Minimal Area Problems and Quantum Open Strings

Barton Zwiebach[*]

Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract. We discuss minimal area problems for surfaces with boundaries and both open and closed string punctures. We define open-closed string diagrams to be surfaces with metrics of minimal area under the condition that any nontrivial Jordan open curve be longer or equal to π and any nontrivial Jordan closed curve be longer or equal to 2π. It is proven that the double of an open-closed string diagram is a closed string diagram of covariant closed string field theory.

1. Introduction and Summary

The fundamental problem in constructing quantum field theories of open and closed strings was the lack of suitable string diagrams. A string diagram is a punctured Riemann surface together with a choice of local analytic coordinates at the punctures. A closed string puncture is a puncture in the interior of the surface, and an open string puncture is one that lies on a boundary component of the surface. Two coordinate systems around a closed string puncture are said to be equivalent if they only differ by a constant phase. The string diagrams must satisfy stringent conditions in order to define a quantum string field theory.

It is typically possible to extract canonical local analytic coordinates from special metrics on punctured surfaces. Thus a way to obtain a string diagram is to find a problem that gives, for any surface, a unique metric on it. As emphasized in [St] it is a fundamental property of minimal area metrics that they are unique. It was recently proposed that closed string diagrams, relevant to punctured surfaces without boundaries, could be defined by demanding the metric to be of minimal area under the condition that all nontrivial homotopy closed curves be longer or equal than a single length, conventionally taken to be 2π [Zw 1]. Such minimal area metrics were found for all punctured Riemann spheres, where they arise from

Jenkins-Strebel (JS) quadratic differentials (quadratic differentials with closed trajectories). The minimal area string diagrams are in fact the diagrams arising from the Feynman rules of the classical closed string field theory [SaZw, KKS]. This allowed to give a proof that the Feynman rules construct the complete moduli spaces of Riemann spheres without overcounting or undercounting [Zw 1]. This minimal area problem defines candidate string diagrams for all higher genus surfaces. Although for every moduli space of higher genus a large subspace has metrics of minimal area that arise from quadratic differentials [Zw 1], some metrics may not arise from quadratic differentials. The minimal area property, however, may well be sufficient to guarantee that the string diagrams arise from a quantum closed string field theory [Zw 2].

The purpose of the present paper is to clarify some aspects of the closed string minimal area problem, and to extend the problem to one relevant to surface with boundaries. We want to discuss a minimal area problem naturally related to the closed string problem. The resulting string diagrams, denoted as *open-closed* string diagrams are defined by the following minimal area problem:

Open-Closed String Diagrams. Consider a genus g Riemann surface with b boundaries, n closed string punctures and m open string punctures $(g, b, n, m \geq 0)$. (If $g = b = 0$, then $n \geq 2$, and if $g = n = 0$, $b = 1$, then $m \geq 2$.) The open-closed string diagram is given by the minimal area metric under the condition that the length of any open Jordan curve of non-trivial homotopy (relative to boundary segments) be greater or equal to π, and that of any non-trivial Jordan closed curve be greater or equal to 2π. $\square$

We have excluded the sphere with one or zero punctures, the disk, and the disk with one open string puncture since these surfaces have no nontrivial curves. Due to the conditions imposed on the metric, the open-closed string diagrams will not have short open curves nor short closed curves. One of the main results in this paper will be to establish the following theorem relating the above open-closed string diagrams to closed string diagrams.

Theorem 1. *A string diagram on an open surface R is an open-closed string diagram if and only if its double is a closed string diagram on the double surface $\hat{R}$.*

The physical implications of the open-closed string diagrams are interesting [Zw 3]. They give rise to a covariant quantum theory of open and closed strings, manifestly factorizable both in open and closed string channels. This theory is different from Witten's open string field theory [Wi], whose string diagrams are built gluing strips of width π together via a symmetric three string vertex. While there is considerable evidence that Witten's theory gives a single cover of all moduli spaces of surfaces with boundaries and punctures [GMW, Og, SB, Sa], it has been correctly emphasized by Samuel [Sa] that no complete proof exists. Since the Witten string diagrams may have small closed curves, a natural minimal area problem for these diagrams would only impose conditions on open curves. We define *open* string diagrams by

Open String Diagrams. Consider a genus g Riemann surface with b boundaries and m open string punctures $(g, m \geq 0, b \geq 1)$. (If $g = 0$ and $b = 1$, then $m \geq 2$.) The open string diagram is given by the minimal area metric under the condition that the length of any Jordan open curve of non-trivial homotopy (relative to boundary segments) be greater or equal to π. $\square$

The proof that the string diagrams of [Wi] solve this minimal area problem, (namely, that they are open string diagrams), together with a new existence and uniqueness theorem for quadratic differentials that establishes the single cover of moduli space, is given in a sequel to this paper [Zw 4]. Actually, the above problem, which imposes conditions on open curves only, can also define string diagrams in the presence of closed string punctures, and it appears that the solution always arises from a quadratic differential [Zw 5].

This paper is organized as follows. In Sect. 2 we begin by defining precisely Feynman diagrams, string diagrams, and explaining their relation. We spell out very explicitly the conditions under which a set of Feynman diagrams give a single cover of moduli space. We consider separately the possibilities that string diagrams arise from quadratic differentials or from minimal area metrics. We then show that if all nontrivial Jordan closed curves on a surface are longer or equal to 2π, then any nontrivial closed curve is longer or equal to 2π. We then discuss nontrivial closed curves on polyhedra, and finally define the non-trivial open curves relevant to the minimal area problems. All of the results in this section will be used in the latter sections or are required in [Zw 4].

In Sect. 3 we establish results that relate homotopy classes of open and closed curves. In Sect. 4 after some brief comments on minimal area problems, we deal with the issue of regularization that arise because the naive area is actually infinite due to the presence of punctures. We extend the definition of reduced area of [Zw 1] for the case when there are boundary components with open string punctures. In Sect. 5 we prove Theorem 1, thus establishing that the open-closed minimal area problem is naturally related to the closed string minimal area problem.

Open-closed string diagrams could have been defined directly from the closed string minimal area problem. Given an open surface R, we double it, find the unique closed string diagram, and cut it back to obtain a unique diagram, which would be defined as the open-closed diagram for R. This way we get a unique diagram for any open surface. Why do we need to establish the minimal area properties of the open-closed string diagrams? The reason is that having those string diagrams does not yet guarantee that they arise from the Feynman rules of a quantum string field theory. The minimal area property essentially guarantees that one can extract vertices which together with propagators build all the string diagrams from Feynman rules. The minimal area property also explains the factorization properties of the off-shell string amplitudes arising from the open-closed string diagrams. It is interesting that open-closed string diagrams can be defined intrinsically without any need for doubling. In the case of light-cone, closed string diagrams are defined via a problem that gives a unique abelian differential for a given surface [GW]. Open string diagrams are defined by doubling. It would be interesting to find an intrinsic definition for these diagrams, and to see if there is more than one possible definition, as it is the case in covariant open string theory.

2. On String Diagrams and Nontrivial Curves

In this section we begin by presenting results that amount to clarification of some points of [Zw 1]. In Sect. 2.1 we define Feynman diagrams and string diagrams, and explain why we need to establish an isomorphism between the set of Feynman diagrams and the set of string diagrams. This isomorphism, together with an

existence and uniqueness theorem for string diagrams guarantees the single cover of moduli space. In Sect. 2.2 we explain why it is sufficient to check that any nontrivial Jordan closed curve is longer or equal to 2π in order to guarantee that any nontrivial closed curve is longer or equal to 2π. Then in Sect. 2.3 we show that certain classes of curves on polyhedra are nontrivial. After discussing all these points related to [Zw 1], we turn in Sect. 2.4 to definitions and basic properties of nontrivial open curves.

2.1. Feynman Diagrams, String Diagrams, and Covering Moduli Space

Since minimal area metrics may not arise always from quadratic differentials, it is important to establish precisely to what degree the uniqueness of the minimal area metric implies the single cover of moduli space by the Feynman diagrams of a string field theory. In order to eludicate this point we must begin by defining string diagrams, Feynman diagrams, and show how they relate to each other.

A string diagram is a Riemann surface with a choice of local analytic coordinates at the punctures, and is denoted as the pair (R, z_i), where R is a punctured Riemann surface and z_i are the local coordinates. When local coordinates can be extracted from a special metric on the surface, a string diagram will be denoted as (R, ϱ), where ϱ is the special metric. When local coordinates arise from a Jenkins-Strebel (JS) quadratic differential (qd) with second order poles at the punctures, the string diagram will be denoted as (R, φ), where φ is the quadratic differential. If we discuss string diagrams that may arise from either quadratic differentials or metrics we will denote them as $(R, \varphi/\varrho)$. Two string diagrams are the same if the underlying surfaces are the same and the local coordinates, or quadratic differentials, or metrics, are the same. Two string diagrams are inequivalent if the underlying surfaces are different, or, in the case the surfaces are the same, if the local coordinates, or the quadratic differentials, or the metrics, differ.

Consider now the Feynman diagrams (FD's) of the classical closed string field theory. They correspond to surfaces built with propagators, which are cylinders of circumference 2π, joined by polyhedra type vertices. Each Feynman diagram is defined by the following data: a collection of external states (labeled punctures), each one representing a semiinfinite tube of circumference 2π, some number of propagators, each with two parameters, length and twist, a number of vertices, each one a polyhedron, whose parameters are the number of faces, the number of vertices, and the edge parameters. The latter give for each edge its length, its two endpoints, and which faces it connects. Finally we must specify which propagators and which external strings connect to which face in each polyhedron. All this data defines the Feynman diagram. We summarize this by $FD = (R, D)$, where D denotes the data, which determines also the Riemann surface R.

We say that two Feynman diagrams FD_1 and FD_2 are *inequivalent* if their data cannot be made to agree by relabeling. We may relabel the polyhedra, and in each polyhedron we may relabel the faces, the vertices and the edges. We may relabel the internal propagators, but we cannot relabel the external states since they correspond physically to different states (we always consider the case of distinguishable punctures). It is a basic combinatorial result that the Feynman rules of a field theory, by means of the symmetry factors, construct each inequivalent Feynman diagram once.

As explained in [Zw 1], each Feynman diagram defines a Riemann surface with a canonically defined JS quadratic differential of special properties. Each of the above cylinders are the characteristic ring domains, and the polyhedra are the

critical graphs. Thus (a) each Feynman diagram (R_i, D_i) is giving us a canonically defined string diagram (R_i, φ_i). Conversely, (b) each JS quadratic differential of special properties (R_i, φ_i) defines a canonical Feynman diagram (R_i, D_i). This is clear because the closed horizontal trajectories determine uniquely the propagators, and the critical graphs determine uniquely the vertices. This allows us to extract the data for the Feynman graph in a canonical way. If we denote by $\mathscr{F}$ the space of all inequivalent Feynman diagrams (R_i, D_i) and by S the space of all inequivalent string diagrams (R_i, φ_i), property (a) defines a map $\psi : \mathscr{F} \mapsto \mathscr{S}$, and property (b) defines a map $\psi^{-1} : \mathscr{S} \mapsto \mathscr{F}$. The way the maps are defined explicitly shows that they are inverses of each other. Namely, $\psi \circ \psi^{-1} = i_{\mathscr{S}}$, and $\psi^{-1} \circ \psi = i_{\mathscr{F}}$, where $i_{\mathscr{S}}$ and $i_{\mathscr{F}}$ are the identity maps on $\mathscr{S}$ and $\mathscr{F}$ respectively. The map ψ is therefore an isomorphism between $\mathscr{F}$ and $\mathscr{S}$.

All of this can be repeated for the case when minimal area metrics define the string diagrams $SD = (R_i, \varrho_i)$. Two well-defined maps must exist: (a) each Feynman diagram must define a canonical minimal area metric, and (b) each minimal area metric must define a canonical Feynman diagram. These two maps must be inverses of each other. Then it follows that we still have the isomorphism between $\mathscr{F}$ and $\mathscr{S}$.

If the special type of qds/metrics can be shown to arise from a problem giving a unique qd/metric for any Riemann surface, then the above isomorphism between $\mathscr{F}$ and $\mathscr{S}$ guarantees a single cover of moduli space $\mathscr{R}$ (the set of inequivalent surfaces). This just means that the forgetful map $\eta : \mathscr{F} \mapsto \mathscr{R}$ defined by $\eta(R_i, D_i) = R_i$, is a one to one and onto map of sets. *Proof.* Consider the forgetful map $\chi : \mathscr{S} \mapsto \mathscr{R}$ defined by $\chi(R_i, \varphi_i/\varrho_i) = R_i$. It is a one to one and onto map of sets, thus an isomorphism. It is one to one because if two different qds/metrics would yield the same Riemann surface, it would be a violation of uniqueness. It is onto because for any surface R_i one can find the corresponding qds/metric. Define now the composition map $\eta = \chi \circ \psi : \mathscr{F} \mapsto \mathscr{R}$. It is a forgetful map: $\chi \circ \psi(R_i, D_i) = \chi(R_i, \varphi_i/\varrho_i) = R_i$ that gives us the Riemann surface associated with the Feynman diagram. Since it is the composition of two isomorphisms, it is an isomorphism. Thus it is one to one and onto, as desired.

For the case of minimal area metrics, as compared to the case when we have JS quadratic differentials, it is harder to find the canonical maps ψ and ψ^{-1}. Under the assumptions of [Zw 2] we can derive vertices such that the Feynman rules construct minimal area metrics, thus defining ψ. We expect that each minimal area metric determines a unique Feynman graph, thus defining ψ^{-1}, since minimal area metrics appear to give rise to foliations by geodesics of length 2π. These foliations generalize the horizontal trajectories of quadratic differentials, and should determine the Feynman diagram. The maps ψ and ψ^{-1} are expected to be inverses of each other. We will not attempt here to establish any of these results.

2.2. Closed Curves with Self-Intersections

Consider a minimal area metric under the condition that any non-trivial homotopy closed Jordan curve (a closed curve without self-intersections) be longer or equal to 2π. It will be shown that, on this minimal area metric *any* non-trivial homotopy simple closed curve is longer or equal to 2π. This includes the possibility of having self-intersections. In fact curves belonging to some non-trivial homotopy classes must have self-intersections. For example, a closed curve going "n" times around a puncture must have at least one self-intersection for $n > 1$. Given a closed

curve with self-intersections, the intersection points determine a set of arcs in the curve. We define a closed subcurve of a closed curve Γ as a closed curve made up with arcs of Γ, where not all arcs need be used, and no arc is used more than once. The reason any nontrivial homotopy closed curve is longer or equal to 2π is that, as will be proven below, any such curve must contain a closed subcurve, without self-intersections, that is also nontrivial. The subcurve must be longer or equal to 2π, and as a consequence the original curve with self-intersections is also longer or equal to 2π. This fact will be necessary in Sect. 5.

Lemma 1. *A nontrivial homotopy closed curve has a nontrivial homotopy closed Jordan subcurve.*

Proof.[1] If the original curve is a Jordan closed curve, this curve is the closed subcurve. If the original curve is not a Jordan closed curve, it has a finite number of intersection points. Remove the set of intersection points, getting a collection of arcs. The number of arcs is a finite number greater or equal to two.

Consider now the set S, whose elements are nontrivial closed curves built joining the above mentioned arcs. Each arc may only be used once, and curves need not use all the arcs. The set S is not empty since the original curve is in S. We demand that the curves in S be parametrized monotonically (such parameterization can be obtained by homotopy). Let l be the lowest number of arcs making up a curve in S. Let $\Gamma(t)$ with $t \in [0,1]$ be a curve in S with l arcs. If Γ is a Jordan closed curve we are done. Assume then that Γ is not a Jordan closed curve. We shall find a contradiction.

If Γ is not a Jordan closed curve, then $l \geq 2$, since it must have at least two arcs. In addition all nontrivial curves (that do not use arcs more than once) must have at least l arcs. Consider a basepoint $P_0 (= \Gamma(0))$ of Γ, and begin following the curve. The curve must intersect itself for the first time at some point $P_1 = \Gamma(b)$, with $0 < b < 1$ (the value of b cannot be one since then the curve would not have intersection points). But since P_1 was reached before, $P_1 = \Gamma(a)$, with $0 \leq a < b < 1$. ($a = 0$ corresponds to $P_0 = P_1$, which is the case when the basepoint is an intersection point.) Then the subcurve $\Gamma'(t) = \Gamma(t)$, $t \in [a,b]$, is a Jordan closed curve.

How many arcs does Γ' have? It may have one, or more, if additional intersection points (occurring later as one traces the curve) happen to lie in the interval traced by $t \in [a,b]$. At any rate it must have less than l arcs since it does not include the arcs that lie in the parameter region $t \in [b,1] \cup [0,a]$. It must therefore be a trivial subcurve. It can then be "homotopy away," and therefore the original curve Γ is homotopic to the curve Γ''' consisting of the remaining part of the curve when Γ' is deleted, namely it consists of $\Gamma(t)$ for $t \in [0,a] \cup [b,1]$. Since $\Gamma \approx \Gamma'''$, the curve Γ''' must be nontrivial. This is not possible, however, since Γ''' has fewer than l arcs (since Γ' has at least one arc). We have reached the desired contradiction and therefore established the lemma. □

2.3. Nontrivial Curves on Polyhedra

We now turn to discuss some nontrivial curves in string diagrams defined by quadratic differentials. Some of the material presented below may be considered as

---

[1] The main idea for the following proof is due to R. Forman and M. Wolf. I thank them for their help

detailed explanation of some facts used in [Zw 1], and some of it will also be of utility in analyzing open string diagrams in [Zw 4].

Given a closed string diagram defined by a meromorphic quadratic differential with closed trajectories, we can pick a closed trajectory from every characteristic ring domain. Each of these curves are nontrivial homotopy Jordan curves, are not homotopic to each other and can be chosen not intersect each other [St]. The surface, of course, must be punctured at the poles of the differential. The critical graph of the quadratic differential may be disconnected; each connected component is a polyhedron [SaZw, KKS, Zw 1]. A polyhedron joins various ring domains of a quadratic differential, and is thought of as representing an interaction vertex in the field theory of closed strings.

We now want to explain why *a closed Jordan edge-path on a polyhedron is a nontrivial curve*. This is just a closed critical path on the surface having no self-intersections (its nontriviality was used in [Zw 1]). Suppose we cut the surface along this curve. If this operation does not separate the surface into two pieces, the curve must be nontrivial (since a trivial Jordan curve separates the surface into a disk and another piece). Suppose then that the curve separates the surface into two pieces. Let the polyhedron on which this curve is defined join n ring domains together ($n \geq 1$). Since the curve goes along a critical path it cannot separate a ring domain into two parts, any ring domain appearing in either of the separate surfaces must appear fully in it. Therefore, since the curve separates locally, it cannot go through a critical edge that has the same ring domain to its two sides. Thus, when it separates, the two resulting surfaces divide the ring domains that converged into the polyhedron, each surface having at least one of the ring domains (since the edge-path has at least one edge). Now we show that neither of the two surfaces can be a disk. Each contains at least one ring domain and the core curve of that ring domain was a nontrivial curve before the surface was separated. Once the surface is cut it must remain nontrivial; if it became trivial it would mean that it bounds a disk in the cut surface, and therefore, also on the original surface, in contradiction with the fact that it was nontrivial to begin with. Thus each of the separate surfaces contains a nontrivial closed Jordan curve, and cannot be a disk. This proves the curve that separated the surface is not trivial.

A case that will be of interest deals with the slightly more complicated curve on polyhedra shown in Fig. 1. The closed edge path is made of two originally disjoint closed edge-paths, (that because of the above argument are nontrivial) joined together by a path that traces back and forth along some edges. This path tracing back and forth may even be of zero length. The resulting edge-path is not a Jordan curve, but a small deformation makes the path into one without self-intersections. This closed path must also be nontrivial, because if it separates it must still carry at least one ring domain into each of the separate surfaces (note, however, that along the repeated edges the same ring domain may appear to both sides). The nontriviality of this type of curves will be used in [Zw 4].
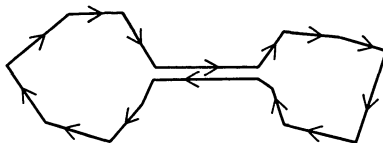

Fig. 1. A closed path on a polyhedron made up by joining two nontrivial Jordan closed edge-paths through a set of edges that are repeated. This closed curve is shown to be of nontrivial homotopy

2.4. Nontrivial Open Curves

In our minimal area problems, we impose conditions on non-trivial homotopy open curves. All open curves must have endpoints on the boundaries of the Riemann surface. The question of non-triviality is whether the curve can be shrinked away keeping the endpoints at the boundary. The endpoints are free to move on the boundary but they cannot slide through punctures lying on the boundary. Such punctures divide boundary components into segments. This criterion for deciding whether or not open curves are trivial is that of homotopy relative to boundary segments:

Definition. Two open curves $A(t)$ and $B(t)$ with $t \in [0, 1] \equiv I$, are homotopic relative to boundary segments, denoted $A \approx B$ (rel. bound.), if $A(0)$ and $B(0)$ are on the same boundary segment Γ_0; $A(1)$ and $B(1)$ are on the same boundary segment Γ_1; and there is a "homotopy $\psi_s(t)$" (a continuous map from $(s, t) \in I \times I$ into the surface) such that $\psi_0(t) = A(t)$, $\psi_1(t) = B(t)$; $\psi_s(0) \in \Gamma_0$, and $\psi_s(1) \in \Gamma_1$.

Namely, there is a homotopy that interpolates between the two curves keeping the endpoints in their respective boundary segments. It then follows that trivial curves must be defined as follows.

Definition. An open curve with boundary endpoints is trivial if it is homotopic (rel. bound.) to the open curve $w(t) = P_0$, $t \in I$, where P_0 is a fixed point on a boundary.

It is clear that an open curve whose endpoints lie on disconnected boundary components is non-trivial. Similarly, an open curve whose endpoints lie on different segments of the same boundary cannot be trivial. It is also useful for our purpose to define homotopy for open curves with *fixed* endpoints (which may or may not lie on a boundary).

Definition. Two open curves $A(t)$ and $B(t)$, with $t \in I$, are homotopic with fixed endpoints, denoted $A \approx B$ (fixed endp.), if $A(0) = B(0)$, $A(1) = B(1)$, and there is a homotopy $\psi_s(t)$ such that $\psi_0(t) = A(t)$, $\psi_1(t) = B(t)$, and $\psi_s(0) = A(0)$, $\psi_s(1) = A(1)$, for all $s \in I$.

That is, the two curves must have the same endpoints and be continuously deformable into each other. Let us conclude this section by discussing an alternative definition for a trivial open curve (with boundary endpoints). An open curve A with endpoints P_0 and P_1 on the same boundary component determines two segments s_1 and s_2 on the boundary both beginning at P_0 and ending at P_1. The open curve A is trivial if and only if it is homotopic to s_1 with s_1 free of punctures, or, it is homotopic to s_2 with s_2 free of punctures, or both. Namely:

Lemma 2. *An open curve A on a Riemann surface with boundaries is trivial if and only if it is homotopic (fixed endp.) to a puncture-free boundary segment determined by the curve.*

Clearly if A is homotopic (fixed endp.) to such a boundary segment it is homotopic (rel. bound.) to that segment. The segment, however is homotopic (rel. bound.) to a point in the boundary segment. It follows that A is homotopic (rel. bound.) to a point in the boundary, thus A is trivial. If A is trivial, by definition it is homotopic (rel. bound.) to a point on the boundary segment that contains the two endpoints. But this point is then homotopic (rel. bound.) to the segment joining the endpoints lying on the boundary segment.

3. Relating Open and Closed Curves

In this section we establish a few useful results relating open and closed curves. We say an open curve is a Jordan curve if it has no self-intersections, and if only its endpoints lie on boundary components. Let us begin with two simple lemmas.

Lemma 3. *Consider a closed curve $\Omega(t)$, $t \in I$, and fix two points on the curve, $Q_1 = \Omega(t_1)$, and $Q_2 = \Omega(t_2)$, with $0 < t_1 < t_2 \leq 1$. Then, the open curve extending from Q_1 to Q_1 can be replaced by any other open curve homotopic to it (fixed endp.) without changing the homotopy type of the closed curve.*

This is easily proven by writing a homotopy for the two closed curves using the homotopy that relates the two open curves. This result will be useful in Sect. 5. Consider now the following property of trivial open curves.

Lemma 4. *Consider an open Riemann surface R and a trivial open curve A on it. Glue this surface to its mirror image R^* and denote the doubled surface by $\hat{R}$. Let A^* denote the open curve which is the mirror image of A. Then $A \approx A^*$ (fixed endp.) on the surface $\hat{R}$.*

The proof goes as follows. Since A is trivial there is a (fixed endp.) homotopy $\psi_s(t)$ that relates A to a boundary segment Γ_i whose endpoints are those of A, and that is free of punctures. Thus $A \approx \Gamma_i$ (fixed endp.). Consider the antiholomorphic map $\mathscr{C}$, which is the automorphism of $\hat{R}$ that exchanges R and R^*. It follows that $\mathscr{C} \circ \psi_s(t)$ is a (fixed endp.) homotopy between A^* and Γ_i, thus $A^* \approx \Gamma_i$ (fixed endp.). As a consequence $A \approx A^*$ (fixed endp.).

The main results we want to establish concern the fate of nontrivial open curves on a surface with boundaries, when the surface is glued to its mirror image. There are several important properties, which we have proven in the following lemmas. They all refer to an open Riemann surface R, its mirror image R^* and the doubled surface $\hat{R}$.

Lemma 5. *A nontrivial homotopy Jordan closed curve in R is a nontrivial homotopy closed curve in $\hat{R}$ (this applies in particular to the boundary components of R).*

Proof. Let Ω be a nontrivial closed curve in R. Suppose it is a trivial closed curve in $\hat{R}$. Then Ω must bound a disk D on the glued surface. Some of D must lie on R^*, since otherwise D would lie totally in R and Ω would be trivial in R. Since the boundary of D, namely Ω, lies totally in R, it follows that D must contain R^* completely. Since R and R^* are connected, R^* must be a subdisk in D. But if R^* is a disk, R, its mirror image, is also a disk, but there are no nontrivial closed curves on a disk. Thus, we have a contradiction and Ω must be nontrivial in $\hat{R}$ $\quad\square$

Lemma 6. *Let A be a Jordan open curve in R and let $\hat{A} = A \cup A^*$ be the Jordan closed curve obtained by joining A^*, the mirror image of A, to A. A is nontrivial if and only if $\hat{A}$ is nontrivial.*

Proof. Let us first show that A nontrivial implies $\hat{A}$ nontrivial. Consider a nontrivial open Jordan curve A, with boundary endpoints P_1 and P_2. Assume the two boundary endpoints are in the same boundary component Γ, which is a Jordan closed curve on $\hat{R}$. The curve $\hat{A}$ crosses Γ precisely in the two points P_1 and P_2, which divide Γ into two boundary segments. Assume $\hat{A}$ is trivial on $\hat{R}$. Then it must bound a puncture free disk D on $\hat{R}$. Since Γ crosses $\hat{A}$, and there are only two intersection points, one of the two boundary segments must lie entirely inside D

(Fig. 2a). It follows that A is homotopic to a puncture free boundary segment, and is therefore trivial on R. This is the desired contradiction.

Assume now the two boundary endpoints of A lie on different boundary components, $P_0 \in \Gamma_0$ and $P_1 \in \Gamma_1$. Thus $\hat{A}$ intersects Γ_0 once only (in P_0) and Γ_1 only once (in P_1). Assume $\hat{A}$ is trivial in $\hat{R}$. Then it bounds a disk D. But then the closed curve Γ_0 must lie partially in and partially out of D, since it intersects $\hat{A}$, which is the boundary of D. This is impossible since Γ_0 and $\hat{A}$ intersect only once (Fig. 2b). This is the desired contradiction. We have therefore established that A nontrivial implies $\hat{A}$ nontrivial.

Now the other case. If $\hat{A}$ is nontrivial it must imply that A is nontrivial. Suppose A is trivial. Then A together with a puncture free boundary segment Γ bound a disk D lying fully on R. It follows from the existence of the mirror map that A^* is also homotopic to Γ, and A^* and Γ bound a disk D^*, lying on R^*, which is the mirror image of D. The union of D and D^* is a disk $\hat{D}$ whose boundary is $A \cup A^* = \hat{A}$. But if $\hat{A}$, bounds a disk, $\hat{A}$ is trivial, in contradiction with our initial assumption. $\quad\square$
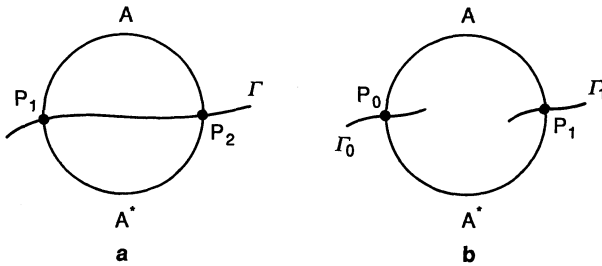

Fig. 2a, b. Understanding why the double $\hat{A}$ of a nontrivial Jordan open curve A is a nontrivial closed curve. The endpoints of A are P_0 and P_1, and may lie on the same boundary component (**a**), or in different boundary components (**b**)
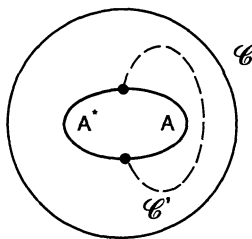

Fig. 3. Understanding why the nontrivial closed curve $\hat{A}$ cannot be homotopic to a boundary component $\mathscr{C}$ that it does not intersect

Lemma 7. *The curve $\hat{A}$ (of Lemma 6) is not homotopic, in $\hat{R}$, to any boundary component of R it does not intersect.*

Proof. The nontrivial curve $\hat{A}$ cannot be homotopic to a boundary component $\mathscr{C}$ of R that it does not intersect. Suppose it is, then $\hat{A}$ and $\mathscr{C}$ must bound a puncture free annulus, as shown in Fig. 3. The heavy dots on $\hat{A}$, represent the two intersections it must have with boundary components of R. These points divide $\hat{A}$ into A and A^*. Since the boundary components that cut $\hat{A}$ must enter the annulus, they must

actually join, making up the curve $\mathscr{C}'$. They cannot leave the annulus since they cannot intersect $\mathscr{C}$ (boundary components never intersect). But then, the two intersection points lie on a common boundary component $\mathscr{C}'$, and then A is homotopic to a puncture free segment of a boundary component, since A and $\mathscr{C}'$ bound a disk. This shows A is a trivial open curve, in contradiction with the original assumption, and establishes the desired result. $\square$

Lemma 8. *Let A and B be two nontrivial nonintersecting and nonhomotopic open Jordan curves in R. Then $\hat{A}$ and $\hat{B}$ (constructed by doubling) are two nontrivial, nonintersecting and nonhomotopic Jordan closed curves in $\hat{R}$.*
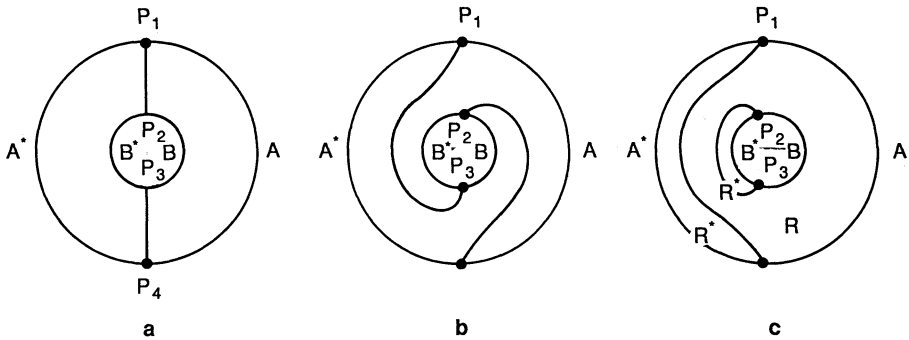

Fig. 4a–c. Showing that two nontrivial nonhomotopic and nonintersecting open Jordan curves A and B, lead, by doubling, to two nontrivial nonhomotopic and nonintersecting closed Jordan curves $\hat{A}$ and $\hat{B}$. The points P_1, P_2, P_3, and P_4 are the intersections of the curves with the boundary components. In the three cases shown here we assume $\hat{A} \approx \hat{B}$ and we find contradictions

Proof. The fact that $\hat{A}$ and $\hat{B}$ are nontrivial closed curves was shown in Lemma 6. They are clearly nonintersecting Jordan closed curves. We now have to show they cannot be homotopic. Assume they are, then they must bound a ring domain, as shown in Fig. 4a. Both $\hat{A}$ and $\hat{B}$ intersect the boundary of R two times each. Thus we have four intersection points that have to be joined by curves, representing parts of boundary components of R, that do not intersect each other. Note that the points, labeled P_1, P_2, P_3, and P_4, divide the curves $\hat{A}$ and $\hat{B}$ into their respective halves A and A^*, and B and B^*.

Let us first show that any other boundary component of R, that does not intersect $\hat{A}$ or $\hat{B}$ cannot show up in this annulus. Assume a boundary component shows up. Since it does not intersect either $\hat{A}$ or $\hat{B}$, it must be fully contained in the annulus. Since it cannot be homotopic to either $\hat{A}$ or $\hat{B}$ (Lemma 7), it must just bound a disk inside the annulus. This, however implies that the boundary component is a trivial closed curve, in contradiction with Lemma 5. Thus, no other boundary component can appear in the annulus.

We need, therefore, to concentrate only in the boundary components that intersect the curves. Point P_1 may be joined to P_2, and P_3 to P_4, as shown in Fig. 4a. If this is the case, it would follow that A^* and B^* are homotopic, since they are the opposite sides of a quadrilateral whose other two sides are boundary components. This is in contradiction with our original assumption. Note that one can perform Dehn twists, namely P_1 can join P_2 after going around the curve $\hat{B}$ n times (clockwise or counterclockwise) and similarly for the curve joining P_3 to P_4.

This does not change our conclusion, since Dehn twists do not alter the structure of the quadrilaterals.

The next possibility, that P_1 join P_3 and that P_2 join P_4, as shown in Fig. 4b, is also inconsistent. The region in the annulus near A^* must belong to R^*, and the region near B must belong to R. In this type of connection, however, these two regions are in the same quadrilateral. This is impossible, since R and R^* must always be separated by a boundary component.

The last possibility is that P_1 joins P_4, and that P_2 joins P_3. This also leads to contradictions. The joining can only be done in two different ways, with both curves to the left or to the right of the inner boundary of the annulus (otherwise the regions R and R^* are not separated properly). One of these cases is shown in the Fig. 4c. However, the pattern shows that A^* and B^* are trivial, since they are homotopic to puncture free boundary segments. This is a contradiction and concludes our proof of the lemma. $\square$

4. Minimal Area Problems

With all the definitions in hand, we can now discuss the minimal area problems posed in the introduction. In the open-closed problem, the Riemann surface has punctures both in the boundary components and in the interior of the surface. No curve can be moved through a puncture in the interior of the surface, and no endpoint can be moved through a boundary puncture. There are length conditions both on nontrivial Jordan open curves and on nontrivial Jordan closed curves. In the open string minimal area problem stated in the introduction there are only punctures in the boundary components of the surface. The main difference between open-closed and open string diagrams is that in the latter there are no conditions on the lengths of closed curves. In open string diagrams one can have closed curves that are very tiny and whose length can go to zero.

For the case of tree diagrams, where the relevant surfaces are disks with punctures on the boundary, there are no closed curves of non-trivial homotopy. Therefore open-closed and open string diagrams are identical. For higher genus this is not always the case. Any open string diagram that happens to have all its nontrivial Jordan closed curves longer than 2π will be also an open-closed diagram. But whenever there are closed curves shorter than 2π on the open string diagram, the open-closed diagram must be different.

Regularization of the Area. Since we will be dealing with surfaces with punctures that can lie on boundary components and will demand that non-trivial open curves should not be shorter than π, the area of the minimal area metric will be infinite and requires regularization. We need coordinates around the punctures in order to be able to subtract away the leading divergence. The case of punctures inside the surface was studied in [Zw 1]. For a puncture P_i on a boundary component Γ of the surface we must have a local coordinate z_i vanishing at the puncture $z_i(P_i) = 0$. The coordinate z_i must map a neighborhood of P_i in the surface into the upper half plane with the boundary going into the real axis. An important property of such a coordinate is that in the double surface one can define a local coordinate around P_i that transforms by complex conjugation under the antiholomorphic map $\mathscr{C}$ which sends the surface into its double keeping the boundary fixed pointwise. This means that if $\mathscr{C}$ takes $P \to \hat{P}$, then $z_i(\hat{P}) = [z_i(P)]^*$. Coordinates around boundary punctures will always be assumed to be of this kind.

Consider a Riemann surface R with boundary components, punctures P_i $(i, 1, 2, ..., m)$ on the boundaries and a metric ϱ (recall the length element dl is given by $dl = \varrho \, |dz|$). The case when there are also punctures that are not on the boundary adds no complication so it will not be treated explicitly in the following. Associate to every puncture a constant A_i and demand that non-trivial open curves homotopic to P_i be longer or equal to A_i. In order to satisfy this condition typically the metric ϱ near a puncture diverges as

$$\varrho(z_i) \approx \frac{A_i}{\pi} \frac{1}{|z_i|} \qquad (|z_i| \to 0). \tag{4.1}$$

In order to obtain a finite value for the area for metrics with such behavior one defines the reduced area $\mathscr{A}_R(\varrho, A_i)$ as follows:

$$\mathscr{A}_R(\varrho, A_i) \equiv \lim_{r \to 0} \left[\int\!\!\int_{R - \Sigma H_i(r)} \varrho^2 dx\, dy + \sum_i \frac{A_i^2}{\pi} \ln r \right], \tag{4.2}$$

where $H_i(r)$ denotes the half-disk $H_i(r) = \{|z_i| \leq r\} \cap \{\mathrm{Im}(z_i) \geq 0\}$ around the ith puncture, and the integral in the above expression extends over the surface R with the half-disks $H_i(r)$ removed. The second term subtracts away the leading divergences of the integral in the limit as $r \to 0$. As was the case for closed string punctures, the reduced area is dependent of the choice of coordinates around the punctures. Under a change of coordinates, however, the reduced area varies by a constant independent of the metric. Thus a metric of least reduced area is so for any choice of local coordinates.

The above definition leads to a simple behavior for the reduced area when the surface is doubled. For ordinary area, if a surface with boundaries and a finite area metric is glued to its mirror image, and the metric is extended via the map $\mathscr{C}$ it is clear that the area of the doubled surface is twice that of the original surface. We show next that the same holds for reduced area.

Consider the doubled surface $\hat{R}$ with a metric and coordinates defined by use of the map $\mathscr{C}$. To each puncture P_i we now associate the constant $2A_i$. The reduced area, (Eq. (2.1.8) in [Zw 1]) now reads

$$\mathscr{A}_{\hat{R}}(\varrho, 2A_i) = \lim_{r \to 0} \left[\int\!\!\int_{\hat{R} - \Sigma D_i(r)} \varrho^2 dx\, dy + \sum_i \frac{(2A_i)^2}{2\pi} \ln r \right], \tag{4.3}$$

where $D_i(r) = \{|z_i| \leq r\}$, are now full disks around the punctures in the doubled surface. It follows that

$$A_{\hat{R}}(\varrho, 2A_i) = 2\mathscr{A}_R(\varrho, A_i), \tag{4.4}$$

because our definition of the metric and the coordinates on the doubled surface is tailored to make the expression in brackets in Eq. (4.3) twice that of Eq. (4.2) for any value of r. Thus the reduced area of the doubled surface obtained by extending coordinates and metric with the antiholomorphic map is twice that of the original surface with boundaries.

5. Relating Open-Closed and Closed String Diagrams

In this section our aim is to establish the relation between the open-closed string diagrams, and the closed string diagrams. A proof of Theorem 1 will be given, thus

showing that the double of an open-closed string diagram is a closed string diagram, and that a closed string diagram for a surface $\hat{R}$, which is the double of an open surface R, induces on R an open-closed string diagram. In order to establish this result we will proceed in stages. Two preliminary results are needed.

Lemma 9. *Consider an open Riemann surface R and a metric (not necessarily of minimal area) satisfying the length conditions of the open-closed problem. On the doubled surface with a metric extended by the map $\mathscr{C}$ every non-trivial Jordan closed curve is longer or equal to 2π.*

Proof. Denote the two components of the doubled surface exchanged by the map $\mathscr{C}$ by $(+)$ and $(-)$. Any non-trivial closed curve that remains in one of the two sectors is automatically longer or equal to 2π. We need only concern ourselves with the closed Jordan curves that cross the boundary components separating the $(+)$ and $(-)$ regions.

The intersection points of the curve with the boundaries divide the curve in segments. There are two exclusive possibilities: (a) at least one segment goes from one boundary component to another boundary component, and (b) all segments being and end on the same boundary component. Case (a) is easily taken care of. Suppose the boundary segment goes from boundary component Γ_1 to Γ_2. Since the complete curve is closed there must be at least another segment from Γ_2 to another boundary component (perhaps Γ_1). There are, therefore, at least two segments going from one boundary component to another. Such segments are non-trivial Jordan open curves; as a consequence each is longer or equal to π and the original closed curve must be longer or equal to 2π.

Consider now case (b), where the endpoints of all segments lie on the same boundary component. A segment is nontrivial if it is not homotopic to a puncture free boundary segment. There are three mutually exclusive possibilities: (b0) no segment is nontrivial,[2] (b1) one segment is nontrivial, and (b2) two or more segments are nontrivial. The idea is to force the curve to lie on a single component of the doubled surface. Since any trivial segment is homotopic (fixed endp.) to its image under $\mathscr{C}$ (Lemma 4), a trivial segment can be replaced by its image without changing the length or the homotopy type of the closed curve (Lemma 3). The resulting nontrivial closed curve may have self-intersections, Lemma 1 guarantees it must be longer or equal to 2π, thus establishing that the original curve is also longer or equal to 2π. Consider the case (b0): here we just reflect all the $(-)$ segments the whole curve lie on the $(+)$ component. In case (b1) pick the nontrivial segment, say it lies on the $(+)$ component, then flip all the $(-)$ segments. Finally, in case (b2) there is no need to flip segments. Each nontrivial segment is guaranteed to be or equal to π. With two such segments the closed curve is assured to be longer or equal to 2π. $\square$

Lemma 10. *Consider a closed string diagram on a Riemann surface $\hat{R}$, which is the double of an open Riemann surface R. Then*
  (i) *The metric on $\hat{R}$ is invariant under the antiholomorphic automorphism $\mathscr{C}$ of $\hat{R}$.*
  (ii) *On the open surface R the induced metric satisfies the length conditions of the open-closed problem.*
(iii) *The metric induced on R defines an open-closed string diagram.*

---

[2] Such curve need not be trivial, an example is a closed curve in $\hat{R}$ homotopic to the boundary component in question, and crossing it an even number of times

Proof. Part (i) holds because of uniqueness of minimal area metrics [St, Zw 1]. If the metric was not invariant, the automorphism would yield a different metric satisfying all the length conditions and with the same area. This is impossible. Consider now (ii): Any non-trivial Jordan open curve on R can be reflected into the double in order to form a Jordan closed curve. This closed curve must be non-trivial (Lemma 6) and therefore longer or equal to 2π. As a consequence, the open curve must be longer or equal to π. Any nontrivial closed Jordan curve on R is a nontrivial Jordan closed curve on $\hat{R}$ (Lemma 5), and therefore must be longer or equal to 2π. Finally, consider (iii). Let $2A$ be the reduced area of the surface $\hat{R}$, as computed with the help of suitable fixed local coordinates around the punctures, for the minimal area metric. Due to Eq. (4.4), the reduced area of R is A. Assume the metric induced on R is not of minimal area. Then there is another metric on R with lower reduced area $A' < A$ satisfying the open-closed length conditions. Because of Lemma 9 if this metric is extended to the double it would satisfy the length conditions on closed curves and would have reduced area $2A' < 2A$, in contradiction with the assumption that the original metric was of minimal area. This establishes (iii) and concludes our proof of the lemma. $\square$

It is now straightforward to establish the following result.

Lemma 11. *Given an open-closed string diagram on a Riemann surface R, double the surface and extend the metric to the double $\hat{R}$ using the map $\mathscr{C}$. On $\hat{R}$ the resulting metric defines a closed string diagram.*

Proof. Let A be the reduced area of R. The doubled surface $\hat{R}$ will have reduced area $2A$, and due to Lemma 9, all non-trivial closed curves will be longer or equal to (2π). Assume this is not the metric of minimal area. Then the metric of minimal area will have reduced area $2A' < 2A$, and as a consequence of Lemma 10 (iii), it induces on R a metric of minimal area, with area [Eq. (4.4)] $A' < A$, in contradiction with the assumption that the original metric on R was of minimal area. $\square$

The above result says that the double of an open-closed string diagram is a closed string diagram. Moreover a closed string diagram on the double of a surface R induces on R an open-closed string diagram (Lemma 10 (iii)). All in all this shows that a string diagram on an open surface R is an open-closed string diagram if and only if its double on the doubled surface $\hat{R}$ is a closed string diagram. This is the content of Theorem 1.

References

[BS]   Bluhm, R., Samuel, S.: Nucl. Phys. B **323**, 337 (1989)

[GMW]   Giddings, S., Martinec, E., Witten, E.: Modular invariance in string field theory. Phys. Lett. **176B**, 362 (1986)

[GW]   Giddings, S., Wolpert, S.: A triangulation of moduli space from light-cone string theory. Commun. Math. Phys. **109**, 177 (1987)

[KKS]   Kugo, T., Kunitomo, H., Suehiro, K.: Non-polynomial closed string field theory. Phys. Lett. **226B**, 48 (1989)

[KS]   Kugo, T., Suehiro, K.: Nonpolynomial closed string field theory: action and gauge invariance. Nucl. Phys. B **337**, 434 (1990)

[Og]   Ogura, W.: Combinatorics of strings and equivalence of Witten's string field theory to
          Polyakov's string theory. Prog. Theor. Phys. **79**, 936 (1988)
[Sa]   Samuel, S.: Solving the open bosonic string in perturbation theory. Nucl. Phys. B **341**,
          513 (1990)
[SaZw]   Saadi, M., Zwiebach, B.: Closed string field theory from polyhedra. Ann. Phys. **192**, 213
          (1989)
[St]   Strebel, K.: Quadratic differentials. Berlin, Heidelberg, New York: Springer 1984
[Wi]   Witten, E.: Noncommutative geometry and string field theory. Nucl. Phys. B **268**, 253
          (1986)
[Zw 1]   Zwiebach, B.: How covariant closed string theory solves a minimal area problem.
          Commun. Math. Phys. **136**, 93 (1991); Consistency of closed string polyhedra from
          minimal area. Phys. Lett. B **241**, 343 (1990)
[Zw 2]   Zwiebach, B.: Quantum closed strings from minimal area. Mod. Phys. Lett. A **2**, No.
          32, 2753 (1990)
[Zw 3]   Zwiebach, B.: Quantum open string theory with manifest closed string factorization.
          Phys. Lett. B **256**, 22 (1991); The Covariant Open-Closed String Theory I, II. MIT
          preprints, MIT-CTP-1909, 1910, to appear
[Zw 4]   Zwiebach, B.: A proof that Witten's open string theory gives a single cover of moduli
          space. Commun. Math. Phys. **142**, 193–216 (1991)
[Zw 5]   Zwiebach, B.: Interpolating open-closed string field theories. MIT preprint, MIT-
          CTP-1911 (to appear)

Communicated by S.-T. Yau