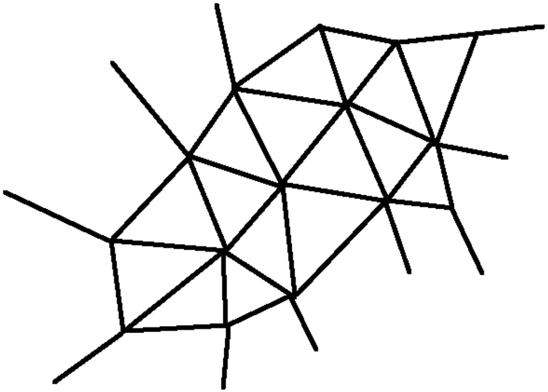
Open and Closed Topological Strings In Two Dimensions

Edward Witten

IAMP, Montreal, July 28, 2018

In the 1980's, several different approaches were developed to what one might call two-dimensional quantum gravity. I will first recall one approach that involves a discrete approximation and the large N limit of a matrix integral. (Among others, early contributions were by Weingarten; F. David; Kazakov; Ambjorn, Durhuus, and Frohlich; Kazakov, Kostov, and Migdal and then the models were solved by Douglas and Shenker; Gross and Migdal; and Brezin and Kazakov.)

Just to make a discrete approximation to geometry, one can consider, for example, a random triangulation of a two-manifold with T triangles, where T is going to be very large:

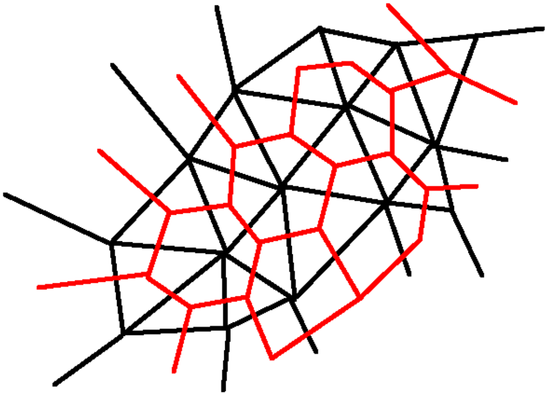


We consider each triangle to be, for example, an equilateral triangle with a side of length a. Then we hope that in some sort of limit with $a \to 0$ and $T \to \infty$, random triangulations will generate some sort of reasonable model of two-dimensional quantum gravity.

Very concretely, to study this we have to count triangulations of a given two-manifold Σ with T triangles, in the limit of large T. The answer turns out to be something like $\exp(cT)T^s(1+\dots)$ where c and s are constants; s, but not c, depends on the topology of Σ. The leading exponential $\exp(cT)$ is "renormalized" away (since the area of the surface is a multiple of T, one can view this as the renormalization of the cosmological constant). The physics is then in s, as well as further corrections in the series.

It is actually a little more convenient to consider not the triangulation itself but the dual graph

Triangulation.png



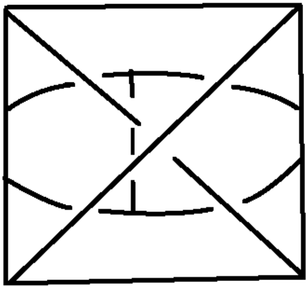
The dual graph is *trivalent*, though it is built from a variety of polygons (the original graph was built from triangles, but had vertices of all orders).

There is a simple and convenient way to count trivalent graphs, if one does not care about using them to triangulate a two-manifold. One looks at the asymptotic expansion in powers of λ of the integral

$$I(\lambda) = \int_{-\infty}^{\infty} \mathrm{d}x \; \exp\left(-\frac{1}{2}x^2 + \frac{\lambda}{3!}x^3\right).$$

When one expands this in powers of λ, one generates Feynman diagrams with cubic vertices – i.e., trivalent graphs. The propagator is 1, so the integral just counts trivalent graphs (each one weighted by the inverse of the order of its symmetry group).

These graphs aren't quite what we want, since they are abstract trivalent graphs, not triangulations of a two-dimensional surface:



Counting such abstract graphs is not what we want, but anyway they are easy to count as the number of trivalent graphs with V vertices (and no external lines) is just the coefficient of λ^V in

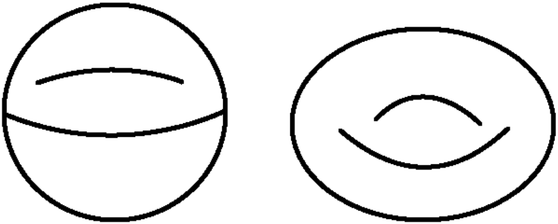
$$\int_{-\infty}^{\infty} \mathrm{d}x \, \exp\left(-\frac{1}{2}x^2 + \frac{\lambda}{3!}x^3\right)$$

or

$$c_V = \int_{-\infty}^{\infty} \mathrm{d}x \left(\frac{\lambda x^3}{3!}\right)^V \exp(-x^2/2).$$

(We can do this integral in closed form, but we can also just do a saddle point evaluation for large V.)

We actually want to count, not abstract trivalent graphs, but such graphs that an be drawn on the surface of a two-manifold, which we'll take for the moment to be a closed Riemann surface without boundary, with some given genus:



How to count graphs that are drawn on such a two-manifold was explained by 't Hooft over 40 years ago. One simply replaces the real variable x by an $N \times N$ hermitian matrix M and considers a matrix version of the integral:

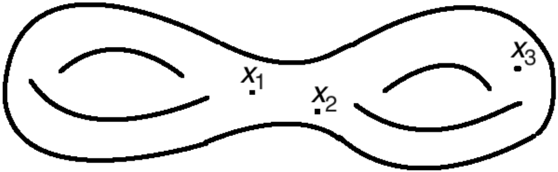
$$\int \mathrm{d}M \exp\left(\mathrm{Tr}(-M^2/2 + \lambda M^3/3!)\right).$$

Now one expands the integral in powers of λ and $1/N$. The coefficient of $N^{2-2g}\lambda^V$ is the number of trivalent graphs with V vertices that can be drawn on a Riemann surface of genus g. As first shown by Brezin, Parisi, Itzykson, and Zuber in 1977, this integral can be effectively analyzed for large N by random matrix methods that go back to Wigner, Dyson, and Mehta, among others.

The contribution of the physicists whom I cited at the start was to show that one gets a more interesting result if, while taking N to infinity, one adjusts λ to a critical value at which the perturbation expansion diverges. In this "scaling limit," the number of triangles in a typical triangulation diverges and one gets what is believed to be a good model of two-dimensional gravity. The model is completely soluble in this limit and the solution was given almost 30 years ago by Douglas and Shenker; Gross and Migdal; and Brezin and Kazakov, in terms of the solutions of certain KdV and/or Virasoro equations.

That is about as much as I will be able to say about this approach at the moment. A second approach to two-dimensional gravity involves a conventional string theory with a certain matter system (a Liouville field). This is actually an important part of the subject, but we do not have time for it today.

There is a third approach, which was also found to be equivalent to the first two, that I want to describe today. This involves intersection theory on the moduli space of Riemann surfaces. Let $\mathcal{M}_{g,n}$ be the moduli space of Riemann surfaces Σ of genus g with n marked points or punctures that we will call $x_1, \ldots, x_n$:



Associated to each of the points x_i is a $U(1)$ gauge field which is the connection on its tangent space and this has a first Chern class which I will call $c_1^{(i)}$. It is the first Chern class of a "Berry-like connection" associated to the i^{th} point. The "correlation functions of topological gravity" are

$$\int_{\mathcal{M}_{g,n}} \prod_{i=1}^{n} (c_1^{(i)})^{q_i}$$

with any integers q_i such that

$$\sum_i q_i = 3g - 3 + n.$$

(Otherwise the integral is 0 for trivial reasons.)

It turns out that these correlation functions are equivalent to what one can compute in the matrix model. Cutting a few corners in the explanation, in the matrix model one can compute

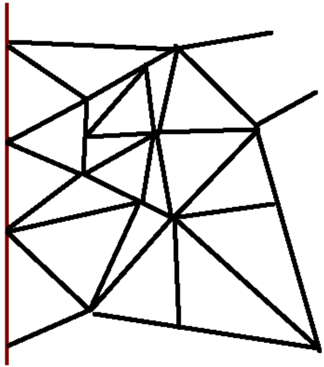
$$\langle \mathrm{Tr}\, M^{q_1}\, \mathrm{Tr}\, M^{q_2} \cdots \mathrm{Tr}\, M^{q_n} \rangle$$

and this is equivalent to computing

$$\int_{\mathcal{M}_{g,n}} \prod_{i=1}^{n} (c_1^{(i)})^{q_i}$$

This assertion is my conjecture from around 1990, proved by Kontsevich and then in other ways by Maryam Mirzakhani; Okounkov and Pandharipande; and Kazarian and Lando. I will not try to explain any of the proofs today.

Instead I will tell you about another side of the story. What happens if the surface Σ has a boundary? In the approach based on triangulations of moduli space, there is no problem. There is no difficulty to triangulate a Riemann surface with boundary:



Although it is a little more subtle, there is also no problem to find a matrix model that counts triangulations of a Riemann surface with boundary. One merely adds "vector" degree ψ, $\overline{\psi}$ of freedom to the matrix model and considers an integral such as

$$\int \mathrm{d}M \, \mathrm{d}\psi \, \mathrm{d}\overline{\psi} \ \exp\left(\mathrm{Tr}(-M^2/2 + \lambda M^3/3!) + \overline{\psi}(M-u)\psi\right).$$

Equivalently, one can integrate out $\psi, \overline{\psi}$ and get such an integral for M only. For example, if $\psi, \overline{\psi}$ are fermions, the result of integrating them out is

$$\int \mathrm{d}M \ \exp\left(\mathrm{Tr}(-M^2/2 + \lambda M^3/3!)\right) \det\left(M-u\right).$$

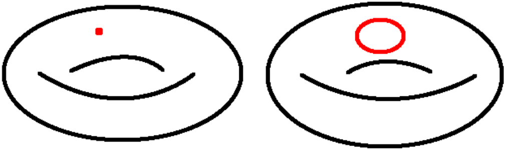
It is also relatively clear, though possibly a little more subtle, how to incorporate boundaries in the string theory description.

But what happens in the approach based on topological gravity? Here we have a problem because there is an anomaly. The anomaly is that the moduli space of Riemann surfaces Σ with boundary is *unorientable* (even if Σ itself is orientable) and therefore the topological correlation functions

$$\int_{\mathcal{M}_{g,n}} \prod_{i=1}^{n} (c_1^{(i)})^{q_i}$$

don't make sense.

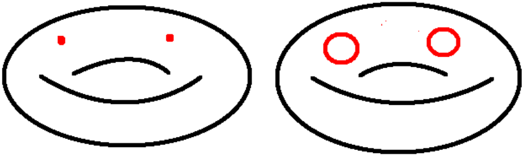
To understand why the moduli space of Riemann surfaces with boundary is unorientable, let us compare a Riemann surface with a marked point to one with a hole:



Adding a puncture to a Riemann surface adds two real moduli – say $\mathrm{Re}\, x$ and $\mathrm{Im}\, x$, where x is its position. Adding a hole adds three real moduli, which one can think of as $\mathrm{Re}\, x$, $\mathrm{Im}\, x$, and b, where again x is the position and b is the size of the hole.

Adding a puncture to a complex Riemann surface does not affect the orientability of its moduli space, because there is no problem in the sign of a two-form $\mathrm{d}x\mathrm{d}\overline{x}$ that “soaks up” the extra moduli. Similarly there is no problem if we add *one* hole because there is no sign problem for the three-form $\mathrm{d}x\mathrm{d}\overline{x}\mathrm{d}b$.

The problem comes with two or more holes:



For punctures, the two-forms $\mathrm{d}x\mathrm{d}\overline{x}$ are bosonic and commute, so we don't have to decide on an ordering of the punctures. But for holes, we have instead a three-form $\mathrm{d}x\mathrm{d}\overline{x}\mathrm{d}b$ and these *anticommute* for different holes. So for a Riemann surface with $h > 1$ holes, to try to orient its moduli space, we would have to pick an ordering of the holes, modulo an even permutation.

Not having a way to do this is a kind of "gravitational anomaly" which means that one cannot make sense of the sign of the path integral measure.

Actually, if Σ is a Riemann surface with two or more holes, it has a diffeomorphism (preserving the orientation of Σ) that exchanges two of the holes, and therefore the moduli space of conformal structures on Σ is actually *unorientable*, not just lacking in a natural orientation.

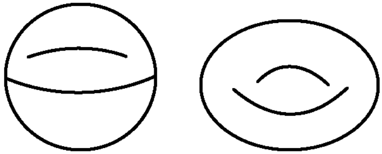
Twenty-five or so years ago, it troubled me that in two of the three approaches to two-dimensional gravity, one can allow Σ to have a boundary, and in the third one seemingly cannot. But not seeing what to do about it, I eventually gave up and moved on.

To my surprise, these difficulties were overcome a couple of years ago by mathematicians Pandharipande, Solomon, and Tessler (arXiv:1507.04951); the proof was completed by Buryak and Tessler (arXiv:1501.07888). In arXiv:1804.03275, R. Dijkgraaf and I have compared the PST-BT answer to the matrix integral that I described earlier for a Riemann surface with boundary, and we also described what they did in a somewhat more physical language. I will use the rest of the time today to try to explain that last point.

The unorientability of the moduli space is a kind of "gravitational anomaly": there is a diffeomorphism of Σ – namely one that exchanges two holes – that reverses the sign of the integration measure that we are trying to use in computing correlation functions. When this happens in physics, we try to compensate by adding a matter system that will cancel the anomaly. In the present context, that matter system is going to have to be a two-dimensional topological quantum field theory (TQFT).

However, the topological field theory that we add has to have the property that it is not just anomaly-free (i.e. well-defined) but trivial on an orientable Riemann surface. Otherwise, we would change the theory on a closed surface. To be more exact, the theory we add must be trivial on a closed surface up to a renormalization of the string coupling constant g_{st}. A g-loop amplitude is proportional to g_{st}^{2g-2}, and a renormalization of the string coupling constant would multiply a closed-string amplitude in genus g by c^{2g-2} for some constant c. So the topological field theory we are looking for must have for its genus g partition function on a closed Riemann surface c^{2-2g} for some c.

The "answer" turns out to be, in part, that Σ must be endowed with a spin structure. A spin structure is a choice of sign when a fermion is parallel-transported around a topologically nontrivial loop.



The Levi-Civita connection defines parallel transport for integer spin, but for half-integer spin there are extra sign choices when one goes around noncontractible loops. On the sphere there is only one spin structure, but on the torus there are $2^2 = 4$ of them.

In two dimensions, there is a rather trivial topological field theory in which one simply sums over spin structures, weighting each one by a factor of $1/2$ (this is a special case of dividing by the order of the automorphism group; 2 is the number of automorphisms of a spin structure). In genus g, there are 2^{2g} spin structures, so the partition function of the theory that I mentioned is $\frac{1}{2}2^{2g} = 2^{2g-1}$. This is not of the form c^{2-2g} for any c, so we need something else.

A spin structure in two dimensions can be either "even" or "odd" – we recall the definition shortly. There is a topological field theory in which the spin structures are weighted by an extra factor of $(-1)^w$, where w is 0 or 1 for an even or odd spin structure. The partition function is then

$$\frac{1}{2}\sum_s (-1)^{w(s)}$$

where s labels a spin structure. The number of even spin structures is $\frac{1}{2}(2^{2g}+2^g)$ and the number of odd ones is $\frac{1}{2}(2^{2g}-2^g)$. The partition function is then

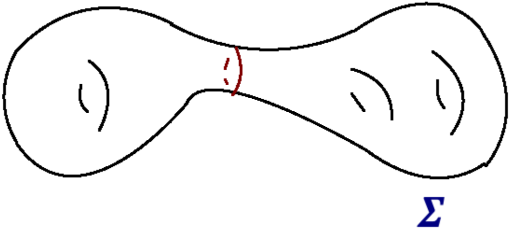
$$\frac{1}{2^2}\left(\left(2^{2g}+2^g\right)-\left(2^{2g}-2^g\right)\right)=\frac{1}{2}2^g=(\sqrt{2})^{2g-2}$$

and this finally is of the desired form.

The topological field theory that I have just introduced is related to the Kitaev chain of Majorana fermions in condensed matter physics, and it is also related to the high temperature phase of the Ising model. Unfortunately there is not really time to explain those points today.

So far I have described the topological field theory with the factor $(-1)^w$ on a closed Riemann surface. Actually, this theory has an anomaly on a Riemann surface with boundary – which is what we want, of course. One definition of the $\mathbb{Z}_2$-valued invariant w on a Riemann surface Σ without boundary is as follows. View Σ as a complex Riemann surface and let $K^{1/2}$ be a square root of the canonical bundle $K \to \Sigma$, corresponding to a spin structure. Then the dimension mod 2 of $H^0(\Sigma, K^{1/2})$ is a deformation invariant and this is w. One way to prove the deformation invariance is to interpret the $\overline{\partial}$ operator on $K^{1/2}$ as a Dirac operator and interpret w as a "mod 2 index" in the sense of Atiyah and Singer.

This definition of w does not make sense if Σ has a boundary, because there is no chiral Dirac operator on a manifold with boundary, and indeed there is no invariant $(-1)^w$ if Σ has a boundary. However, even if Σ has a boundary, it is possible to define $(-1)^w$ as a unit vector in a real line associated to the boundary. One way to do this is to use a gluing relation in index theory:



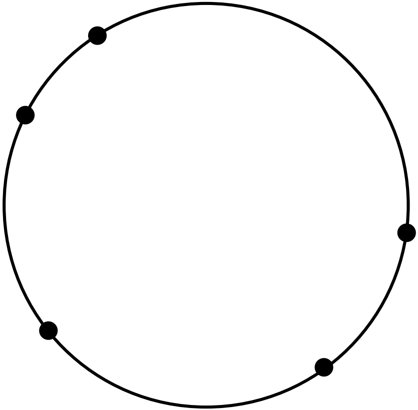
What I'd like to say in order to finish the lecture is that $(-1)^w$ is naturally a section of the orientation bundle of the unoriented (and unorientable) moduli space $\mathcal{M}_{g,n,n',h}$ of Riemann surfaces with holes and boundary and bulk punctures. In that case an improved version of the definition of the correlation functions would be

$$\frac{1}{2}\sum_s \int_{\mathcal{M}_{g,n,n',h}} (-1)^{w(s)} \tau_{a_1}\tau_{a_2}\cdots\tau_{a_n}.$$

Actually, it isn't quite true $(-1)^w$ is a section of the orientation bundle. What is true is a little more complicated. This regrettably also makes it a little hard to finish my lecture properly and without getting into more details.

The rule that PST-BT give to finish canceling the anomaly and to give a proper definition of correlation functions is as follows: The boundary of a two-manifold is, of course, a 1-manifold. In 1 dimension, there is only one gamma matrix γ_1 obeying $\gamma_1^2 = 1$, so it can be represented by a 1×1 real matrix, and hence (to use a fancy language) the spin bundle of a 1-manifold is a real vector bundle of rank 1. As such it has, up to homotopy, two real trivializations: a trivialization is a choice of what is the “positive” direction in the real line bundle.

The rule of PST-BT is that the spin bundle of the boundary is everywhere trivialized away from marked points, but the sign of the trivialization is required to “jump” when one crosses a boundary marked point:



PST-BT show that when what I said earlier is supplemented with this rule about local trivializations and their jumping, the anomaly cancellation is completed and the problem with orientations goes away. They further show that a problem with the boundary of the moduli spaces (which I did not have time to describe) goes away, so the correlation functions become well-defined.

Finally, they showed that these correlation functions can be explicitly calculated: they obey an "open" version of the Virasoro and KdV equations via which the correlation functions for closed surfaces were computed almost 30 years ago.

Instead of going into all this, I want to use the remaining time to give a few hints about how Dijkgraaf and I gave a more physical interpretation to this story. Unfortunately it is difficult to do this without assumig some familiarity with supersymmetric field theories and their twisting to make topological field theories..

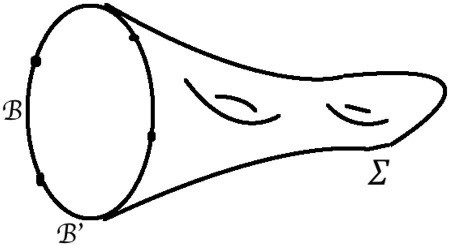
The basic approach is as follows. The theory that is trivial on a Riemann surface without boundary and becomes anomalous when there is a boundary can be realized by taking an $N = 2$ supersymmetric Landau-Ginzburg theory with a single chiral superfield ϕ and a superpotential

$$W(\phi) = im\phi^2.$$

The superpotential only has one critical point – at $\phi = 0$ – which is why the theory is "trivial." Since W is quasihomogeneous – in fact, homogeneous – we can make a topological field theory by "twisting." As W is supposed to have spin 1 after twisting, ϕ has spin $1/2$. Therefore the twisted theory requires a choice of spin structure. It can be shown that on a closed surface Σ with a given spin structure, the partition function is $(-1)^w$. (There is a unique classical solution, $\phi = 0$, and the sign of the fermion determinant gives the answer $(-1)^w$.)

If we want to study this theory on a Riemann surface with boundary, we need to pick a "brane." Roughly speaking there is only one natural brane in the problem – a Lefschetz thimble associated to the unique critical point at $\phi = 0$. To be more exact, this is the *support* of the only natural branes. In general, in a theory with a target space X, the choice of a Lagrangian submanifold $L \subset X$ is not enough to determine a brane. We also need (among other things, which aren't relevant here) to pick an *orientation* of L. So in our problem there are two natural branes $\mathcal{B}$ and $\mathcal{B}'$, which correspond to the Lefschetz thimble with one or the other orientation. Neither one is more natural than the other.

To get the PST-BT answer, one considers the theory with one brane of each type. In the twisted theory, the branes have a special meaning:



Away from the boundary, after twisting, ϕ is a spinor field on Σ, a section of its complex spin bundle; the meaning of the brane is that along the boundary, ϕ is a section of the (real) spin bundle of the boundary. Each boundary segment is labeled by $\mathcal{B}$ or $\mathcal{B}'$ corresponding to a choice of orientation of the spin bundle of the boundary.

Since the spin bundle of the boundary is a real bundle of rank 1, a choice of its orientation is the same as a trivialization of this spin bundle (up to homotopy) and thus we have the PST-BT rule: away from boundary punctures, the spin bundle of the boundary is trivialized. The rest of their rule – the trivialization jumps across each puncture – means that the operators inserted at boundary punctures are of type $\mathcal{B}\mathcal{B}'$ or $\mathcal{B}'\mathcal{B}$, not of type $\mathcal{B}\mathcal{B}$ or $\mathcal{B}'\mathcal{B}'$.

The absence of $\mathcal{B}\mathcal{B}$ or $\mathcal{B}'\mathcal{B}'$ insertions has a simple explanation: the only local BRST invariant operator of type $\mathcal{B}\mathcal{B}$ or $\mathcal{B}'\mathcal{B}'$ would be a multiple of the identity, and its 1-form descendant vanishes. So the only boundary insertions that are interesting are the ones of type $\mathcal{B}\mathcal{B}'$ or $\mathcal{B}'\mathcal{B}'$. Moreover one can show that globally, one cannot distinguish $\mathcal{B}\mathcal{B}'$ from $\mathcal{B}'\mathcal{B}$, so there is essentially only one type of boundary insertion and one boundary coupling parameter, as in the theory of PST-BT.