_0 + u_\Delta da_h \right| \\ &\leq \frac{1}{2} \left| \int_{T_{\gamma_i}} \|du_0\|_h^2 + 2u_0 da_h \right| + \frac{1}{2} \left| \int_{T_{\gamma_i}} u_0 da_h \right| + \frac{1}{2} \left| \int_{T_{\gamma_i}} \|du_\Delta\|_h^2 + u_\Delta da_h \right| \\ &\leq 2|W_{T_{\gamma_i}}(h_{\gamma_i}, h)| + \frac{1}{2} \left| \int_{T_{\gamma_i}} u_0 da_h \right| + \frac{1}{2} \int_{T_{\gamma_i}} \|du_\Delta\|_{h_{T_h}}^2 da_{h_{T_h}} + \frac{1}{2} \int_{T_{\gamma_i}} u_\Delta da_h . \end{aligned}$$

However we have already seen in the proof of Lemma A.13 that the integral of  $u_0$  on  $T_{\gamma_i}$  is bounded by a fixed constant, say  $C'_0$ . Using Lemma A.13 and Lemma A.14, it follows that

$$|W_{T_{\gamma_i}}(h_{Th}, h)| \leq 2C_0 + \frac{C'_0}{2} + \frac{C_2}{2} + \frac{C_2}{2} \text{Area}(T_{\gamma_i}, h) .$$

In addition

$$|W_{\partial X \setminus (T_{\gamma_1} \cup \dots \cup T_{\gamma_m})}(h_{Th}, h)| \leq \frac{C_1^2 + C_1}{4} \text{Area}(\partial X, h) ,$$

and the result follows.  $\square$

### A.15 Asymptotic behavior of the renormalized volume when pinching a compressible curve

Finally we give here a more precise asymptotic description of the behavior of  $V'_R$  when a compressible curve is pinched. This analysis can be extended to the case where two or more compressible curves are pinched, with a dominant term which is a sum of terms corresponding to each pinched curve.

**Theorem A.15.** *Let  $(c_t)_{t \in [0,1]}$  be a smooth curve in  $\mathcal{T}_{\partial X}$ , with  $\lim_{t \rightarrow 1} c_t$  a point in the Weil-Petersson compactification of  $\mathcal{T}_{\partial X}$  corresponding to a hyperbolic metric with one simple compressible closed curve  $\gamma$  pinched. Then, as  $t \rightarrow 1$ ,*

$$V'_R(c_t) \sim \frac{-\pi^3}{L_{c_t}(\gamma)} ,$$

where  $L_{c_t}(\gamma)$  is the length of  $\gamma$  in the hyperbolic metric  $c_t$  on  $\partial X$ .

Note that the proof actually shows a little more: in the case where several curves are pinched so as to have (asymptotically) constant length ratio,  $V'_R$  is equivalent to a sum of terms corresponding to each of those short curves.

In the next lemma, we consider  $T_\gamma$  as a subset of  $\partial C(X)$ , using the nearest-point projection from  $\partial_\infty X$  to  $\partial C(X)$ .

**Lemma A.16.** *There exists a constant  $C_3 > 0$  such that if the Margulis tube  $T_\gamma$  contains a maximal segment of length  $2L$  for the induced metric  $m$  in the support of the measured bending lamination  $l$ , then all segments of  $l$  in  $T_\gamma$  have length in  $[2L - C_3, 2L + C_3]$ , and the extremal length of  $\gamma$  satisfies*

$$|\text{Ext}(\gamma) - \frac{\pi}{L}| \leq C_3 .$$

*Proof.* Let  $c$  and  $c'$  be two maximal segments in  $T_\gamma$  in the support of  $l$ . Let  $c_-, c_+$  be the endpoints of  $c$  on  $\partial T_\gamma$ , and similarly let  $c'_-, c'_+$  be the endpoints of  $c'$ , with  $c'_-$  on the same boundary component of  $T_\gamma$  as  $c_-$ . Since the boundary components of  $T_\gamma$  have length  $\epsilon_0$  for  $h$ ,  $c_-$  and  $c'_-$  are at distance at most  $\epsilon_0/2$  for  $h$ , and similarly for  $c_+$  and  $c'_+$ .

Lemma A.12 therefore shows that  $c_-$  and  $c'_-$  are also at distance at most  $e^{C_1} \epsilon_0/2$  in the Thurston metric  $h_{Th}$ , and similarly for  $c_+$  and  $c'_+$ . But then it follows that they are

also at distance at most  $e^{C_1}\epsilon_0/2$  in the induced metric  $m$  on the  $\partial C(X)$ , which is smaller than  $h_{Th}$ . This constant therefore also bounds their hyperbolic distance in  $X$ .

Since  $c$  and  $c'$  are geodesics for the hyperbolic metric on  $X$ , it follows from the triangle inequality that the hyperbolic length of  $c$  and  $c'$  are close:

$$|L(c') - L(c)| \leq e^{C_1}\epsilon_0 .$$

This proves the first point.

For the second point, let  $c_0$  be a geodesic segment of  $C(X)$ , of length  $2L$ , centered at a point close to the geodesic representative of  $\gamma$  in  $\partial C(X)$ . The same argument as above for  $c$  and  $c'$  shows that there exists a constant  $C > 0$  such that the orthogonal projection of  $T_\gamma$  on the geodesic containing  $c_0$  is within the set of points at distance at most  $C$  from  $c_0$  – we denote this extended segment by  $c_{0+}$ . Conversely, this orthogonal projection of  $T_\gamma$  contains the set of points of  $c_0$  at distance at least  $C$  from the endpoints, denoted here by  $c_{0-}$ .

Let  $N^1c_{0+}$  be the unit normal space to  $c_{0+}$ , that is, the set of unit vectors orthogonal to  $c_{0+}$ . We consider the exponential map  $\exp_\infty : N^1c_{0+} \rightarrow \partial_\infty X$  sending a vector  $n \in N^1c_{0+}$  to the endpoint at infinity of the geodesic ray defined by  $n$ . If  $\epsilon_0$  is small enough, then  $\exp_\infty$  is a diffeomorphism on its image, which is an annulus in  $\partial_\infty X$  containing  $T_\gamma$ .

An explicit computation – using for instance the Poincaré model of  $\mathbb{H}^3$  – shows that  $\text{Mod}(\exp_\infty(c_{0-})) = \frac{L-C}{\pi}$  while  $\text{Mod}(\exp_\infty(c_{0+})) = \frac{L+C}{\pi}$ . Since the modulus is increasing under inclusion, it follows that

$$\frac{L-C}{\pi} \leq \text{Mod}(T_\gamma, h) \leq \frac{L+C}{\pi} .$$

Since the conformal structure in the “thick” part of  $\partial X$  remains bounded, standard arguments then show that the modulus of  $\gamma$  in  $\partial X$  differs from its modulus in  $T_\gamma$  by bounded quantity, that is, replacing  $C$  if necessary, any annulus  $A$  extending  $T_\gamma$  in  $\partial X$  satisfies

$$\text{Mod}(T_\gamma, h) \leq \text{Mod}(A, h) \leq \text{Mod}(T_\gamma, h) + C .$$

Therefore the extremal length of  $\gamma$  in  $\partial X$  satisfies

$$\frac{\pi}{L+2C} \leq \text{Ext}_{\partial X}(\gamma) \leq \frac{\pi}{L-C} ,$$

and the result follows. □

*Proof of Theorem A.15.* We assume that the hyperbolic length of  $\gamma$  goes to zero as  $t \rightarrow 1$ . It then follows from results of Maskit [68] that the extremal length of  $\gamma$  for  $c_t$  is equivalent to

$$\text{Ext}_{c_t}(\gamma) \sim \frac{L_{c_t}(\gamma)}{\pi} .$$

It then follows from Lemma A.16 that if the Margulis tube around  $\gamma$  contains a maximal segment of length  $2L_t$  in the support of the measured bending lamination, then

$$L_t \sim \frac{\pi}{\text{Ext}_{c_t}(\gamma)} \sim \frac{\pi^2}{L_{c_t}(\gamma)} .$$

However, the total length of the measured bending lamination in the “thick” part of  $\partial C(X)$  (the set of points where the injectivity radius is larger than  $\epsilon_0/2$ ) is uniformly bounded. Moreover, the intersection with  $\gamma$  of the measured bending lamination  $l_t$  on the boundary of the convex core converges to

$$i(\gamma, l_t) \rightarrow 2\pi ,$$

as can be seen by considering the intersection of  $\partial C(X)$  with the plane  $P$  considered in Section A.6. As a consequence, since the length of every segment of  $l_t$  in  $T_\gamma$  has length approximately  $2L_t$ , the length of  $l_t$  for the induced metric  $m_t$  on the convex core behaves as

$$L_{m_t}(l_t) \sim 2\pi \cdot 2L_t = 4\pi L_t .$$

and, as a consequence,

$$V'_R \sim -\frac{L_{m_t}(l_t)}{4} \sim -\pi L_t \sim -\frac{\pi^3}{L_{c_t}(\gamma)} .$$

□

## References

- [1] J. Maldacena, “The Large  $N$  Limit Of Superconformal Field Theories and Supergravity,” *Adv. Theor. Math. Phys.* **2** (1998) 231-52, hep-th/9711200.
- [2] E. Witten and S.-T. Yau, “Connectedness of the Boundary in the AdS/CFT Correspondence,” *Adv. Theor. Math. Phys.* **3** (1999) 1635-1655, hep-th/9910245.
- [3] J. Maldacena and L. Maoz, “Wormholes in AdS,” *JHEP* **02** (2004) 053, hep-th/0401024.
- [4] P. Saad, S. H. Shenker, and D. Stanford, “JT Gravity As A Matrix Integral,” arXiv:1903.11115.
- [5] S. R. Coleman, “Black Holes as Red Herrings: Topological Fluctuations and the Loss of Quantum Coherence,” *Nucl. Phys.* **B307** (1988) 867-882.
- [6] S. B. Giddings and A. Strominger, “Loss of Incoherence and Determination of Coupling Constants in Quantum Gravity,” *Nucl. Phys.* **B307** (1988) 854-866.
- [7] D. Marolf and H. Maxfield, “Transcending the Ensemble: Baby Universes, Spacetime Wormholes, and the Order and Disorder of Black Hole Information,” *JHEP* **08** (2020) 044, arXiv:2002.08950.
- [8] S. W. Hawking and D. Page, “Thermodynamics of Black Holes in Anti de Sitter Space,” *Commun. Math. Phys.* **87** (1983) 577.
- [9] E. Witten, “Anti de Sitter Space and Holography,” *Adv. Theor. Math. Phys.* **2** (1998) 253-291, hep-th/9802150.
- [10] N. Beisert, C. Ahn, L. F. Alday, Z. Bajnok, J. M. Drummond, L. Freyhult, N. Gromov, R. A. Janik, V. Kazakov, T. Klohe, G. P. Korchemsky, C. Kristjansen, M. Magro, T. McLoughlin, J. A. Minahan, R. I. Nepomechie, A. Rej, R. Roiban, S. Schafer-Nameki, C. Sieg, M. Staudacher, A. Torrielli, A. A. Tseytlin, P. Vieira, D. Volin, and K. Zoubos, “Review Of AdS/CFT Integrability: An Overview,” arXiv:1012.3982.
- [11] E. Witten, “A Note On Complex Spacetime Metrics,” arXiv:2111.06514.

- [12] J. Cotler and K. Jensen, “AdS<sub>3</sub> Gravity and Random CFT,” JHEP **04** (2021) 033, arXiv:2006.08648.
- [13] H. Maxfield and G. J. Turiaci, “The Path Integral of 3d Gravity Near Extremality,” arXiv:2006.11317.
- [14] S. H. Shenker and D. Stanford, “Black Holes and the Butterfly Effect,” JHEP **03** (2014) 067, arXiv:1306.0622.
- [15] E. Witten, “Why Does Quantum Field Theory In Curved Spacetime Make Sense? And What Happens to the Algebra of Observables in the Thermodynamic Limit?” arXiv:2112.11614.
- [16] J. M. Maldacena, “Eternal Black Holes in Anti de Sitter,” JHEP **04** (2003) 021, arXiv:hep-th/0106112.
- [17] P. Saad, S. H. Shenker, and D. Stanford, “A Semiclassical Ramp in SYK and in Gravity,” arXiv:1806.06840.
- [18] J. Chandra, S. Collier, T. Hartman, and A. Maloney, “Semiclassical 3D Gravity As An Average of Large  $c$  CFT’s,” arXiv:2203.06511.
- [19] J. D. Brown and M. Henneaux, “Central Charges in the Canonical Realization of Asymptotic Symmetries: An Example From Three-Dimensional Gravity,” Commun. Math. Phys. **104** (1986) 207-26.
- [20] G. ’t Hooft, “A Planar Diagram Theory For Strong Interactions,” Nucl. Phys. **B72** (1974) 461.
- [21] M. Bañados, C. Teitelboim, and J. Zanelli, “The Black Hole in Three Dimensional Spacetime,” Phys. Rev. Lett. **69** (1992) 1849-51, arXiv:hep-th/9204099.
- [22] A. Strominger, “Black Hole Geometry From Near Horizon Microstates,” JHEP 02(1998) 009, hep-th/9712251.
- [23] M. Henningson and K. Skenderis, “The Holographic Weyl Anomaly,” arXiv:hep-th/9806087.
- [24] R. Dijkgraaf, J. Maldacena, G. W. Moore, and E. Verlinde, “A Black Hole Farey Tail,” hep-th/0005003.
- [25] J. Manschot and G. W. Moore, “A Modern Farey Tail,” arXiv:0712.0573.
- [26] A. Maloney and E. Witten, “Quantum Gravity Partition Functions In Three Dimensions,” JHEP 02(2010) 029, arXiv:0712.0155.
- [27] T. Dray and G. ’t Hooft, “The Effect of Spherical Shells of Matter On The Schwarzschild Black Hole,” Commun. Math. Phys. **99** (1985) 613.
- [28] R. Haag, N. M. Hugenholtz, and M. Winnink, “On The Equilibrium States in Quantum Statistical Mechanics,” Commun. Math. Phys. **5** (1967) 215-236.
- [29] J. Cardy, A. Maloney, and H. Maxfield, “A New Handle On Three-Point Coefficients: OPE Asymptotics From Genus Two Modular Invariance,” JHEP **10** (2017) 136, arXiv:1705.05855.
- [30] C. Vafa, “The String Landscape and the Swampland,” arXiv:hep-th/0509212.
- [31] P. Saad, S. Shenker, and D. Stanford, “Wormholes Without Averaging,” arXiv:2103.16754.
- [32] B. Mukhametzhanov, “Factorization and Complex Couplings in SYK and Matrix Models,” arXiv:2110.06221.

- [33] K. Krasnov. “Holography and Riemann surfaces”. *Adv. Theor. Math. Phys.* 4 (2000) 929.
- [34] K. Krasnov. “Three-dimensional gravity, point particles and Liouville theory”. *Classical Quantum Gravity* 18 (2001) 1291.
- [35] L. Takhtajan and P. Zograf. “On uniformization of Riemann surfaces and the Weil-Petersson metric on the Teichmüller and Schottky spaces”. *Mat. Sb.* 132 (1987) 303. English translation in *Math. USSR Sb.* 60:297-313, 1988.
- [36] L. A. Takhtajan and L.-P. Teo. “Liouville action and Weil-Petersson metric on deformation spaces, global Kleinian reciprocity and holography”. *Comm. Math. Phys.* 239 (2003) 183.
- [37] K. Krasnov and J.-M. Schlenker. “On the renormalized volume of hyperbolic 3-manifolds”. *Comm. Math. Phys.* 279 (2008) 637.
- [38] K. Krasnov and J.-M. Schlenker. “The Weil-Petersson metric and the renormalized volume of hyperbolic 3-manifolds”. In “Handbook of Teichmüller theory. Volume III”, volume 17 of *IRMA Lect. Math. Theor. Phys.*, 779–819 Eur. Math. Soc., Zürich 2012.
- [39] J.-M. Schlenker. “The renormalized volume and the volume of the convex core of quasifuchsian manifolds”. *Math. Res. Lett.* 20 (2013) 773. Corrected version available as arXiv:1109.6663v4.
- [40] S. Kojima and G. McShane. “Normalized entropy versus volume for pseudo-Anosovs”. *Geom. Topol.* 22 (2018) 2403.
- [41] B. Loustau. “Minimal surfaces and symplectic structures of moduli spaces”. *Geom. Dedicata* 175 (2015) 309.
- [42] K. Krasnov and J.-M. Schlenker. “A symplectic map between hyperbolic and complex Teichmüller theory”. *Duke Math. J.* 150 (2009) 331.
- [43] J. F. Brock and K. W. Bromberg. “Inflexibility, Weil-Petersson distance, and volumes of fibered 3-manifolds”. *Math. Res. Lett.* 23 (2016) 649.
- [44] S. Moroianu. “Convexity of the renormalized volume of hyperbolic 3-manifolds”. *Amer. J. Math.* 139 (2017) 1379.
- [45] F. Vargas Pallete. “Local convexity of renormalized volume for rank-1 cusped manifolds”. *ArXiv e-prints* (2015).
- [46] F. Vargas Pallete. “Continuity of the renormalized volume under geometric limits”. *ArXiv e-prints* (2016).
- [47] F. Vargas Pallete. “Additive continuity of the renormalized volume under geometric limits” (2017).
- [48] C. Ciobotaru and S. Moroianu. “Positivity of the renormalized volume of almost-Fuchsian hyperbolic 3-manifolds”. *Proc. Amer. Math. Soc.* 144 (2016) 151.
- [49] M. Bridgeman, K. Bromberg and F. Vargas Pallete. “The Weil-Petersson gradient flow of renormalized volume on a Bers slice has a global attracting fixed point”. *arXiv preprint arXiv:2105.01207* (2021).
- [50] M. Bridgeman, J. Brock and K. Bromberg. “Schwarzian derivatives, projective structures, and the Weil-Petersson gradient flow for renormalized volume”. *Duke Math. J.* 168 (2019) 867.
- [51] M. Bridgeman and R. D. Canary. “Renormalized volume and the volume of the convex

- core”. *Ann. Inst. Fourier (Grenoble)* 67 (2017) 2083.
- [52] F. Vargas Pallete. “Upper bounds on Renormalized Volume for Schottky groups”. arXiv preprint arXiv:1905.03303 (2019).
- [53] M. Bridgeman, J. Brock and K. Bromberg. “The Weil-Petersson gradient flow of renormalized volume and 3-dimensional convex cores”. arXiv preprint arXiv:2003.00337 (2020).
- [54] W. P. Thurston. “Three-dimensional geometry and topology.” (1980). Originally notes of lectures at Princeton University, 1979. Recent version available on <http://www.msri.org/publications/books/gt3m/>.
- [55] R. D. Canary. “Pushing the boundary”. In “In the tradition of Ahlfors and Bers, III”, volume 355 of *Contemp. Math.*, 109–121 Amer. Math. Soc., Providence, RI2004.
- [56] W. Thurston. “Minimal stretch maps between hyperbolic surfaces”. Arxiv preprint math/9801039 (1998).
- [57] J. S. Birman and C. Series. “Geodesics with bounded intersection number on surfaces are sparsely distributed”. *Topology* 24 (1985) 217.
- [58] W. P. Thurston. “Hyperbolic structures on 3-manifolds, II: Surface groups and 3-manifolds which fiber over the circle”. arXiv preprint math/9801045 (1998).
- [59] A. A. Borisenko. “Complete 1-dimensionnal surfaces of non-positive extrinsic curvature in a riemannian space”. *Math. USSR Sb.* 33 (1977) 485.
- [60] T. Sugawa. “Uniform perfectness of the limit sets of Kleinian groups”. *Trans. Amer. Math. Soc.* 353 (2001) 3603.
- [61] R. D. Canary. “The conformal boundary and the boundary of the convex core”. *Duke Math. J.* 106 (2001) 193.
- [62] M. Bridgeman and R. D. Canary. “From the boundary of the convex core to the conformal boundary”. *Geom. Dedicata* 96 (2003) 211.
- [63] C. R. Graham and E. Witten. “Conformal anomaly of submanifold observables in AdS/CFT correspondence”. *Nuclear Phys. B* 546 (1999) 52.
- [64] I. Kra and B. Maskit. “Remarks on projective structures”. In “Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978)”, volume 97 of *Ann. of Math. Stud.* Princeton Univ. Press, Princeton, N.J.1981 343–359.
- [65] C. Guillarmou, S. Moroianu and F. Rochon. “Renormalized volume on the Teichmüller space of punctured surfaces”. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* 17 (2017) 323.
- [66] Z. Nehari. “The Schwarzian derivative and schlicht functions”. *Bull. Amer. Math. Soc.* 55 (1949) 545.
- [67] M. Bridgeman and R. D. Canary. “Bounding the bending of a hyperbolic 3-manifold”. *Pacific J. Math.* 218 (2005) 299.
- [68] B. Maskit. “Comparison of hyperbolic and extremal lengths”. *Ann. Acad. Sci. Fenn. Ser. A I Math.* 10 (1985) 381.
- [69] F. Bonahon and J.-P. Otal. “Laminations mesurées de plissage des variétés hyperboliques de dimension 3”. *Ann. Math.* 160 (2004) 1013.