

A New Formulation of String Theory

D. Friedan

Enrico Fermi and James Franck Institutes and Department of Physics, University of Chicago, Chicago, Illinois 60637, U.S.A.

Received August 29, 1986

Abstract

Quantum string theory is written as integrable analytic geometry on the universal moduli space of Riemann surfaces.

1. Introduction

In this lecture I describe recent work of Stephen Shenker and myself, reformulating two dimensional conformally invariant quantum field theory [1] and string theory [2]. This work grew out of a line of investigation of string theory whose beginnings were reported in Ref. [3]. Our main goal is to express string theory in an abstract geometric language which makes no reference to spacetime, motivated by the expectation that string theory, as a theory of quantum gravity, will produce spacetime dynamically.

String theory is interesting, despite any direct experimental contact with reality, because it is the only available theory which seems both self-consistent and capable of exhibiting the basic features of the low energy world: quantum mechanics, gravity, gauge forces, light fermions, chiral asymmetry and so on. But string theory as presently understood is an incomplete theory. It is limited to perturbation theory in weakly curved spacetimes; it requires an *a priori* choice of (partially compactified) spacetime and it permits too many such choices.

There is apparently no direct experimental data on the process of string compactification. For the moment, we are forced to do what Nambu has called *postmodern physics* [4]. In such a remarkable circumstance, having an incomplete theory so far in advance of experiment, the most promising strategy is to investigate the formal structure of the existing theory so as to expose its most fundamental properties, in the hope that this will lead to effective methods of extracting predictions from the theory and perhaps even to new ways of doing theoretical physics. We should look for formal models which might provide guidance, playing a role in this post-modern physics analogous to that performed, for example, by superconductivity in electro-weak theory.

We would especially like to rewrite string theory so as to have some hope of being able to control it mathematically, as an integrable quantum system, for example. Debatable prejudices about the goal of physics aside, it is becoming difficult to imagine explaining the smallness of the cosmological constant without such precise mathematical control.

The crucial constraint in this enterprise is that an abstract reformulation must exactly reproduce the mathematical structure of the existing recipe for perturbatively calculating the string *S*-matrix in weakly curved spacetimes by integrating the partition function of the conformally invariant string world surface over the moduli space of all Riemann surfaces [5].

A traditional approach to nonperturbative string theory starts by writing a classical field theory of string and then attempts semi-classical calculations. We have not taken this approach, although string field theory was useful as a trial ground for thinking about the abstract structure of string theory [6]. The main drawback of string field theory is that it turns away from the most beautiful aspect of string theory, duality, which can be interpreted as conformal invariance of the string world surface. A second drawback of string field theory is that it makes arbitrary and unnecessary extrapolation from the on-shell content of string theory to obtain an off-shell formulation. This is unnecessary because string theory is a self-contained, potentially complete theory of physics.

Our starting point is the equivalence between the string equation of motion and the condition of conformal invariance on the two dimensional quantum field theory of the string world surface. Solving the string equation of motion is analogous to the problem of finding all two dimensional conformal field theories [7–10]. In particular, we take guidance from the unitary discrete series [9, 10] of conformal field theories which describe universality classes of two dimensional critical phenomena. We learn from their properties how to reformulate two dimensional conformal field as analytic geometry on the space of Riemann surfaces [11]. A geometric formulation of perturbative string theory then follows immediately. As the setting we define the *universal moduli space* of Riemann surfaces, written \bar{R} . The universal moduli space contains all compact but not necessarily connected Riemann surfaces, without punctures or boundaries. \bar{R} is an effectively compact and connected analytic space, so its global analytic geometry is highly constrained. The fundamental object of abstract string theory is a hermitian metric h_{ab} in an infinite dimensional holomorphic vector bundle W over \bar{R} . The covariant derivative D associated with the metric h is defined by $Dh = 0$ or $\partial h_{ab} - h_{ac}A_b^c = 0$ (a covariant derivative in W has the form $D = \partial + A$, $\bar{D} = \bar{\partial}$ where A is the connection (1, 0)-form). The quantum equation of motion of string is the condition of zero curvature:

$$F = [\bar{D}, D] = \bar{\partial}A = 0. \quad (1)$$

This means that, in principle, quantum string theory is integrable.

To do nonperturbative string theory it will be necessary to construct a completion of the universal moduli space. Making such a completion might be thought of as adding to the moduli space some appropriate class of infinite genus surfaces. The integrable form of the abstract string equation of motion (1) will remain unchanged. Solving string theory will then be reduced, in principle, to algebra. The hardest problem will be to extract useful information from an abstract solution; to find, for example, a concrete semi-classical interpretation in terms of spacetime geometry. We

must hope that the integrability of the theory will allow some encouraging exact results in advance of a complete solution. For example, we might imagine using topological properties of the completion of universal moduli space to demonstrate nonperturbatively the exact vanishing of the cosmological constant.

For the present, we have expressed as analytic geometry on universal moduli space only the bosonic string perturbation theory and ordinary conformal field theory on finite genus Riemann surfaces. The perturbative bosonic string theory is only formal, because the quantum corrections are divergent. Supersymmetric string theories, on the other hand, are expected to have finite perturbative expansions. But formal study of the bosonic theory provides all the basic ideas we need. To write fermionic string theory as analytic geometry, it is only necessary to replace two dimensional conformal field theory by two dimensional superconformal field theory, and ordinary Riemann surfaces by super Riemann surfaces; the formal structure remains the same. The bosonic string theory as integrable analytic geometry on universal moduli space will be transcribed directly into fermionic string theory as integrable super analytic geometry on the universal super moduli space of super Riemann surfaces [12].

This reformulation of conformal field theory has potential applications beyond string theory, because two dimensional conformal field theory is the underlying calculus of several areas in physics and mathematics, including two dimensional critical phenomena, the representation theory of affine algebras and sporadic groups and the classical geometry of Calabi–Yau spaces. These can now be seen as aspects of a single subject – integrable analytic geometry on the universal moduli space of (super) Riemann surfaces

While we were carrying out this work, several groups were investigating string theory in flat spacetime in terms of the analytic and algebraic geometry of moduli spaces of Riemann surfaces [13] and Martinec was discussing several ideas closely connected to aspects of the present work [14].

2. The string S-matrix

The existing recipe for calculating perturbative string scattering amplitudes calls for integrating the partition function of a two dimensional conformal field theory over the space of all Riemann surfaces. This is a generalization of the recipe originally developed from the Veneziano formula and the Koba–Nielsen integral. It gives the perturbative S-matrix of strings propagating and scattering in some fixed background. In the original formulation of string theory the background was Minkowski spacetime. More generally, the background spacetime is some manifold M of dimension d .

In the first quantized formalism, the string S-matrix is calculated by performing a reparametrization invariant path integral over string histories, which are surfaces in spacetime. A natural gauge fixing procedure is to introduce an intrinsic metric on the world surface, and to impose the conformal gauge on the intrinsic metric. The conformal classes of surface metrics are exactly the Riemann surfaces. In the conformal gauge the string path integral factors into an integral over the space of Riemann surfaces and an integral over maps each Riemann surface into spacetime. Let $m = (m^1, \dots, m^n)$ be local analytic coordinates for the space of Riemann surfaces, the moduli space. Choose a local analytic coordin-

ate z on the surface m , so the intrinsic surface metric takes the form $ds^2 = e^{f(\bar{z}, z)} |dz|^2$. A surface in spacetime is written $x^\mu(\bar{z}, z)$. For each Riemann surface m , the integral over maps into spacetime has the form

$$Z_{st}(\bar{m}, m) = \int Dx e^{-A[x]} \quad D = \prod_{z \in m} d^d x(\bar{z}, z), \quad (2)$$

where $A[x]$ is a local functional of the map $x^\mu(\bar{z}, z)$. The most general local action functional has the form

$$A[x] = \int_m d\bar{z} dz \left[\frac{1}{T} g_{\mu\nu}(x) \bar{\partial} x^\mu \partial x^\nu + \dots \right]. \quad (3)$$

The functional integral (2) is the partition function of the general two dimensional nonlinear model, the two dimensional quantum field theory whose action is $A[x]$. The spacetime metric $T^{-1}g_{\mu\nu}(x)$ appears as a coupling in the two dimensional field theory. The infinitely many couplings indicated by the ellipse are tensor fields on spacetime, of higher and higher rank as more and more derivatives of $x^\mu(\bar{z}, z)$ appear in the action. The couplings in the surface action $A[x]$ express the background in which the strings move and interact. This is the only point in the calculation of the S-matrix where the background enters.

In conformal gauge the string partition function also includes a contribution from the surface ghost fields:

$$Z(\bar{m}, m) = Z_{gh}(\bar{m}, m) Z_{st}(\bar{m}, m). \quad (4)$$

The surface ghost fields are free chiral fermions $b(z)$, $c(z)$ and $\bar{b}(\bar{z})$, $\bar{c}(\bar{z})$. The ghost field c represents infinitesimal reparametrizations of the surface and the field b represents infinitesimal variations of the conformal gauge slice. The zero modes of b correspond to the nontrivial variations of the gauge slice, which are the infinitesimal variations in moduli space. Since there is no coupling between the two chiralities, except through the conformal anomaly, the ghost partition function is formally

$$Z_{gh}(\bar{m}, m) = \left| \int DbDc e^{-\int_m b\bar{\partial}c} \right|^2. \quad (5)$$

Because of the zero modes of the b field, the string partition $Z(\bar{m}, m)$ transforms as a density on the space of Riemann surfaces. The integral of the partition function over the space of Riemann surfaces gives the S-matrix. Let M_g be the moduli space of smooth, compact, connected Riemann surfaces of genus g , without boundary or punctures [11]. The generating functional for the connected part of the S-matrix is

$$S_{\text{conn}} [T^{-1}g_{\mu\nu}(x), \dots] = \sum_{g=0}^{\infty} \int_{M_g} \lambda^{2g-2} Z(\bar{m}, m) \quad (6)$$

where λ is the string coupling constant. From now on we absorb the factor λ^{2g-2} into the partition function.

The functional S_{conn} gives the string S-matrix because its variations with respect to the couplings $[(1/T)g_{\mu\nu}(x), \dots]$ are generalized Koba–Nielsen formulas for scattering amplitudes as integrals over the locations of vertex operators and over the moduli space. For example, varying the spacetime metric gives the integrated vertex operator

$$\delta A[x] = \int d^2z \delta g_{\mu\nu}(x) \bar{\partial} x^\mu \partial x^\nu. \quad (7)$$

In flat spacetime, the Fourier transform is the familiar form of the vertex operator:

$$\delta g_{\mu\nu}(x) = \int dk \xi_{\mu\nu}(k) e^{ik \cdot x}. \quad (8)$$

Let us note a few basic facts about the moduli spaces M_g . For a connected Riemann surface, the genus g is an integer which counts the number of handles. M_0 is a single point, the Riemann sphere. M_1 is the space of tori, a branched cover of the complex plane. For $g > 1$, M_g is a noncompact complex analytic space of complex dimension $n = 3g - 3$. The moduli spaces M_g are almost everywhere smooth manifolds. But for $g > 0$ each M_g contains a finite number of singular points where smoothness breaks down. The singularities are not especially pathological; at worst they are only corners in the moduli space. The singular points are the surfaces with discrete conformal symmetries. Technically, the moduli spaces M_g are V -manifolds, sometimes called orbifolds. At its generic smooth points M_g looks like a complex vector space C^n . At the singular points it looks like C^n modulo the action of a finite group of linear transformations.

So far we have described the connected part of the S -matrix in terms of a sum of integrals over the various moduli spaces M_g . As a motivation for our eventual definition of the universal moduli space of Riemann surfaces, we write the generating functional for the full S -matrix as an integral over a single space:

$$S = e^{S_{\text{conn}}} = \int_R Z(\tilde{m}, m) \quad (9)$$

where R is the space of all compact, smooth Riemann surfaces, without boundary or punctures, but not necessarily connected. The derivation of eq. (9) from eq. (6) is a matter of elementary combinatorics. The moduli space R consists of infinitely many disconnected components indexed by multiplicities $\{n_g\}$, where $n_g \geq 0$ is the number of connected components of genus g in the surface:

$$R = \prod_{g=0}^{\infty} \left(\bigcup_{n=0}^{\infty} \text{Sym}^n(M_g) \right) = \bigcup_{\{n_g\}} \left(\prod_{g=0}^{\infty} \text{Sym}^{n_g}(M_g) \right) \quad (10)$$

where $\text{Sym}^n(M_g)$ is the n -fold symmetric product of M_g with itself, and $\text{Sym}^0(M_g)$ is a single point. This symmetrization is due to the indistinguishability of the various connected components of a surface. If $m = m_1 \cup m_2 \cup \dots \cup m_N$ is a surface with connected components m_1, m_2, \dots then the partition function of m is

$$Z(\tilde{m}, m) = \prod_{i=1}^N Z(\tilde{m}_i, m_i). \quad (11)$$

The integral of the partition function over each connected component of R is therefore

$$\int_{\prod_{g=0}^{\infty} \text{Sym}^{n_g}(M_g)} Z(\tilde{m}, m) = \prod_{g=0}^{\infty} \frac{1}{n_g!} \left(\int_{M_g} Z(\tilde{m}, m) \right)^{n_g}. \quad (12)$$

Summing over the multiplicities $\{n_g\}$ gives eq. (9) for the generating functional of the full S -matrix.

3. String theory and conformal field theory

The recipe for the string S -matrix makes sense only if the string partition function is well-defined on the space of Riemann surfaces. This means that it must be independent of the local scale factor $e^{f(z,z)}$ in the two dimensional metric, which is exactly the condition of anomaly free conformal invariance

on the two dimensional quantum field theory of the world surface. Conformal invariance in quantum field theory is equivalent to the vanishing of the trace of the stress-energy tensor, up to a possible conformal anomaly. The conformal anomaly in two dimensions is proportional to the scalar curvature of the surface. On the plane or on the cylinder the stress-energy tensor is traceless.

Locality implies that conformal invariance is equivalent to scale invariance which in turn is equivalent to the vanishing of the renormalization group β -function, since the β -function is the infinitesimal variation of the effective couplings of the field theory with respect to the two dimensional scale μ . In the general nonlinear model (2) the scaling behavior of the couplings $[(1/T)g_{\mu\nu}(x), \dots]$ can be calculated perturbatively in T , where $T \rightarrow 0$ is the limit of large distance in spacetime. The β -function for the spacetime metric is

$$\beta_{\mu\nu} = \mu \frac{\partial}{\partial \mu} \frac{1}{T} g_{\mu\nu} = R_{\mu\nu} + O(T). \quad (13)$$

The renormalization group fixed point equation $\beta_{\mu\nu} = 0$ thus becomes, at large distance in spacetime, Einstein's equation for the gravitational field. More generally, there is exactly one β -function for each coupling in the two-dimensional field action. In the limit $T \rightarrow 0$ the β -function for each coupling becomes a local differential operator on the corresponding spacetime tensor field. The equation of two dimensional conformal invariance thus has exactly the form of an equation of motion on the string background, where the background is encoded as the coupling of the two dimensional field theory.

In the operator interpretation of a conformal field theory on the cylinder, or on the conformally equivalent punctured plane, the moments of the traceless stress-energy tensor $T(z) = \sum_n z^{-n-2} L_n$, $\bar{T}(\bar{z}) = \sum_n \bar{z}^{-n-2} \bar{L}_n$, generate two commuting Virasoro algebras

$$[L_m, L_n] = (m - n)L_{m+n} + \frac{c}{12} (m^3 - m)\delta_{m+n,0}. \quad (14)$$

The number c is the coefficient of the conformal anomaly. The operator L_n generates the infinitesimal conformal transformation $z \rightarrow z + \epsilon z^{n+1}$, and \bar{L}_n generates the complex conjugate transformation.

The Virasoro algebra was the language in which conformal invariance was originally equated with physical meaningfulness of the string S -matrix. In string theory, the punctured plane is the world surface of a free string. Anomaly free conformal invariance implies gauge invariance of the string interaction under the local conformal transformations generated by the Virasoro algebra. This gauge invariance ensures the decoupling of spacetime ghosts to give a unitary S -matrix. The vanishing of the conformal anomaly in spacetime string theory fixes the dimension of spacetime;

$$c = c_{gh} + c_{st} = 0 \quad c_{st} = d \quad c_{gh} = -26. \quad (15)$$

Unitarity of the S -matrix can be interpreted as the absence of net flux of string into, or out of, the vacuum. Thus the string S -matrix is unitary if and only if the background is a solution of the quantum equation of motion. But string can be identified with its own background, since the scattering states of string can be identified with variations of the two dimensional couplings which encode the background. This is

one way of seeing that string theory is potentially a complete theory of physics. Since unitarity of the string S -matrix can be considered as the quantum equation of motion of string, as a condition on the background, and unitarity is equivalent to conformal invariance of the world surface, it follows that the string equation of motion is the condition of conformal invariance in two dimensional quantum field theory. What we gain by this rewriting of the string equation of motion is the possibility of abstracting it from spacetime. All we need do is replace the couplings of the classical two dimensional action (3) by some alternate parametrization of the two dimensional quantum field theory. We then interpret the integral (9) of the partition function as a functional of all possible variations of the background which preserve conformal invariance, giving the generating functional for an abstract generalization of the string S -matrix.

One possible abstract parameterization of conformal field theories is given by the algebra of two dimensional quantum fields. There already exists a program, called the conformal bootstrap, to find all possible algebras of quantum fields satisfying the condition of conformal invariance [7, 8]. The bootstrap idea is to attack two dimensional conformal field theory on the plane, or equivalently on the Riemann sphere, with first principles. Conformal invariance is expressed by the action of the Virasoro algebra, two dimensional locality by real analyticity of the correlation functions on the Riemann sphere, and two dimensional crossing symmetry by single-valuedness of the correlation functions (modular invariance on the punctured sphere). In string theory, correlation functions on the Riemann sphere translate into tree-level scattering amplitudes. Conformal invariance, two dimensional locality and two dimensional crossing symmetry are equivalent to basic physical properties of the tree-level string S -matrix: unitarity, spacetime locality, and duality. The conformal bootstrap describes a conformal field theory by the highest weights h_i, \bar{h}_i of the Virasoro algebra (the lowest L_0, \bar{L}_0 eigenvalues in each irreducible subrepresentation) and the operator product coefficients $C(ijk)$ of the primary conformal fields corresponding to the irreducible subrepresentations. It then attempts to translate the basic conditions of conformal field theory into algebraic conditions on the numbers h_i, \bar{h}_i and $C(ijk)$.

This program has several drawbacks as an abstract formulation of string theory. The weights and operator product coefficients do constitute a minimal set of ingredients for conformal field theory, but it has turned out to be extremely difficult to find the explicit algebraic bootstrap conditions. Also, the weights and operator product coefficients are very distant from the physically interesting two dimensional quantities, since they give only tree-level information about the string. Finally, this version of the conformal bootstrap is too mechanical and un-geometric to be aesthetically attractive.

4. The $c < 1$ discrete series

In the last few years it has become clear that the problem of describing all conformal field theories is much more tractable if $c < 1$ than in the general case [8–10]. In particular, the condition of unitarity (positivity of the metric) on representations of the Virasoro algebra restricts the possible values of $c < 1$ and the possible highest weights to a discrete series of rational numbers [9]. Recently it was shown that modular

invariance of the partition function in genus 1 puts stringent restrictions on the multiplicities of highest weights in the $c < 1$ discrete series [10]. The form of the genus 1 partition function and of the genus 1 modular invariance equation for the discrete series were highly suggestive to us. The $c < 1$ discrete series, whose physical application is the description of universality classes of two dimensional critical phenomena, can serve as a model setting from which to derive ideas about abstract string theory.

Genus 1 Riemann surfaces can be parametrized by points τ in the upper half plane $\text{Im } \tau > 0$. The torus parametrized by τ is the complex plane with each point w identified with $w + 1$ and $w + \tau$ [Fig. 1(a)]. Two such tori τ and τ' are the same Riemann surface if and only if they are related by a modular transformation $\tau' = (a\tau + b)/(c\tau + d)$ with a, b, c, d integers satisfying $ad - bc = 1$. The moduli space M_1 is thus the quotient of the upper half plane by the modular group Γ_1 . The modular group is generated by $T(\tau) = \tau + 1$ and $S(\tau) = -1/\tau$. The moduli space M_1 can be pictured by drawing the standard fundamental domain for Γ_1 in the upper half plane and identifying its edges by the transformations T and TS [Fig. 1(b)]. The points $e^{\pi i/3}$ and i on the boundary of the fundamental domain are the only two singular or orbifold points in M_1 . They are only fixed points of nontrivial modular transformations.

The partition function of a conformal field theory is the same for the two tori τ and τ' because the flat metrics on these tori, inherited from the complex plane, are related by a global scale factor $c\tau + d$. In the operator interpretation, writing $q = e^{2\pi i\tau}$, the modular invariant partition function is

$$Z(\bar{\tau}, \tau) = \text{tr} (q^{\varepsilon_0 + L_0} \bar{q}^{\varepsilon_0 + \bar{L}_0}). \quad (16)$$

The number $\varepsilon_0 = -c/24$ is the universal ground state energy of conformal field theory [15]. The appearance of ε_0 is explained by the fact that the Virasoro generators L_0 and \bar{L}_0 are normalized for the coordinate $z = e^{2\pi i w}$. In the punctured z plane, the torus τ is made by identifying z with qz . The shift of the ground state energy is due to the conformal anomaly for the transformation $w \rightarrow z$.

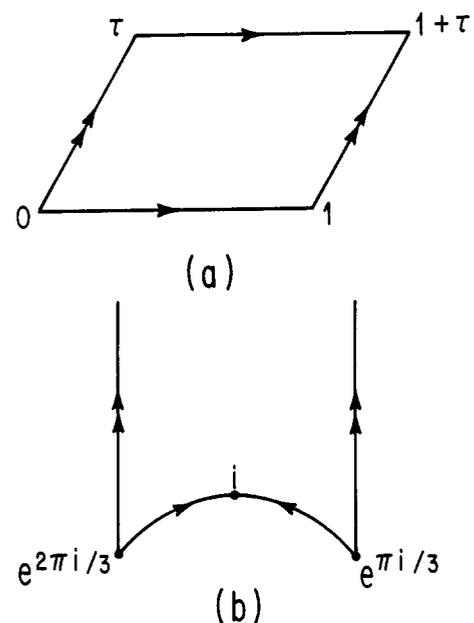


Fig. 1. (a) The torus parametrized by τ . (b) The standard fundamental domain of the genus 1 modular group Γ_1 acting on the upper half τ plane.

For c in the discrete series let $\{h_a\}$ be the finite set of highest weights allowed by unitarity and let

$$\chi^a(\tau) = \text{tr}(q^{\bar{L}_0 + L_0}) \quad (17)$$

be the character of the irreducible Virasoro representation of highest weight h_a . Explicit formulas have been found for these characters [16]. The partition function is written in terms of the characters by summing over the irreducible subrepresentations of the two Virasoro algebras:

$$Z(\bar{\tau}, \tau) = \bar{\chi}^{\bar{a}}(\bar{\tau}) h_{\bar{a}b} \chi^b(\tau) \quad (18)$$

where $h_{\bar{a}b}$ is the integer multiplicity of the highest weights $\bar{h} = h_{\bar{a}}$, $h = h_b$ in the conformal field theory, and the summation convention is applied to the indices \bar{a} , b .

Under the modular transformation T

$$\chi^a(\tau + 1) = e^{2\pi i(h_a + \bar{h}_a)} \chi^a(\tau) \quad (19)$$

so invariance of the partition function under T is the condition that $h_{\bar{a}b} = 0$ unless the spin $h_{\bar{a}} - h_b$ (the eigenvalue of the rotation operator $L_0 - \bar{L}_0$) is an integer. Under S , the characters transform linearly:

$$\chi^a(-1/\tau) = S_b^a \chi^b(\tau), \quad (20)$$

where S_b^a is a matrix of algebraic numbers. The characters $\chi^a(\tau)$ thus form a closed collection of multi-valued analytic functions on the moduli space M_1 . The modular invariance condition on the partition function (18) is now reduced to the invariance of the matrix $h_{\bar{a}b}$:

$$h_{\bar{a}b} = \bar{S}_a^{\bar{c}} h_{\bar{c}d} S_b^d. \quad (21)$$

This is a Diophantine equation because S_b^a is algebraic while $h_{\bar{a}b}$ is integer. For the first few values of c in the discrete series there seem to be very few solutions [10].

The genus 1 partition function offers several lessons. First, it is possible to learn much about conformal field theory by studying the partition function alone. Second, given the modular properties of the Virasoro characters, modular invariance of the partition function is a powerful principle. Third, the real analyticity of the partition function is obtained by writing it as a hermitian product of locally analytic, multi-valued functions on moduli space. Fourth, the single-valuedness or modular invariance of the partition function is obtained from the modular invariance of the hermitian metric.

The possibility of expressing conformal field theory through the partition function alone is very important for string theory. We already know that the string S -matrix can be written entirely in terms of the partition function. Moreover, we do not want to specify a spacetime for the theory. This means that we must avoid specifying in advance the Hilbert space of the two dimensional conformal field theory or the space of quantum fields. For a surface with boundary, the boundary conditions on each boundary circle are identified with the Schrodinger picture Hilbert space. Punctures are asymptotic forms of boundaries, where the boundary conditions introduce quantum fields at the punctures to give correlation functions. A correlation function can be thought of as the partition function of a surface with punctures. Therefore, to avoid knowing in advance the possible boundary conditions, we should work only with closed surfaces, without boundaries or punctures. We want to think of correlation functions as derivative quantities in conformal field

theory; the essential object is the partition function on all closed surfaces, without boundary or puncture.

5. The partition function and the universal moduli space

In order to generalize from genus 1 to higher genus it is useful to translate the structure of the genus 1 partition function into geometric language. The characters $\chi^a(\tau)$ form a linear representation of the modular group Γ_1 , generated by the actions of T and S described in eqs. (19) and (20) above. The modular group is the fundamental group of the moduli space M_1 . A representation of the fundamental group defines a flat holomorphic vector bundle W over M_1 , whose parallel transport matrices around closed loops are exactly the original representation. The characters $\chi^a(\tau)$ can be interpreted as the components, in a basis $\{e_a\}$ of locally constant sections of W , of a global holomorphic section $\chi = \chi^a(\tau)e_a$ of W over M_1 . The modular invariant matrix $h_{\bar{a}b}$ is the component matrix $h_{\bar{a}b} = h(\bar{e}_{\bar{a}}, e_b)$ of a globally defined flat hermitian metric h in W . Described invariantly, the partition function is $Z = h(\bar{\chi}, \chi)$.

This is the geometric structure which we generalize to higher genus. We will write the partition function for surfaces of arbitrary genus in the form

$$Z(\bar{m}, m) = \bar{\Psi}^{\bar{a}}(\bar{m}) h_{\bar{a}b} \psi^b(m) = h(\bar{\Psi}, \Psi). \quad (22)$$

(Recall that $m = (m^1, m^2, \dots)$ are local analytic coordinates on moduli space.) The functions $\psi^a(m)$ are multi-valued analytic functions of the moduli and the hermitian matrix $h_{\bar{a}b}$ is constant. Geometrically, the functions $\psi^a(m)$ are the components of a holomorphic section ψ of a holomorphic vector bundle W over moduli space and the constants $h_{\bar{a}b}$ are the components of a flat hermitian metric in W . Strictly speaking this is correct in genus $g > 1$ only for $c = 0$. For $c \neq 0$ the partition function $Z(\bar{m}, m)$ is a section of a real line bundle $\bar{E}_c \otimes E_c$, where E_c is a certain holomorphic line bundle over moduli space; ψ is a section of $E_c \otimes W$ and h is only projectively flat, i.e., its curvature tensor is an ordinary $(1, 1)$ form times the identity matrix in W . These subtleties are treated in [1]. Here we limit the discussion to conformal field theory with $c = 0$, since the combined spacetime-ghost system of the string has $c = 0$.

The crucial physical properties of the partition function are real analyticity and single-valuedness or modular invariance. They are obtained from the combination of the flatness of the hermitian metric h and the complex analyticity of the section ψ . Flatness means that there is a local basis of W in which the components of the metric are constant. In that basis, the complex analyticity of the components $\psi^a(m)$ guarantees that the partition function is real analytic. Then single-valuedness is guaranteed by the fact that h and ψ are globally defined over moduli space.

Flatness and analyticity are potentially effective geometric properties. In a connected space, a flat hermitian metric is equivalent to a unitary representation of the fundamental group. On a connected, compact space there are very few globally defined holomorphic sections of a vector bundle. For example, the only globally defined analytic functions are constant. Unfortunately, the moduli spaces M_g of connected Riemann surfaces are not compact and are not connected to each other. To make the geometry of conformal field theory effective we need a way of compactifying the moduli spaces

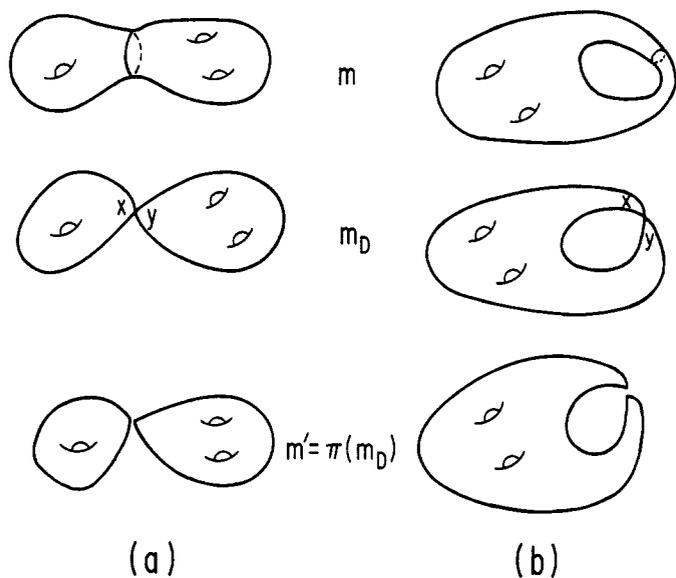


Fig. 2. (a) Forming a node in a disconnecting channel. (b) Forming a node in a handle.

and of connecting them together. Both goals are accomplished by introducing Riemann surfaces with nodes [11].

Two elementary processes change the connectivity of a surface (Fig. 2). These are the two possible ways of forming a node and then removing it. A node is a point where the surface has a coordinate neighbourhood consisting of two disks with their origins identified (Fig. 3). When the node is removed, the two coordinate disks are separated, and a smooth surface is left, except for punctures at two points x and y corresponding to the origins of the two coordinate disks. In the first process [Fig. 2(a)] a connected surface m of genus g develops a node in a channel (tube) which disconnects the surface. The surface with node is labelled m_D . When the node is removed the resulting surface m' is the union of two connected components, m_1 of genus g_1 and m_2 of genus g_2 , with $g = g_1 + g_2$. In the second process [Fig. 2(b)] the connected surface m of genus g develops a node which lies in a handle. Removing the node leaves a connected surface m' of genus $g - 1$.

Various combinations of the two elementary processes of node removal produce surfaces with multiple nodes (Fig. 4).

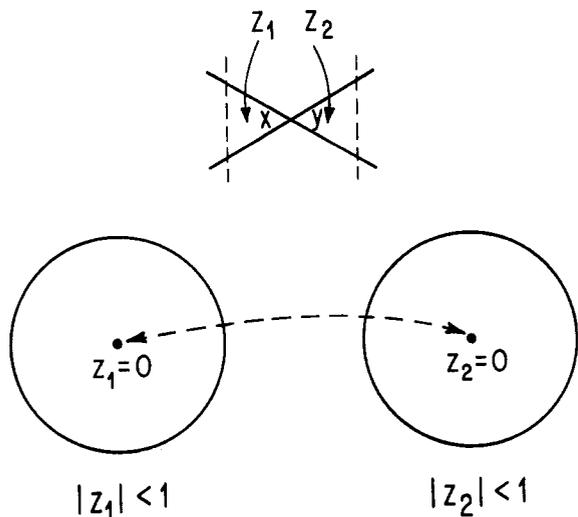


Fig. 3. A coordinate neighborhood of a node as two disks with their origins identified.

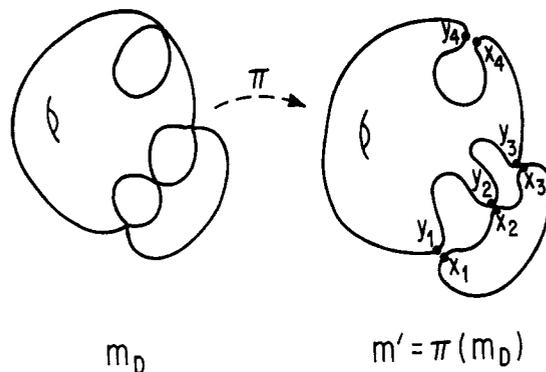


Fig. 4. A surface m_D with multiple nodes, and the corresponding smooth surface $m' = \pi(m_D)$ with the nodes removed.

Each node α has an independent coordinate neighborhood consisting of a double disk. Removing the node α leaves two punctures, at x_α and y_α . If all the nodes are removed from a surface m_D , and all the resulting punctures are erased, the result is a smooth surface we call $\pi(m_D)$ (Fig. 4). The surfaces m_D with nodes are parametrized by the moduli of the smooth surfaces $m' = \pi(m_D)$ and by the locations of the pairs of punctures (x_α, y_α) on m' which are joined to form the nodes: $m_D = (m', x_1, y_1, x_2, y_2, \dots)$.

M_g can be compactified by adding surfaces with nodes. The surfaces which are to be added are called the *stable* Riemann surfaces with nodes (connected, compact and of genus g). Stability, for a surface with nodes of genus $g > 1$, means that when all the nodes are removed, the resulting surface with punctures has no continuous conformal symmetries. That is, each connected component with the topology of the sphere in the resulting surface must have at least three punctures. The surface m_D in Fig. 4 is an example of a

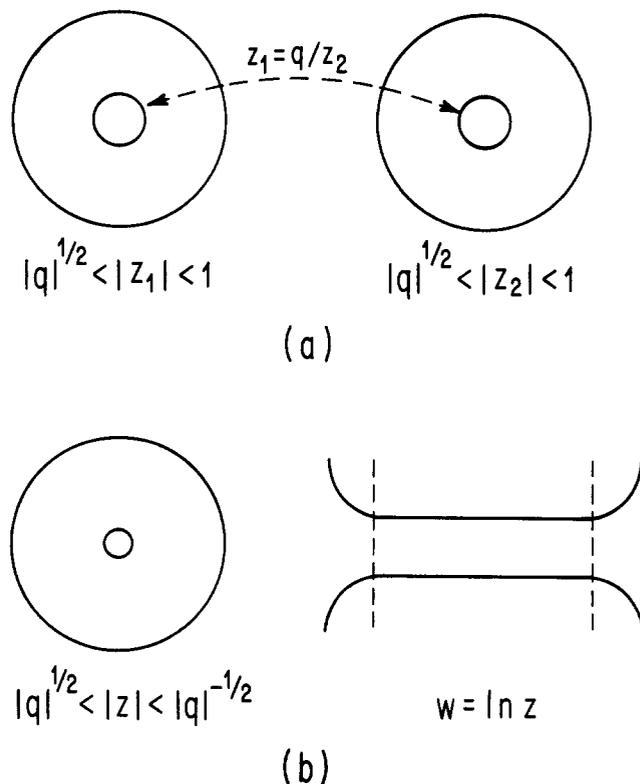


Fig. 5. (a) The neighborhood of an opened node as two annuli with their inner circles identified. (b) The neighborhood of an opened node as a large annulus or long cylinder.

stable surface. The space of stable surfaces with nodes which are connected, compact and of genus g is written D_g . The union $\bar{M}_g = M_g \cup D_g$ is a compact V -manifold or orbifold, called the stable moduli space. D_g is called the compactification divisor. D_g is a divisor in \bar{M}_g because it is carved out locally by the vanishing of one analytic coordinate function, $q = 0$.

The coordinate q parametrizes the opening of a node (Fig. 5). Let $m_D = (m', x, y)$ be a surface with a node. Pick a coordinate neighborhood of the node as pictured in Fig. 3. From each of the two coordinate disks $|z_i| < 1$ remove an inner disk $|z_i| \leq |q|^{1/2}$. Then identify the inner circles $|z_i| = |q|^{1/2}$ by $z_2 = q/z_1$ [Fig. 5(a)]. The conformal structure is held fixed outside the original neighborhood of the node, $|z_i| < 1$. The result is a smooth surface (except for whatever other nodes are present elsewhere on the surface). The surface near m_D are parametrized by $m = (m_D, q) = (m', x, y, q)$. The points $q = 0$ correspond to the surfaces with the node closed, i.e., $m_D = (m', x, y, 0)$. A useful coordinate to use for the neighborhood of each opened node is the annulus $|q|^{1/2} < |z| < |q|^{-1/2}$ [Fig. 5(b)]:

$$z = \begin{cases} q^{1/2}/z_2 & \text{if } |q|^{1/2} < |z| \leq 1 \\ q^{-1/2}z_1 & \text{if } 1 \leq |z| < |q|^{-1/2}. \end{cases} \quad (23)$$

The conformal transformation $w = \ln z$ makes the opened node into a long cylinder or tube [Fig. 5(b)]. The simplest example occurs in genus 1. The point $q = e^{2\pi i\tau} = 0$ corresponds to the limit of the torus in which it becomes the Riemann sphere punctured at 0 and ∞ [cf. eq. (23)]. Adding the point $q = 0$ to M_1 makes the compact space \bar{M}_1 which is a branched cover of the sphere.

As a channel is closed, as $q \rightarrow 0$, the neighbourhood of the opened node becomes the infinite cylinder or the punctured plane. By two dimensional locality, the behavior of the conformal field theory near the midpoint of such a tube is essentially independent of the boundary conditions at the ends of the tube provided by the conformal field theory on the rest of the surface. Deep inside the tube only the ground state component of the boundary condition survives. This means that we only have to understand conformal field theory on the infinite cylinder in order to understand the behavior of the partition function near the compactification divisor.

As q varies, the surface outside the neighborhood of the opened node is held fixed, so the boundary conditions at the ends of the tube, provided by the conformal field theory outside the tube, are independent of q (Fig. 6). If the tube does not separate the surface [Fig. 2(b)] then the boundary conditions at the ends of the tube are correlated. They are described by a density matrix $\rho(\bar{m}_D, m_D)$, in the operator representation of the field theory on the cylinder. If the tube does separate the surface [Fig. 2(a)] the boundary conditions at the two ends of the tube are uncorrelated, and can be represented by a density matrix ρ which is a pure state. The partition function of the surface $m = (m_D, q)$ near m_D is

$$Z(\bar{m}, m) = \text{tr} [q^{L_0} \bar{q}^{\bar{L}_0} \rho(\bar{m}_D, m_D)]. \quad (24)$$

The trace can be expanded into a sum over a complete set of intermediate states $|\phi\rangle$, which are eigenstates of L_0 and \bar{L}_0 with eigenvalues h_ϕ and \bar{h}_ϕ :

$$Z(\bar{m}, m) = \sum_{\phi} q^{h_\phi} \bar{q}^{\bar{h}_\phi} \langle \phi | \rho(\bar{m}_D, m_D) | \phi \rangle. \quad (25)$$

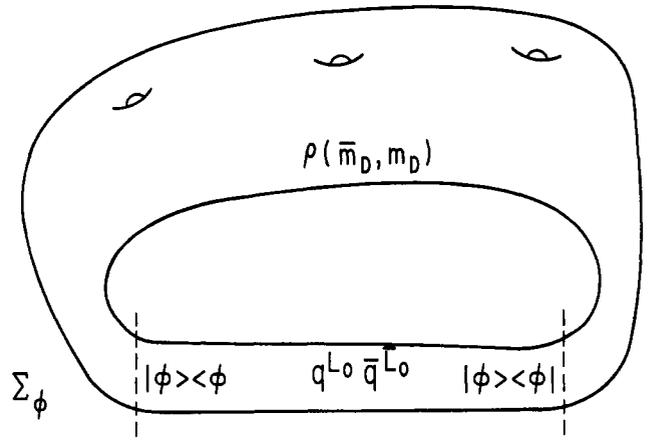


Fig. 6. The sum over states in an almost closed channel.

To interpret the expectation value of q in the state ϕ , turn the picture around and regard $\langle\phi|$ and $|\phi\rangle$ as providing asymptotic boundary conditions at the punctures x and y in the smooth surface $m' = \pi(m_D)$. These asymptotic boundary conditions produce at x and y the scaling field ϕ corresponding to the state $|\phi\rangle$, so $\langle\phi|\rho(\bar{m}_D, m_D)|\phi\rangle$ is the unnormalized two point function of the field ϕ on the surface m' :

$$\langle\phi|\rho(\bar{m}_D, m_D)|\phi\rangle = Z(\bar{m}', m') \langle\phi(x)\phi(y)\rangle_{m'}. \quad (26)$$

At $q = 0$ the only intermediate state which contributes is the SL_2 invariant ground state $|0\rangle$, with $h = \bar{h} = 0$, which corresponds to the identity operator 1. Therefore the partition function on the compactification divisor satisfies the fundamental factorization identity

$$Z(\bar{m}_D, m_D) = Z(\bar{m}', m') \text{ or } Z = Z \circ \pi. \quad (27)$$

In words, the partition function of a surface with node is equal to the partition function of the surface with the node removed and the resulting punctures forgotten. The partition function is continuous under change of connectivity by the processes of Fig. 2.

So far we have only looked at a single node. The surfaces with exactly one node form the generic part of the compactification divisor. The surfaces with multiple nodes form the self-intersections of the compactification divisor (Fig. 7). Since independent, non-overlapping coordinate neighborhoods can be chosen for each node, and since the opening of each node ignores what is happening outside the neighborhood of that node, the opening of the nodes is described by independent coordinates q_α . Near a surface m_D with multiple nodes, the moduli space is parametrized by $m = (m_D, q_1, q_2, \dots)$, where m_D itself is parametrized by the smooth surface $m' = \pi(m_D)$ and by the locations of the punctures in m' . A sum over intermediate states $|\phi_\alpha\rangle$ can be performed in each node α separately, giving

$$Z(\bar{m}, m) = Z(\bar{m}', m') \sum_{\{\phi_\alpha\}} \left(\prod_{\alpha} q^{h_{\phi_\alpha}} \bar{q}^{\bar{h}_{\phi_\alpha}} \right) \left\langle \prod_{\alpha} \phi_\alpha(x_\alpha) \phi_\alpha(y_\alpha) \right\rangle_m. \quad (28)$$

In particular, the ground state factorization condition $Z = Z \circ \pi$ holds on all of the self-intersections of the compactification divisor. The partition function is continuous under all changes of connectivity produced by the formation and removal of nodes.

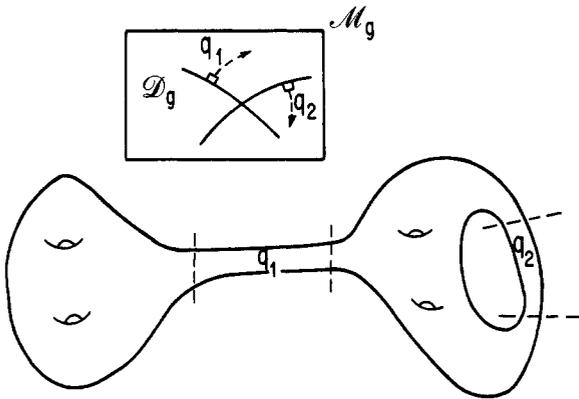


Fig. 7. A surface with multiple nodes as a self-intersection of the compactification divisor.

The independence of the coordinates q_x for the opening of multiple nodes, i.e., the transversality of the self-intersections of the compactification divisor, should be a key ingredient in the eventual proof of finiteness for the perturbative fermionic string in any supersymmetric background. This transversality means that it will suffice to demonstrate finiteness separately for each q_x integral in a neighborhood of 0. Thus the proof will depend only on properties of the appropriate superconformal field theory on the cylinder [12].

The factorization equation (28) can be used to reconstruct the correlation functions of the quantum fields from the behavior of the partition function near the compactification divisor. The essential properties of the correlation functions follow from the properties of the partition function. Real analyticity of the partition function makes the correlation functions also real analytic. Modular invariance of the partition function makes the correlation functions crossing symmetric, since permutations of punctures can be realized by modular transformations. The correlation functions will automatically be SL_2 invariant, since they are expressed in terms of modular invariants of the punctures.

The local conformal invariance of the reconstructed theory is expressed through the dependence of the partition function on the moduli. Differentiating with respect to the moduli inserts smeared stress-energy tensors in correlation functions. To see local conformal symmetry in the cylinder, attach a surface of large genus to the boundaries of the cylinder. Local conformal transformations of the cylinder are approximated by the many independent variations of the moduli for the handles outside the cylinder.

The reconstruction of the correlation functions cannot be done directly, because the correlation functions which appear in eq. (28) are special in that they contain each field ϕ_x twice. However, nodes which separate the surface m_D give correlation functions on $m' = \pi(m_D)$ which are products of correlation functions on the connected components of m' . The paired fields $\phi_x(x_x)$ and $\phi_x(y_x)$ occur on different components. In this way a large class of products of correlation functions on the Riemann sphere can be obtained. We conjecture that this is enough information to reconstruct the correlation function themselves. There could be some difficulty in the reconstruction when multiple fields ϕ have the same dimensions h_ϕ, \bar{h}_ϕ , as when there are internal symmetries in the conformal field theory. But this should be no more serious than the difficulty of reconstructing a quantum field theory with charges from the correlation functions of the neutral fields.

Any given correlation function appears as a factor on the right-hand side of the factorization equation (28) for many different surfaces m_D . It is the ground state factorization condition $Z = Z \circ \pi$ which ensure the consistency of the correlation functions obtained at different points in the compactification divisor. Suppose two surfaces m_D and m'_D factor into $\pi(m_D) = m_0 \cup m_1$ and $\pi(m'_D) = m_0 \cup m_2$, with the same punctures on m_0 . We check that the reconstructed correlation functions on the two copies of m_0 agree by finding a third surface m''_D which factors into $\pi(m''_D) = m_0 \cup m_3$, again with the same punctures on m_0 , such that m_3 can be deformed into either m_1 or m_2 by closing the appropriate channels and removing nodes. This guarantees that the correlation functions on m_0 reconstructed near m_D and near m'_D are the same as those reconstructed near m''_D and therefore are consistent with each other.

The ground state factorization condition $Z = Z \circ \pi$ only makes sense if the partition function is defined on a moduli space closed under the process of forming and removing nodes. The collection of stable moduli spaces \bar{M}_g of connected surfaces will clearly not do, since the removal of nodes can produce a disconnected surface. The smallest moduli space which is closed under these processes is the space of all stable, compact, but not necessarily connected Riemann surfaces:

$$\bar{R} = \prod_{g=0}^{\infty} \left(\bigcup_{n=0}^{\infty} \text{Sym}^n(\bar{M}_g) \right). \tag{29}$$

This is what we call the *universal moduli space*. The partition function is clearly defined on \bar{R} . The surfaces with nodes in \bar{R} form the universal compactification divisor $D = \bar{R} - R$. The map π which eliminates the nodes is a map from D to R . This is pictured along with the inclusions of R and D in \bar{R} in the following diagram

$$\begin{array}{ccc} D & \rightarrow & \bar{R} \\ \pi \downarrow \nearrow & & \\ & & R \end{array} \tag{30}$$

\bar{R} has infinitely many connected components in the naive topology inherited from R . But if we allow continuous paths which move through the compactification divisor via the map π , then \bar{R} is a *connected space*.

By now it should seem natural to make \bar{R} into a connected analytic space by taking as local analytic functions the functions f which are analytic on each naive component and which satisfy $f = f \circ \pi$ on the compactification divisor. With this analytic structure, eq. (30) is a commuting diagram of analytic maps. With this analytic structure \bar{R} is effectively compact as well. It is easy to see that the only global analytic functions are the constants, since a global analytic function f must be constant on all the naive components of \bar{R} , since they are compact, and the constants must all be equal by $f = f \circ \pi$. The fundamental physical conditions of real analyticity, modular invariance and ground state factorization are subsumed into the condition that the partition function be a globally defined real analytic function on universal moduli space.

We have another motivation for introducing this universal moduli space. If there is to be any use in describing quantum string in terms of analytic geometry on moduli space nonperturbative string effects must be included. As long as the

connected components of the surfaces all have finite genus we are only doing the perturbation theory. The moduli of infinite genus surfaces should in some sense be the limit $g \rightarrow \infty$ of the moduli spaces of finite genus surfaces. But the moduli spaces \bar{M}_g are not suitable for taking this limit, since for $g < g'$ there is no inclusion $\bar{M}_g \rightarrow \bar{M}_{g'}$. \bar{R} on the other hand can be written as a limit of sequence of finite dimensional spaces, essentially corresponding to the string perturbation expansion. We expect that nonperturbative string theory can be done as analytic geometry on some completion of \bar{R} in this limit. Call this hypothetical completion \bar{R}_∞ . Intuitively, \bar{R}_∞ should be very regular. The completion of the compactification divisor, D_∞ , should be essentially dense in \bar{R}_∞ , since the partition function would be zero unless almost all channels were nearly closed. The commuting diagram (30) should have interesting consequences for the geometry of \bar{R}_∞ .

6. The hermitian metric and the holomorphic section

We now construct, given a conformal field theory, a holomorphic vector bundle W universal moduli space, a hermitian metric h_{ab} in W and a holomorphic section $\psi^a(m)$ of W , such that the partition function is $Z = h(\bar{\psi}, \psi)$. Write the set of primary conformal fields $\{\phi_k\}$, with corresponding highest weights $\{\bar{h}_k, h_k\}$.

We do the construction first on M_g . Pick a stable surface m_D with $3g - 3$ nodes, such that when all the nodes are removed the resulting components consist of $2g - 2$ spheres, each containing exactly three punctures. Such a surface is pictured in Fig. 8, where the spheres are represented by vertices and the nodes by lines. There are a finite number of these *maximally stable surfaces*. The partition function near m_D can be expressed effectively in terms of the operator product coefficients and the highest weights. Let $\{q_\alpha\}$ be the coordinates describing the opening of the nodes labelled by $\alpha = 1, \dots, 3g - 3$. Then $m = (q_1, \dots, q_{3g-3})$ is a complete set of local analytic coordinates on \bar{M}_g near m_D . Label the spheres or vertices by the triplets (α, β, γ) of channels which they connect to.

Calculate the partition function by summing over a complete set of irreducible representations of the Virasoro algebra in each channel. For each term in the sum let k_α label the

irreducible representation occurring in the channel α , and write $k = (k_1, \dots, k_{3g-3})$, so the sum is over all possible values of the multi-index k . Let Z_k be the contribution to the partition function of the representations labelled by k . Write $\bar{h}(k) = (\bar{h}_{k_1}, \dots, \bar{h}_{k_{3g-3}})$ and $h(k) = (h_{k_1}, \dots, h_{k_{3g-3}})$ for the corresponding highest weights. Now calculate Z_k by summing over the states in each of the irreducible representations k_α . These states are generated by the L_n and the \bar{L}_n acting on the highest weight states. Now apply the general factorization equation (28) and use (in principle) the Virasoro commutation relations. The crucial point is that the L_n and the \bar{L}_n commute, so all combinations of the descendents in L_n and the descendents in \bar{L}_n occur. The result is that each contribution of irreducible representations k is the product of a holomorphic function and an anti-holomorphic function:

$$Z_k(\bar{m}, m) = \left[\prod_{\substack{\text{vertices} \\ (\alpha, \beta, \gamma)}} C(k_\alpha, k_\beta, k_\gamma) \right] \bar{\psi}^{\bar{h}(k)}(\bar{m}) \psi^{h(k)}(m) \quad (31)$$

where the $C(k_\alpha, k_\beta, k_\gamma)$ are the three point functions on the sphere or, equivalently, the operator product coefficients of the primary fields:

$$C(k_\alpha, k_\beta, k_\gamma) = \langle \phi_{k_\alpha}(\infty) \phi_{k_\beta}(1) \phi_{k_\gamma}(0) \rangle = \langle \phi_{k_\alpha} | \phi_{k_\beta}(1) | \phi_{k_\gamma} \rangle. \quad (32)$$

The locally analytic functions $\psi^{h(k)}(m)$ are *generalized characters*. In principle, they can be calculated in power series in the q_α using only the commutation relations of the Virasoro algebra L_n and the values of the $3g - 3$ highest weights $h(k)$.

Now, let $b = (b_1, \dots, b_{3g-3})$ be a multi-index, where b_i ranges over all the *values* of the highest weights h which occur in the theory, and similarly let $\bar{a} = (\bar{a}_1, \dots, \bar{a}_{3g-3})$ with \bar{a} ranging over all the values of \bar{h} occurring in the theory. The sum over all contributions (31) gives the partition function in the form

$$Z(\bar{m}, m) = \bar{\psi}^{\bar{a}}(\bar{m}) h_{ab} \psi^b(m). \quad (33)$$

where

$$h_{ab} = \sum_{\substack{k \\ \bar{h}(k)=\bar{a} \\ h(k)=b}} \prod_{\text{vertices} (\alpha, \beta, \gamma)} C(K_\alpha, k_\beta, k_\gamma). \quad (34)$$

This construction works directly only in some neighbourhood of each of the maximally stable surfaces. We presume that the locally analytic functions $\psi^a(m)$ can be analytically continued to give multi-valued functions on the whole of moduli space, and that they are closed under analytic continuation. Equivalently, we assume that the power series for the $\psi^a(m)$ converge in large enough neighborhoods of the maximally stable surfaces m_D to collectively cover M_g , and that the functions corresponding to different m_D are linearly related on the overlaps of these neighborhoods. These technical assumptions are not likely to be easily proved.

Given these assumptions, the multi-valued functions $\psi^a(m)$ define a representation of the fundamental group of M_g , the modular group Γ_g . Equivalently, they define a flat vector bundle W over M_g , and are the components of a holomorphic section ψ of W . The single-valuedness of the partition function then requires that h_{ab} be an invariant hermitian inner product in the representation of Γ_g , or equivalently a flat hermitian metric in W (dividing out by the null space of h_{ab} if there is one).

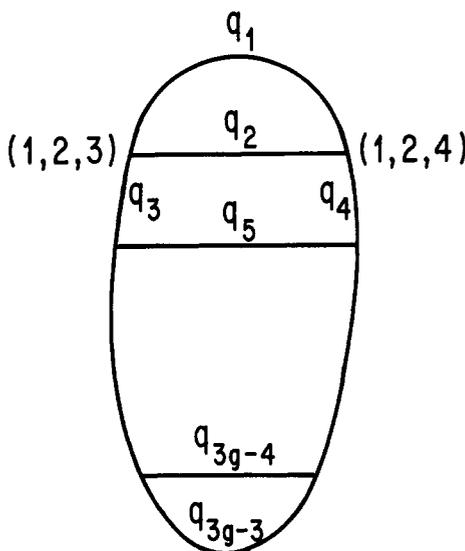


Fig. 8. Diagram of a maximally stable surface of genus g constructed from $2g - 2$ spheres (the vertices) connected at $3g - 3$ nodes (the lines).

The hermitian metric and holomorphic section have now been constructed (modulo technical assumptions) over the moduli spaces M_g and thus, by taking tensor products, over R . But to make use of the effective connectedness and compactness of the universal moduli space \bar{R} we need to fit together the vector bundles W over the various connected components of R to give a single vector bundle over \bar{R} . To see how $\psi^a(m)$ and h_{ab} behave when a channel is closed, simply include that channel as one of the nodes in a maximally stable surface m_p and study the behavior near m_D . It will be evident that ψ and h individually satisfy factorization identities. For example, if m_D is a surface with a node which separates it into two components m_1 and m_2 , then there are constants F_{bc}^a such that

$$\begin{aligned}\psi^a(m_D) &= F_{bc}^a \psi^b(m_1) \psi^c(m_2) \\ h_{ab}(m_D) \bar{F}_{\bar{c}_1 \bar{c}_2}^{\bar{a}} F_{d_1 d_2}^b &= h_{\bar{c}_1 d_1}(m_1) h_{\bar{c}_2 d_2}(m_2).\end{aligned}\quad (35)$$

Such factorization constants for a vector bundle W over R are exactly what is needed to define a vector bundle over \bar{R} , with ψ a global holomorphic section and h a flat metric. The analogue of $f = f \circ \pi = \pi^* f$ for functions is a rule $\psi = \pi^* \psi$ for vector valued functions, which expresses the value of ψ on D in terms of its value on $R = \pi(D)$. Equation (35) does exactly this, since the tensor product construction gives $\psi^b(m_1) \psi^c(m_2)$ as the value of ψ on $\pi(m_D) = m_1 \cup m_2$. Equation (35) obviously implies that $Z = h(\bar{\psi}, \psi)$ is a real analytic function on \bar{R} .

To apply this formulation of $c = 0$ conformal field theory to string theory we need to take account of the zero modes of the ghost system. The moduli space is the gauge slice for the conformal gauge. Therefore there is one zero mode b_i of the ghost field $b(z)$ for every tangent direction in moduli space. The ghost fermionic functional integral is zero unless these zero modes are absorbed by a factor $\prod_i b_i dm^i \bar{b}_i d\bar{m}^i$ introduced as an integrand. This is why the string partition function is a density on moduli space. The b_i contribute to the holomorphic section $\psi^a(m)$ for the string; the \bar{b}_i contribute to $\bar{\psi}$. Therefore the string holomorphic section ψ is proportional to $\prod_i dm^i$, i.e., it transforms as a holomorphic half-density on moduli space. In the standard notation for differential forms on analytic spaces, the holomorphic half-densities are sections of the line bundle $K = \wedge^{\max} T^* R^{1,0}$. Thus the string holomorphic section ψ is a section of $K \otimes W$. This is exactly what is needed to make $Z = h(\bar{\psi}, \psi)$ a density, since the metric h absorbs the factors \bar{W} and W to leave Z as a section of $|K|^2$.

The ghost zero modes also are responsible for modifying the factorization condition for ψ . Suppose m_D is a surface with a node. Write $m' = \pi(m_D)$ for the smooth surface with the node removed, write (x, y) for the punctures on m' left by the node, and write q for the coordinate parametrizing the opening of the node. As we have seen, (m', x, y, q) are coordinates for the surfaces near m_D . In conformal theory, the factorization condition on ψ would be of the form, for example, of eq. (35). More generally, $\psi(m', x, y, 0) = \pi^* \psi(m')$, expressing ψ at m_D in terms of ψ at $m' = \pi(m_D)$. In string theory the factorization condition expresses a residue of ψ at $q = 0$ in terms of ψ at m' :

$$\psi(m', x, y, q) \sim q^{-2} dq dx dy \psi(m'). \quad (36)$$

The abstract string S -matrix is given by the integral

$$S[h] = \int_{\bar{R}} h(\bar{\psi}, \psi), \quad (37)$$

considered as a functional of the flat hermitian metrics h , which are the backgrounds solving the string quantum equation of motion. For bosonic string this can only be a formal expression since, by eq. (36), the integral is always divergent. We assume, in writing $S[h]$ as a function of the metric alone, that for any sensible background metric h , the section ψ will be uniquely determined, at least to the extent that it affects S .

7. A topological interpretation

It is possible to write an even more gauge invariant version of this abstract string theory. For any flat hermitian connection in a vector bundle W we can write an exterior derivative on differential forms in \bar{R} with coefficients in W :

$$Q = dm^i D_i + d\bar{m}^{\bar{i}} \bar{D}_{\bar{i}} \quad (38)$$

which satisfies

$$Q^2 = dm^i d\bar{m}^{\bar{i}} F_{\bar{i}j} = 0. \quad (39)$$

The hermitian metric h in W gives a hermitian product $h(\bar{\omega}, \omega)$ on differential forms ω with coefficients in W , using the antisymmetric product of differential forms. The value of $h(\bar{\omega}, \omega)$ is an ordinary differential form on \bar{R} . The inner product

$$\langle \bar{\omega} | \omega \rangle_h = \int_{\bar{R}} h(\bar{\omega}, \omega) \quad (40)$$

makes at least formal sense. With respect to this inner product, again only formally in the bosonic theory,

$$Q^\dagger = Q. \quad (41)$$

Equations (39) and (41) give an even more abstract version of the string equation of motion.

It is easily checked that the inner product (40) is formally invariant under $\omega \rightarrow \omega + Q\eta$, and is null on forms $\omega = Q\eta$, so it is really an inner product on cohomology classes with coefficients in W . It is also easy to see that any holomorphic half-density ψ with coefficients in W satisfies

$$Q\psi = 0 \quad (42)$$

so the string holomorphic section represents a cohomology class $[\psi]$ with coefficients in W . The S -matrix depends only on the class $[\psi]$. The uniqueness condition on ψ is the requirement that the cohomology be one dimensional.

The final point is based on conversations with D. Kazhdan and E. Martinec (and of course S. H. Shenker). The inner product on cohomology given by eq. (40) determines an intersection matrix on the dual homology theory. So the string S -matrix can be interpreted as an intersection number.

8. Conclusion

I will end by mentioning what seem to be some interesting directions in which to pursue this abstract formulation of string theory. Most obvious is the extension to fermionic string [12]. As mentioned previously, this is a straightforward transcription of the bosonic theory to the super moduli space of super Riemann surfaces. As an immediate application, we expect to be able to use $N = 2$ superconformal invariance and the structure of the super moduli space of

$N = 2$ super Riemann surfaces [17] to prove abstract finiteness and renormalization theorems to all orders in perturbation theory. Such theorems would apply in particular to perturbative fermionic string theory in flat spacetime and also in any supersymmetric compactification.

Another question which can be investigated within perturbation theory is the derivation of eq. (37) for the S -matrix from the quantum equation of motion $F = 0$ or $Q^2 = 0$. The relation between the two is at the moment only indirect. A basic issue which we have not addressed is how, and when, to interpret an abstract solution of string theory as a description of spacetime geometry. This requires understanding the characteristic *intrinsic* properties of the conformal field theories which are nonlinear models for weakly curved spacetimes. The abstract formulation of Wick rotation can also be addressed in the perturbative theory. And of course there are several technical assumptions which need to be proved.

The deepest problems have to do with constructing and interpreting a non-perturbative theory. The main technical task is to find a good completion of universal moduli space. This might be done by studying Riemann surfaces of infinite genus, but it probably not necessary to do so. The string theory is defined entirely in terms of intrinsic analytic geometry on universal moduli space. The theory has forgotten that this analytic space parametrizes Riemann surfaces. So it should be possible to complete universal moduli space directly, without reference to surfaces. This might suggest that interpreting the theory as a theory of strings in spacetime would not be sensible except in the perturbative approximation.

The first step is to understand the criteria which should govern such a completion. Then, given a completion, we would ask if the nonperturbative string equation of motion has a unique solution. It might be that there is a unique non-perturbative solution over the completion of universal moduli space in a sense similar to that in which there are far fewer von Neumann algebras than C^* -algebras. There is no real reason to suppose that a nonperturbative theory of gravity should allow for Wick rotation and the concomitant manifold of solutions, or even a quantum mechanical interpretation. A related question is whether there is a sense in which a perturbative solution can be completed to give a nonperturbative solution, or whether perturbative solutions can be measured as approximations to the nonperturbative solution(s). This is connected to the need for an abstract theory of measurement, since the validity of a perturbative approximation should depend on the regime it is applied to, i.e., to the observables of interest. It would seem necessary to address these foundational issues in string theory in order to understand how to ask the crucial question of whether there is a regime of non-perturbative string theory susceptible to a semi-classical interpretation in terms of spacetime geometry and, in particular, which spacetime geometry of the many now available [18].

Underlying these questions is the most basic one: whether

the metric h and the holomorphic section ψ should be interpreted as structures more basic than the Hilbert space and hamiltonian of quantum mechanics, whether the equation of flatness on universal moduli space is a fundamental principle, whether this geometric formalism of abstract string theory provides a new formulation of physics.

References

1. Friedan, D. and Shenker, S. H., The Analytic Geometry of Conformal Field Theory, EFI preprint 86-18A (1986), Nucl. Phys. B (to appear).
2. Friedan, D. and Shenker, S. H., Phys. Lett. **175B**, 287 (1988).
3. Friedan, D., Qiu, Z., and Shenker, S. H., in Proceedings of the Santa Fe Meeting of the APS Division of Particles and Fields, October 31–November 3, 1984 (Edited by T. Goldman and M. Nieto), World Scientific (1985).
4. Nambu, Y., Directions of Particle Physics, EFI preprint 85-71 (1985), to appear in the proceedings of Meson 50, Kyoto International Symposium (1985), and Science, Two Cultures and Postwar Japan, to appear in the proceedings of the Rikkyo Symposium on Postwar Japan.
5. For references on string theory and in particular on conformal invariance in string theory see, for example, Dual Theory (Edited by M. Jacob), North Holland (1974); Superstrings, the First Fifteen Years, (Edited by J. Schwarz), World Scientific (1985); Proceedings of the Workshop on Unified String Theories, Institute for Theoretical Physics, Santa Barbara, July 29–August 16, 1985 (Edited by M. Green and D. Gross), World Scientific (1986); Friedan, D., Martinec, E., and Shenker, S. H., Nucl. Phys. **B271**, 93 (1986).
6. Friedan, D., Nucl. Phys. **B271** 540 (1986); Phys. Lett. **162B**, 102 (1985).
7. Polyakov, A. M., Pisma ZhETP **12**, 538 (1970) [JETP Lett. **12**, 381 (1970)]; ZhETP **66**, 23 (1974) [JETP **39**, 10 (1974)].
8. Belavin, A. A., Polyakov, A. M., and Zamolodchikov, A. B., J. Stat. Phys. **34**, 763 (1984); Nucl. Phys. **B241**, 333 (1984).
9. Friedan, D., Qiu, Z., and Shenker, S. H., in Vertex Operators in Mathematics and Physics (Edited by J. Lepowsky *et al.*), Springer-Verlag (1984); Phys. Rev. Lett. **52**, 1575 (1984).
10. Cardy, J., Nucl. Phys. **B270**[FS16], 186 (1986).
11. For background on the moduli spaces of connected Riemann surfaces see, for example, Wolpert, S., Ann. Math. **118**, 491 (1983); Amer. J. Math. **107**, 1485 (1985). For background on analytic geometry see, for example, Griffiths, P. and Harris, J., Principles of Algebraic Geometry, Wiley (1978).
12. Friedan, D., Martinec, E., and Shenker, S. H., Work in progress.
13. Belavin, A. A. and Knizhnik, V. G., Phys. Lett. **168B**, 201 (1988); Alvarez-Gaume, L., Ginsparg, P., Moore, G., and Vafa, C., Phys. Lett. **171B**, 155 (1986); Manin, Yu., Phys. Lett. **172B**, 184 (1986); Catenacci, R., Conalpa, M., Martellini, M. and Reina, C., Phys. Lett. **172B**, 328 (1986); Bost, J. B. and Jolicoeur, T., Phys. Lett. **174B**, 273 (1986); Seiberg, N. and Witten, E., Princeton preprint (1986); Dixon, L. and Harvey, J., Nucl. Phys. **B274**, 93 (1986); Alvarez-Gaume, L., Moore, G., and Vafa, C., Comm. Math. Phys. **106**, 40 (1986); Cohen, A., Moore, G., Nelson, P., and Polchinski, J., Harvard preprint (1986); Gomez, C., Phys. Lett. **175B**, 32 (1986).
14. Martinec, E., Phys. Lett. **171B**, 189 (1986), Princeton preprint (1986).
15. Blote, H., Cardy, J., and Nightingale, M., Phys. Rev. Lett. **56**, 742 (1986); Affleck, I., Phys. Rev. Lett. **56**, 746 (1986). This ground state energy was first calculated in free field theory in Brink, L. and Nielsen, H. B., Phys. Lett. **45B**, 332 (1973).
16. Rocha-Caridi, A., in Vertex Operators in Mathematics and Physics, (Edited by J. Lepowsky *et al.*), Springer-Verlag (1984).
17. Boucher, W., Friedan, D., and Kent, A., Phys. Lett. **172B**, 316 (1986); Friedan, D., Kent, A., Shenker, S. H., and Witten, E., in preparation; Cohn, J. D., $N = 2$ Super Riemann Surfaces, EFI preprint 86-32 (1986).
18. Cf. Banks, T., Nucl. Phys. **B249**, 332 (1985).