

Perturbative superstring theory revisited

EDWARD WITTEN

Abstract: Perturbative superstring theory is revisited, with the goal of giving a simpler and more direct demonstration that multi-loop amplitudes are gauge-invariant (apart from known anomalies), satisfy space-time supersymmetry when expected, and have the expected infrared behavior. The main technical tool is to make the whole analysis, including especially those arguments that involve integration by parts, on supermoduli space, rather than after descending to ordinary moduli space.

1	Introduction	219
2	A measure on moduli space	222
2.1	The vacuum amplitude	222
2.2	Reducing to moduli space	224
2.3	Integration over moduli space	228
2.4	Vertex operator insertions	231
2.4.1	Conformal vertex operators	231
2.4.2	Integration on moduli space	233
2.4.3	Gauge invariance	234
2.4.4	More general vertex operators	235
2.4.5	Open and/or unoriented strings	236
2.5	Integrated vertex operators	237
2.5.1	Two types of vertex operator	237
2.5.2	The derivation	238
2.5.3	More on gauge invariance	242

arXiv: [1209.5461](https://arxiv.org/abs/1209.5461)

Received September 6, 2018.

2.6	Some variations on the theme	244
3	A measure on supermoduli space	245
3.1	A short account of super Riemann surfaces	246
3.1.1	Worldsheets	246
3.1.2	Deformations	248
3.1.3	Action and ghosts	250
3.2	The extended action and the integration measure	252
3.2.1	A lightning review of integration on supermanifolds	253
3.2.2	Bosonic integration as an algebraic operation	255
3.2.3	The zero-mode coupling	258
3.3	Where to integrate	260
3.3.1	Reduction to supermoduli space	260
3.3.2	A hard-to-avoid detail	261
3.4	Integration over a slice	264
3.4.1	More on changes of coordinates	267
3.5	A basis of odd moduli	269
3.6	Integrating out odd variables	273
3.6.1	General procedure	273
3.6.2	The picture-changing operator	275
3.6.3	Picture number	278
4	The Neveu-Schwarz sector	280
4.1	Neveu-Schwarz vertex operators	281
4.1.1	Other picture numbers?	284
4.2	Scattering amplitudes for NS states	285
4.3	Integrated NS vertex operators	288
4.3.1	More on gauge invariance	290

4.4	Tree-level scattering	291
4.5	Integration over odd moduli	293
5	The Ramond sector	296
5.1	Ramond insertion points and superconformal algebra	296
5.1.1	The alternative	301
5.1.2	A pair of Ramond punctures	301
5.2	Matter vertex operators in the Ramond sector	303
5.3	Ghost vertex operators in the Ramond sector	305
5.3.1	The usual description of the $\beta\gamma$ system	308
5.3.2	The definition of the ghosts	310
5.4	Ramond amplitudes	312
5.4.1	The simplest examples	314
5.5	Duality	316
5.5.1	Massless fermions in ten dimensions	318
6	The Propagator	320
6.1	Open-string propagators	323
6.1.1	Bosonic open strings	323
6.1.2	The NS propagator	326
6.1.3	Picture number and the propagator	328
6.1.4	The Ramond propagator	331
6.2	Closed-string propagators	334
6.2.1	Closed bosonic strings	334
6.2.2	Closed superstrings	337
6.3	The Feynman $i\epsilon$	339
6.3.1	The propagator in lorentz signature	339
6.3.2	An interesting analogy	342
6.3.3	Closed strings	343

6.4	The on-shell and infrared behavior of string theory	344
6.4.1	Overview	344
6.4.2	The separating case	346
6.4.3	The nonseparating case	347
6.4.4	The leading singularity	349
6.4.5	Factorization of superstring amplitudes	351
6.4.6	Picture numbers and picture-changing operators	354
6.5	The states that contribute to the pole	356
6.5.1	Bosonic open strings	357
6.5.2	Open superstrings	360
6.5.3	Analog for closed strings	362
6.5.4	Another look at the pole of the propagator	364
6.6	The Deligne-Mumford compactification and the integration cycle	367
6.6.1	The Heterotic string	367
6.6.2	An example	370
6.6.3	Other superstring theories	371
7	BRST anomalies, massless tadpoles, and all that	374
7.1	BRST anomalies	374
7.2	Review of field theory	377
7.2.1	The importance of spacetime supersymmetry	379
7.2.2	A subtlety at zero momentum	381
7.3	Back to string theory	382
7.4	Gauge symmetry breaking	386
7.4.1	Examples	389
7.4.2	Some miscellaneous remarks	392

	Perturbative superstring theory revisited	217
	7.4.3 Restricting to the massless S -matrix	393
7.5	Massless tadpoles in string theory	394
7.6	What happens when the tadpole vanishes	395
	7.6.1 Conditionally convergent integrals	395
	7.6.2 Another infrared regulator	402
	7.6.3 Wavefunction renormalization	405
7.7	More detail on the anomaly	406
8	Spacetime supersymmetry and its consequences	411
8.1	Massless particles and conserved charges	412
	8.1.1 Gauge theory	412
	8.1.2 Gravity and supergravity	416
8.2	Gauge symmetries in string theory and conserved charges	418
	8.2.1 Manifest symmetries of the worldsheet theory	418
	8.2.2 Spacetime supersymmetry	421
	8.2.3 The supersymmetric Ward identity	424
	8.2.4 Comparison to the standard description	428
	8.2.5 More on momentum conservation	430
	8.2.6 The supersymmetry algebra	431
8.3	Vanishing of massless tadpoles	435
9	Open and/or unoriented strings	438
9.1	Overview	438
	9.1.1 Orientifolds and D-branes	439
	9.1.2 What we will learn	440
9.2	Fundamentals	441
	9.2.1 Geometry	441
	9.2.2 BRST anomalies	444

9.2.3	Spacetime supersymmetry for open and/or unoriented superstrings	445
9.3	NS-NS and R-R tadpoles	448
9.3.1	Preview	448
9.3.2	Ward identity with R-R vertex operators	449
9.3.3	A factor of momentum	451
9.3.4	When the fermionic gluing parameter disappears	454
9.3.5	The tadpoles	457
9.3.6	What happens when R-R tadpoles cancel?	460
9.4	What happens when R-R tadpoles do not cancel?	463
10	More on the $\beta\gamma$ system	469
10.1	Preliminaries	469
10.1.1	The importance of ghost number	471
10.2	Correlation functions for spin 1/2	472
10.2.1	Elementary fields	472
10.2.2	Delta function operators	473
10.2.3	Multiple delta function insertions	476
10.2.4	Mixed correlation functions	478
10.2.5	Extended delta functions	482
10.3	The $\beta\gamma$ system with zero-modes	483
10.3.1	The minimal delta function insertion	484
10.3.2	More general correlation functions	485
10.3.3	Spin fields?	489
10.3.4	Reparametrization invariance	490
10.3.5	Another interpretation of the delta function insertions	492
A	Pullback of differential forms	493

B	Bosonic string gauge parameters	494
B.1	Examples	494
B.2	General proof	495
C	Superstring gauge parameters	499
C.1	Examples	499
C.2	General proof	501
D	An example of fermion integration	504
D.1	What we have learned	508
D.2	The moduli space as an orbifold	509
	Acknowledgments	510
	References	510

1. Introduction

Powerful covariant methods to compute superstring scattering amplitudes in the RNS formulation were introduced in the 1980s [1] and have been widely used ever since.¹ But some aspects of the formalism have remained slightly opaque. In the present paper, we will revisit this subject, aiming to give simpler and more direct demonstrations that multi-loop superstring amplitudes are gauge-invariant (apart from known anomalies), satisfy space-time supersymmetry when expected, and have the infrared behavior that one would expect of a field theory with the same massless particles and low energy interactions.

These results are more transparent if superstring perturbation theory is formulated in terms of super Riemann surfaces and supermoduli space, rather than aiming to reduce everything to ordinary Riemann surfaces and moduli space. This is not a new idea and most of the pieces of the puzzle have been described long ago. The literature is too vast to be properly summarized here,² but a few relevant points are as follows. The measures on supermoduli space that should be integrated to compute perturbative superstring scattering amplitudes were defined from several points of view [4–6] in the 1980s.

¹Parts of this construction were developed independently in [2].

²The history and development of string theory up to 1984 are recounted in [3].

The link between the picture-changing formalism of [1] and integration over fermionic moduli was made in [7]. The article [8] gives a thorough review of much of what was known in the late 1980s. Some of the ideas that will be important in the present paper were introduced in the 1990s in work that has unfortunately remained little-known [9–11]. Finally, we mention a different kind of milestone; the first completely consistent one-loop computations were performed in [12, 13], and two-loop calculations were first performed in [14].

We start by describing the appropriate measures on moduli space and supermoduli space, and the framework for understanding gauge-invariance of loop amplitudes. We review the bosonic string in section 2 in a way that straightforwardly generalizes to superstrings in section 3. In sections 4 and 5, we extend the superstring analysis to include external vertex operators, among other things incorporating the relevant aspects of the covariant quantization of superstrings [1] and discussing the role of pictures. In section 6, we use the general formalism to compute the string propagator, or in other words the integration measure for a string that propagates almost on-shell for a long proper time. This is a good illustration, and also an important example, since it is the key to understanding the infrared region and explaining why superstring theory has precisely the same singularities as a field theory with the same low energy content. In section 7, we analyze BRST anomalies and introduce the massless tadpoles that present the most subtle challenge for superstring perturbation theory. We study loop-induced symmetry breaking in detail, even though it is fairly rare, because it is a potential failure mode of superstring perturbation theory and it is important to have criteria under which this failure mode does not occur. In section 8, we analyze the spacetime supersymmetry of loop amplitudes and the vanishing of massless tadpoles. The corresponding analysis for open and/or unoriented superstrings involves some new ingredients related to spacetime anomalies and Ramond-Ramond tadpoles and is treated in section 9. Finally, in section 10, we provide more details about the path integral of the bosonic ghost system in the presence of the delta function insertions that are ubiquitous in superstring perturbation theory.

We will generally attempt to minimize the demands on the reader's familiarity with supermanifolds and integration on them and with super Riemann surfaces and their moduli. More detail can be found, for example, in [15, 16] and references cited there.

It is tedious to always repeat everything for all of the open and closed, oriented and unoriented, bosonic and supersymmetric string theories. Our default examples are usually closed oriented bosonic strings and heterotic strings. When this is simplest, we explain a point in the context of open

strings. We only run through the full roster of string theories when there are novelties involved, and otherwise leave generalizations to the reader.

We make several restrictions on the scope of this paper. First and very important, we assume that the reader has already learned the basics about superstring perturbation theory elsewhere, and we only review the basics here to the extent that it seems particularly useful. Second, as explained in section 2.4.1, we consider only the simplest class of vertex operators that suffice for computing the S -matrix. These are vertex operators that are conformal or superconformal primaries with the simplest possible dependence on the ghost fields. Techniques for computing with more general vertex operators (arbitrary BRST-invariant vertex operators that are annihilated by $b_0 - \tilde{b}_0$) are well-established [17, 18], but the necessary details would make the present paper more complicated, possibly without interacting in an interesting way with our main points.

Accordingly, we omit certain questions, such as mass renormalization and an analysis of models in which a perturbative shift in the vacuum is necessary to maintain spacetime supersymmetry, that are difficult to study in a conformally-invariant formalism. Actually, an off-shell formalism appropriate for these questions has been developed in a relatively recent series of papers (these papers were not available when the original version of the present paper was submitted to the arXiv in 2012). For a selection of these papers, see [19–24]. These papers have been written in the language of picture-changing operators rather than in terms of super Riemann surfaces. It would certainly be possible to restate the results of those papers in the super Riemann surface language employed in the present paper, basically by working on a super Riemann surface with a choice of local superconformal coordinates at each puncture. However, this has not yet been done.

Likewise, it is difficult to incorporate expectation values of Ramond-Ramond fields in the superconformal framework, so we do not consider backgrounds with such expectation values. Similarly, we do not make contact with the pure spinor formalism [25].

We will not discuss ultraviolet issues in this paper, since there are none. Modular invariance, which was discovered over forty years ago [26] following the recognition of the special role of 26 dimensions [27], removes the ultraviolet region from superstring perturbation theory, though it took some time for this to be understood.

An informal account of some of the main ideas in this paper can be found in [28].

2. A measure on moduli space

2.1. The vacuum amplitude

We begin by recalling how the path integral of bosonic string theory generates a measure on the moduli space of Riemann surfaces. For standard explanations, see [29–31] or section 5.3 of [32] (whose conventions we generally follow). The explanation that follows is chosen to extend straightforwardly to superstring theory.

As usual, the worldsheet of a bosonic string is a Riemann surface Σ , endowed with a metric tensor g and parametrized by local coordinates σ^i , $i = 1, 2$. We always assume Weyl invariance, so we are really interested in the conformal structure, or equivalently the complex structure, defined by g . The worldsheet theory has matter fields X and ghost fields c^i , b_{ij} ; geometrically, c is a vector field on Σ and b is a symmetric traceless tensor. The gauge-fixed action I is the sum of a matter action I_X and a ghost action I_{gh} ,

$$(2.1) \quad \begin{aligned} I &= I_X + I_{\text{gh}} \\ I_{\text{gh}} &= \frac{1}{2\pi} \int d^2\sigma \sqrt{g} b_{ij} D^i c^j. \end{aligned}$$

One defines an anomalous ghost number charge N_{gh} such that c and b have $N_{\text{gh}} = 1$ and -1 respectively. On a surface³ Σ of genus g , the ghost number anomaly violates N_{gh} by $-(6g - 6)$, meaning that a product of operators of definite ghost number can have an expectation value only if their total ghost number is $-(6g - 6)$.

The gauge-fixed theory has a BRST symmetry with generator Q_B . Apart from the fact that Q_B has ghost number 1 and obeys $Q_B^2 = 0$, its most important property for our purposes is that

$$(2.2) \quad \{Q_B, b_{ij}\} = T_{ij},$$

where T_{ij} is the stress tensor, defined as the response of the action to a change in the metric:

$$(2.3) \quad \delta I = \frac{1}{4\pi} \int d^2\sigma \sqrt{g} \delta g_{ij} T^{ij}.$$

³ The function $F(g|\delta g)$ that we will define vanishes if Σ has any conformal Killing vector fields, since in this case the ghost field c has zero-modes and the integral over those zero-modes will vanish. Hence, until we introduce punctures in section 2.4, the following discussion is non-trivial only if $g > 1$.

In asserting that the action I is BRST-invariant, one views the metric g as a fixed, c -number quantity. However, it will be convenient to introduce a new fermionic variable δg , of $N_{\text{gh}} = 1$, and to extend the BRST symmetry by

$$(2.4) \quad [Q_B, g_{ij}] = \delta g_{ij}, \quad \{Q_B, \delta g_{ij}\} = 0.$$

We take δg_{ij} to be a symmetric traceless tensor (we define δg to be traceless as we are only interested in g up to Weyl transformations).

The formula (2.3) shows that if we allow the metric to vary with an unspecified BRST variation δg_{ij} , then the action I is not invariant. But the fundamental formula (2.2) shows that we can restore BRST symmetry with a simple addition to the action:

$$(2.5) \quad I \rightarrow \widehat{I} = I + \frac{1}{4\pi} \int d^2\sigma \sqrt{g} \delta g_{ij} b^{ij}.$$

Now we define a function⁴ of g and δg by integration over the fields X, b, c :

$$(2.6) \quad \begin{aligned} F(g|\delta g) &= \int \mathcal{D}(X, b, c) \exp\left(-\widehat{I}(X, b, c; g, \delta g)\right) \\ &= \int \mathcal{D}(X, b, c) \exp(-I) \exp\left(-\frac{1}{4\pi} \int_{\Sigma} d^2\sigma \sqrt{g} \delta g_{ij} b^{ij}\right). \end{aligned}$$

Since the action and the integration measure are BRST-invariant, integrating out the fields X, b, c is a BRST-invariant operation, so $F(g|\delta g)$ is BRST-invariant:

$$(2.7) \quad [Q_B, F(g|\delta g)] = 0.$$

The integrand in (2.6) is an inhomogeneous function of δg . However, the function $F(g|\delta g)$ is actually homogeneous in δg with a definite weight. To see this, we expand in powers of δg :

$$(2.8) \quad \exp\left(-\frac{1}{4\pi} \int_{\Sigma} d^2\sigma \sqrt{g} \delta g_{ij} b^{ij}\right) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{1}{4\pi} \int_{\Sigma} d^2\sigma \sqrt{g} \delta g_{ij} b^{ij}\right)^n.$$

Each power of δg accompanies a power of b , and because of the ghost number anomaly, a non-zero contribution comes only from $n = 6g - 6$. So actually,

⁴The vertical bar in $F(g|\delta g)$ is meant to remind us that g is bosonic or even and δg is fermionic or odd. We often write $F(t^1 \dots t^p | \theta^1 \dots \theta^q)$ for a function F that depends on even variables $t^1 \dots t^p$ and odd variables $\theta^1 \dots \theta^q$.

$F(g|\delta g)$ is homogeneous in δg of degree $6g - 6$. (We are limited here to $g > 1$ as we have assumed that c has no zero-modes. In section 2.4, we will make vertex operator insertions and then there is no such restriction.)

Acting as in (2.7) on a function that depends only on g and δg , Q_B can be written in more detail as

$$(2.9) \quad Q_B = \int_{\Sigma} d^2\sigma \sum_{i,j=1,2} \sqrt{g} \delta g_{ij} \frac{\delta}{\delta g_{ij}}.$$

At this point, it is good to remember that on a manifold M with local coordinates x^i , $i = 1, \dots, s = \dim M$, one introduces odd variables dx^i (which are usually called one-forms) and defines the exterior derivative operator

$$(2.10) \quad d = \sum_{i=1}^s dx^i \frac{\partial}{\partial x^i},$$

which obeys $d^2 = 0$. A function $F(x^1 \dots x^s | dx^1 \dots dx^s)$ that is homogeneous of degree p in the odd variables has an expansion

$$(2.11) \quad F(x^1 \dots | \dots dx^s) = \sum_{i_1 < \dots < i_p} F_{i_1 \dots i_p}(x^1 \dots x^s) dx^{i_1} \dots dx^{i_p}$$

and is usually called a p -form. So d maps p -forms to $(p + 1)$ -forms; a p -form ω is said to be closed if $d\omega = 0$ and to be exact if $\omega = d\lambda$ for some $(p - 1)$ -form λ . If we just identify δg_{ij} with dg_{ij} , we can think of the operator on the left hand side of (2.9) as the exterior derivative on the space \mathcal{J} of all of conformal structures on Σ (that is, the space of metrics on Σ modulo Weyl transformations): the integral over Σ and sum over $i, j = 1, 2$ in (2.9) is the analog of the sum over i in (2.10). So in fact, we can view $F(g|\delta g)$ as a $(6g - 6)$ -form on \mathcal{J} . With this interpretation, (2.7) says that this form is closed, $dF = 0$.

2.2. Reducing to moduli space

To get farther, we must consider diffeomorphism invariance. We want to interpret $F(g|\delta g)$ as a form not on the infinite-dimensional space \mathcal{J} , but on the quotient $\mathcal{M}_g = \mathcal{J}/\mathcal{D}$, where \mathcal{D} is the group of orientation-preserving diffeomorphisms of Σ . \mathcal{M}_g is the moduli space of Riemann surfaces of genus g . To show that $F(g|\delta g)$ is the “pullback” from \mathcal{M}_g of a differential form on \mathcal{M}_g , we need to establish two things (see appendix A):

(1) $F(g|\delta g)$ must be \mathcal{D} -invariant. This is actually manifest from the diffeomorphism-invariance of the whole construction.

(2) $F(g|\delta g)$ must vanish if contracted with one of the vector fields on \mathcal{J} that generates the action of \mathcal{D} . Such a vector field is

$$(2.12) \quad g_{ij} \rightarrow g_{ij} + \epsilon(D_i v_j + D_j v_i),$$

where v^i is an ordinary vector field on Σ and ϵ is an infinitesimal parameter. The operation of contraction with this vector field transforms δg_{ij} by

$$(2.13) \quad \delta g_{ij} \rightarrow \delta g_{ij} + \epsilon(D_i v_j + D_j v_i).$$

(It may be helpful to compare to the finite-dimensional formulas (A.1) and (A.2).) For $F(g|\delta g)$ to be invariant under this shift of δg_{ij} means precisely that

$$(2.14) \quad \int_{\Sigma} d^2\sigma \sqrt{g} (D_i v_j + D_j v_i) \frac{\delta}{\delta(\delta g_{ij})} F(g|\delta g) = 0.$$

Explicitly, given the definition (2.6) of $F(g|\delta g)$, the requirement is

$$(2.15) \quad \int \mathcal{D}(X, b, c) \exp(-\widehat{I}) \int_{\Sigma} d^2\sigma \sqrt{g} (D_i v_j + D_j v_i) b^{ij} = 0.$$

This is indeed true, as we learn upon integrating by parts and using the equation of motion

$$(2.16) \quad D_i b^{ij} = 0.$$

So in fact, $F(g|\delta g)$ is the pullback to \mathcal{J} of a $(6g - 6)$ -form on \mathcal{M}_g . This form, moreover, is closed, since $F(g|\delta g)$ is BRST-invariant. Actually, once we know that $F(g|\delta g)$ is a pullback from \mathcal{M}_g , we do not really need BRST symmetry to prove that $dF = 0$. This is automatically true; the dimension of \mathcal{M}_g is $6g - 6$, so a $(6g - 6)$ -form on \mathcal{M}_g is automatically closed. However, as we will see in section 2.4, once we consider vertex operator insertions, the closedness of $F(g|\delta g)$ does give non-trivial information.

The equation (2.16) arises as the classical equation of motion $\delta \widehat{I} / \delta c^j = 0$. The use of the classical equation of motion is really a shorthand for the fact that the path integral is invariant under the change of variables

$$(2.17) \quad c^i \rightarrow c^i + \epsilon v^i,$$

with ϵ a parameter. Under this transformation, the measure is invariant and the action shifts by $\widehat{I} \rightarrow \widehat{I} - (\epsilon/2\pi) \int_{\Sigma} d^2\sigma \sqrt{g} b_{ij} D^i v^j$. To first order in ϵ , the path integral transforms by

$$(2.18) \quad \int \mathcal{D}(X, b, c) \exp(-\widehat{I}) \rightarrow \int \mathcal{D}(X, b, c) \exp(-\widehat{I}) \left(1 + \frac{\epsilon}{2\pi} \int_{\Sigma} d^2\sigma \sqrt{g} b_{ij} D^i v^j \right).$$

The fact that the right hand side is independent of ϵ gives (2.15).

In our derivation, we have not made use of the fact that \mathcal{J} and \mathcal{M}_g have complex structures. In fact, picking local holomorphic and antiholomorphic coordinates⁵ z and \tilde{z} on Σ , we can decompose c^i and b_{ij} into holomorphic components c^z , b_{zz} and antiholomorphic components $\tilde{c}^{\tilde{z}}$, $b_{\tilde{z}\tilde{z}}$, which we often denote as c , b and \tilde{c} , \tilde{b} , respectively. One can define separate anomalous ghost number symmetries for c , b and for \tilde{c} , \tilde{b} . Each has an anomaly $-(3g-3)$. This means in practice that both b and \tilde{b} have $3g-3$ zero-modes on a surface Σ of genus g , so it requires $3g-3$ insertions of b and $3g-3$ insertions of \tilde{b} to get a non-zero path integral. From this, it follows that $F(g|\delta g)$ is a form of degree $(3g-3, 3g-3)$, in other words, its holomorphic and antiholomorphic degree are both $3g-3$.

A few more comments will be helpful background for the generalization to superstrings. By a zero-mode of b , we will mean simply a solution of the classical equation $D_i b^{ij} = 0$. Let us expand b_{ij} in c -number zero-modes \mathbf{b}_{α} , $\alpha = 1, \dots, 6g-6$, and non-zero modes \mathbf{b}'_{λ} :

$$(2.19) \quad b_{ij} = \sum_{\alpha=1}^{6g-6} \mathbf{u}_{\alpha} \mathbf{b}_{\alpha,ij} + \sum_{\lambda} \mathbf{w}_{\lambda} \mathbf{b}'_{\lambda,ij}.$$

Here \mathbf{u}_{α} and \mathbf{w}_{λ} are anticommuting coefficients. When we make this expansion, the coefficients \mathbf{u}_{α} do not appear in the original action $I(X, b, c)$, precisely because the corresponding modes are zero-modes. They do appear in the extended action \widehat{I} :

$$(2.20) \quad \widehat{I} = \dots + \frac{1}{4\pi} \sum_{\alpha=1}^{6g-6} \int_{\Sigma} d^2\sigma \sqrt{g} \delta g_{ij} \mathbf{u}_{\alpha} \mathbf{b}_{\alpha}^{ij}.$$

⁵We write \tilde{z} for what is more commonly called \bar{z} . Since all correlation functions are real-analytic, we can analytically continue slightly away from $\tilde{z} = \bar{z}$ and that is why we adopt this more flexible notation. The ability to make this analytic continuation is more important in superstring theory, in which a condition $\tilde{z} = \bar{z}$ would not be invariant under superconformal transformations.

The part of the path integral that depends on the zero-mode coefficients u_α is therefore particularly simple. This is a factor

$$(2.21) \quad \prod_{\alpha=1}^{6g-6} \int du_\alpha \exp\left(\frac{u_\alpha}{4\pi} \int_\Sigma d^2\sigma \sqrt{g} \delta g_{ij} b_\alpha^{ij}\right) = \prod_{\alpha=1}^{6g-6} \frac{1}{4\pi} \int_\Sigma d^2\sigma \sqrt{g} \delta g_{ij} b_\alpha^{ij}.$$

We have used the fact that if u and v are odd variables, then $\int du \exp(uv) = v$. On the right hand side of (2.21), we see explicitly that $F(g|\delta g)$ is proportional to $6g - 6$ factors of δg . It is instructive to recall that if v is an odd variable, then $v = \delta(v)$. So if we set

$$(2.22) \quad v_\alpha = \frac{1}{4\pi} \int_\Sigma d^2\sigma \sqrt{g} \delta g_{ij} b_\alpha^{ij},$$

then $F(g|\delta g)$ is proportional to

$$(2.23) \quad \prod_{\alpha=1}^{6g-6} \delta(v_\alpha).$$

Actually, (2.23) gives the complete dependence of $F(g|\delta g)$ on δg . To see this, we can assume that c has no zero-modes, since otherwise $F(g|\delta g) = 0$. So the mode expansion of c involves only non-zero modes c_λ with coefficients γ_λ :

$$(2.24) \quad c^i = \sum_\lambda \gamma_\lambda c_\lambda^i.$$

The part of the action \hat{I} that depends on the γ_λ has the general form $-\sum_\lambda m_\lambda w_\lambda \gamma_\lambda$ with non-zero constants m_λ , and the integral over the γ_λ is

$$(2.25) \quad \prod_\lambda \int D\gamma_\lambda \exp(m_\lambda \gamma_\lambda w_\lambda) = \prod_\lambda m_\lambda \cdot \prod_\lambda \delta(w_\lambda)$$

(here $\prod_\lambda m_\lambda$ is the determinant of the non-zero ghost and antighost modes; of course it requires regularization). The integral over w_λ can be performed with the aid of the δ functions in (2.25), so even though in the definition (2.5) of the extended action, the coefficients w_λ do couple to δg , this coupling does not affect the evaluation of $F(g|\delta g)$. Hence (2.21) or (2.23) does give the full dependence of $F(g|\delta g)$ on δg .

2.3. Integration over moduli space

Having understood that $F(g|\delta g)$ is the pullback of a differential form of top degree on \mathcal{M}_g , which we denote by the same name, we can formally define the genus g vacuum amplitude of string theory by integration over \mathcal{M}_g :

$$(2.26) \quad Z_g = \int_{\mathcal{M}_g} F(g|\delta g).$$

The only problem with this formal definition is that \mathcal{M}_g is not compact. This is an essential fact in string theory, since the region at infinity in \mathcal{M}_g is the infrared region that generates singularities – such as poles associated to on-shell particles and cuts associated to unitarity – that are essential to the physical interpretation of the theory. But by the same token, the non-compactness of \mathcal{M}_g makes infrared divergences possible. In bosonic string theory, in attempting to evaluate the integral (2.26), one indeed runs into infrared divergences associated to tachyons and massless scalar tadpoles. The real arena of application of a formula such as (2.26) is superstring theory, where in an appropriate class of tachyon-free vacua with spacetime supersymmetry, a similar formula arises (in a formalism based on super Riemann surfaces rather than ordinary ones) without the infrared divergences.

In addition, it is a little artificial to focus on the vacuum amplitude (2.26). What we really want to compute is the S -matrix, which is obtained by a generalization of what we have so far explained to include external vertex operators. We discuss this in section 2.4.

Here we want to explain a key point that is largely unaffected by these generalizations. We can explicitly evaluate the differential form that has to be integrated over \mathcal{M}_g using any local section of the fibration $\mathcal{J} \rightarrow \mathcal{M}_g$, or equivalently any explicit family of metrics $g_{ij}(\sigma; m_1, \dots, m_p)$ on Σ that depends on modular parameters m_1, \dots, m_p . Here the m_s , $s = 1, \dots, p = \dim \mathcal{M}_g$ are local coordinates on \mathcal{M}_g that we can interpret as moduli of Σ , and $g_{ij}(\sigma; m_1, \dots, m_p)$ is a corresponding family of metrics on Σ . (We call this “a” corresponding family of metrics, not “the” family, since the metric corresponding to a given point in \mathcal{M}_g is only determined up to diffeomorphism and Weyl transformation.)

The family of metrics $g_{ij}(\sigma; m_1, \dots, m_p)$ defines a local slice transverse to the action of the diffeomorphism group on \mathcal{J} , or equivalently a local section \mathcal{S} of the fibration $\mathcal{J} \rightarrow \mathcal{M}_g$. We do not need to worry about the global properties of \mathcal{S} ; the differential form on \mathcal{M}_g that we are trying to evaluate is intrinsically defined, and we can evaluate it in a given region of \mathcal{M}_g using any local slice that we choose.

Once we specify the metric g_{ij} in terms of the modular parameters m_1, \dots, m_p , we can similarly express $\delta g = \{Q_B, g\}$ in terms of m_s and $dm_s = \{Q_B, m_s\}$:

$$(2.27) \quad \delta g_{ij} = \sum_{s=1}^p \frac{\partial g_{ij}}{\partial m_s} dm_s.$$

Accordingly, we can evaluate the term in the action that involves δg :

$$(2.28) \quad \frac{1}{4\pi} \int_{\Sigma} d^2\sigma \sqrt{g} \delta g_{ij} b^{ij} = \frac{1}{4\pi} \sum_{s=1}^p dm_s \int_{\Sigma} d^2\sigma \frac{\partial(\sqrt{g}g_{ij})}{\partial m_s} b^{ij}.$$

After expressing $F(g|\delta g)$ in terms of m_s and dm_s , $s = 1, \dots, p$, we get a differential form $F(m_1, \dots, m_p | dm_1, \dots, dm_p)$ that can be integrated over the variables m_s and dm_s . From the definition of $F(g|\delta g)$, we have

$$(2.29) \quad \begin{aligned} F(g|\delta g) &= \int \mathcal{D}(X, b, c) \exp(-\hat{I}) \\ &= \int \mathcal{D}(X, b, c) \exp\left(-I - \frac{1}{4\pi} \sum_{s=1}^p dm_s \int_{\Sigma} d^2\sigma \frac{\partial(\sqrt{g}g_{ij})}{\partial m_s} b^{ij}\right). \end{aligned}$$

$F(g|\delta g)$ is an inhomogeneous differential form, rather than a form of definite degree, since the right hand side is not of definite degree in dm_1, \dots, dm_p . The only part that can be integrated over \mathcal{M}_g is the part of top degree p . This is obtained by expanding the exponent in (2.29) and picking out the term that is proportional to each of the odd variables dm_1, \dots, dm_p :

$$(2.30) \quad \begin{aligned} &F_{\text{top}}(m_1 \dots | \dots dm_p) \\ &= (-1)^{p(p+1)/2} dm_1 \dots dm_p \int \mathcal{D}(X, b, c) e^{-I} \prod_{s=1}^p \frac{1}{4\pi} \int_{\Sigma} d^2\sigma \frac{\partial(\sqrt{g}g_{ij})}{\partial m_s} b^{ij}. \end{aligned}$$

In short, for each modulus m_s over which one wishes to integrate, one must insert in the path integral a factor of

$$(2.31) \quad \Psi_s = \frac{1}{4\pi} \int_{\Sigma} d^2\sigma \frac{\partial(\sqrt{g}g_{ij})}{\partial m_s} b^{ij}.$$

⁶Once we get down to a finite-dimensional slice, we denote $\{Q_B, m_s\}$ as a one-form dm_s , rather than using the notation δm_s , which is suggestive of infinite dimensions.

This is a standard result (see [30] or eqn. (5.3.9) of [32]). The integral over \mathcal{M}_g is

$$(2.32) \quad (-1)^{p(p+1)/2} \int \mathcal{D}(m_1 \dots | \dots dm_p) dm_1 \dots dm_p \\ \times \int \mathcal{D}(X, b, c) \exp(-I) \Psi_1 \dots \Psi_p.$$

Since the Ψ_i are odd variables, we have $\Psi_i = \delta(\Psi_i)$, and we could equivalently write

$$(2.33) \quad (-1)^{p(p+1)/2} \int \mathcal{D}(m_1 \dots | \dots dm_p) dm_1 \dots dm_p \\ \times \int \mathcal{D}(X, b, c) \exp(-I) \delta(\Psi_1) \dots \delta(\Psi_p).$$

One can perform the Berezin integral over the odd variables dm_s , $s = 1, \dots, p$ and write an equivalent formula

$$(2.34) \quad \int \mathcal{D}(m_1, \dots, m_p) \int \mathcal{D}(X, b, c) \exp(-I) \delta(\Psi_1) \dots \delta(\Psi_p)$$

that involves integration over only the m_s and not the dm_s . This is a more standard version of the formula. Any of the formulas in this paragraph expresses the form that has to be integrated over \mathcal{M}_g to compute the vacuum amplitude of the bosonic string in terms of correlation functions on the Riemann surface Σ .

We used a local slice in writing the formula, but the formula is independent of the slice. Concretely, the mechanism by which this occurs is that the factors $\delta(\Psi_s)$ involve both zero-modes and non-zero modes of b , but when we perform the integral over c , the coefficients of non-zero modes are set to zero, as in the discussion following eqn. (2.23), and only the zero-modes remain. At that point the dependence on the slice drops out. Hopefully, the reader can see that the derivation of (2.34) and the other formulas of the last paragraph is a close relative of the derivation of eqn. (2.23). In each case, the key was the coupling of δg to the zero-modes of b . To get eqn. (2.23), we integrated over those zero-modes to get a function of δg , while to get (2.34), we integrated over δg to get a function of the zero-modes.

2.4. Vertex operator insertions

Typically, we want to calculate not the vacuum amplitude (2.6) but a more general amplitude with insertion of some function $\Omega(X, b, c, g, \delta g)$:

$$(2.35) \quad F_\Omega(g|\delta g) = \int \mathcal{D}(X, b, c) \exp\left(-\widehat{I}(X, b, c, g, \delta g)\right) \Omega.$$

The right hand side of (2.35) is invariant if we transform all variables $\Phi = X, b, c, g$, or δg by $\Phi \rightarrow (1 + \epsilon Q_B)\Phi$, with a small parameter ϵ . In general, Ω may not be invariant under this transformation, so from this invariance we deduce not that $dF_\Omega = 0$, but that

$$(2.36) \quad dF_\Omega + F_{Q_B}\Omega = 0.$$

The usual case is that Ω is a product of vertex operators \mathcal{V}_s , $s = 1, \dots, n$ inserted at points $p_s \in \Sigma$:

$$(2.37) \quad \Omega = \prod_{s=1}^n \mathcal{V}_s(X, b, c; p_s).$$

The points p_s are often called marked points or punctures. In this case, we write $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}$ for F_Ω .

2.4.1. Conformal vertex operators We will consider only the simplest class of operators that suffices for computing the S -matrix. These are operators of the form $\mathcal{V}_s = \tilde{c}V_s$, where V_s is a (1,1) primary field⁷ constructed from the matter fields only. The reason that these are the simplest operators to use is as follows.

To express S -matrix elements in terms of integrals of naturally-defined differential forms on the appropriate finite-dimensional moduli space, we will need the vertex operators to depend only on c and \tilde{c} but not their derivatives. (See section 2.4.2.) An equivalent condition⁸ is that

$$(2.38) \quad b_n \mathcal{V}_s = \tilde{b}_n \mathcal{V}_s = 0, \quad n \geq 0,$$

where b_n and \tilde{b}_n are the antighost modes

$$(2.39) \quad b_n = \frac{1}{2\pi i} \oint dz z^{n+1} b_{zz}, \quad \tilde{b}_n = \frac{1}{2\pi i} \oint d\tilde{z} \tilde{z}^{n+1} \tilde{b}_{\tilde{z}\tilde{z}}.$$

⁷ By an (n, m) primary field, we mean a primary field of holomorphic and anti-holomorphic conformal dimensions n and m .

⁸This condition may have first been stated in eqn. (5.18) of [31]. See also [18].

If $Q_B \mathcal{V}_s = 0$ and \mathcal{V}_s also obeys (2.38), then \mathcal{V}_s is annihilated for $n \geq 0$ by $L_n = \{Q_B, b_n\}$ and by $\tilde{L}_n = \{Q_B, \tilde{b}_n\}$, $n \geq 0$. So \mathcal{V}_s will be a conformal primary of dimension $(0, 0)$.

The most simple formalism also represents the external states by vertex operators of ghost number 2; the usual physical states appear at this ghost number, and this value will lead to differential forms on moduli space of the right degree to be integrated, as we will see. Even without assuming BRST invariance, the most general vertex operator of ghost number 2 that does not involve derivatives of c and \tilde{c} is $\mathcal{V}_s = \tilde{c}cV_s$, where V_s is constructed from matter fields only. Such a \mathcal{V}_s is BRST-invariant if and only if V_s is a conformal primary field of dimension $(1, 1)$, and in this case \mathcal{V}_s is indeed a conformal primary of dimension $(0, 0)$. So these are the simplest vertex operators to use and we will call them conformal vertex operators.

Every physical state of string theory with non-zero momentum⁹ can be represented by a conformal vertex operator, so vertex operators of this type suffice for computing the S -matrix. However, the representation of a physical state by a conformal vertex operator is in general not unique. A matter primary V_s of dimension $(1, 1)$ may be a null vector, in which case the corresponding vertex operator \mathcal{V}_s is a BRST commutator, $\mathcal{V}_s = \{Q_B, \mathcal{W}_s\}$. We will have to demonstrate gauge-invariance of the S -matrix, by which we mean that a conformal vertex operator of the form $\mathcal{V}_s = \{Q_B, \mathcal{W}_s\}$ decouples from the S -matrix. In doing so, an important fact, explained in appendix B, will be that if a conformal vertex operator \mathcal{V}_s is of the form $\{Q_B, \mathcal{W}_s\}$, then \mathcal{W}_s can be chosen to obey the same conditions as \mathcal{V}_s :

$$(2.40) \quad b_n \mathcal{W}_s = \tilde{b}_n \mathcal{W}_s = 0, \quad n \geq 0.$$

This is equivalent to saying that just like \mathcal{V}_s , \mathcal{W}_s is constructed only from c and \tilde{c} and not their derivatives. Given this, \mathcal{W}_s is a conformal primary of dimension $(0, 0)$, just like \mathcal{V}_s . Indeed, we have $L_n \mathcal{W}_s = \{Q_B, b_n\} \mathcal{W}_s = Q_B b_n \mathcal{W}_s + b_n \mathcal{V}_s$, and this vanishes for $n \geq 0$, since $b_n \mathcal{V}_s = b_n \mathcal{W}_s = 0$.

For insertions of conformal vertex operators, (2.36) reduces to $dF_{\mathcal{V}_1, \dots, \mathcal{V}_n} = 0$, since conformal vertex operators are Q_B -invariant not just in the usual

⁹This assertion is part of the BRST version of the no-ghost theorem; see [33–36], or, for example, section 4.2 of [32]. (For the original no-ghost theorem, see [37, 38].) The usual proof holds in compactification to $d \geq 2$ dimensions – or more precisely if the matter conformal field theory has at least 2 free fields. In typical bosonic string compactifications, what we miss by considering only conformal vertex operators is the ability to separate the zero momentum dilaton from the trace of the metric. (See section 7.2.2.) This does not affect the ability to compute the S -matrix.

sense but also in the extended sense. The usual Q_B transformations for an operator constructed just from matter fields X , ghosts c , and the world sheet metric g – such as a conformal vertex operator \mathcal{V}_s – are

$$\begin{aligned}
 \delta c^i &= c^j \partial_j c^i \\
 \delta X &= c^j \partial_j X \\
 (2.41) \quad \delta g_{ij} &= D_i c_j + D_j c_i - g_{ij} D_k c^k.
 \end{aligned}$$

A conformal vertex operator $\mathcal{V}_s(p)$ is invariant under this transformation. In the extended formalism that we use here, only one thing is different: instead of δg_{ij} being defined as in (2.41), it is treated, off-shell, as an arbitrary symmetric traceless tensor. However, on-shell the formula $\delta g_{ij} = D_i c_j + D_j c_i - g_{ij} D_k c^k$ is valid since this is the equation of motion for the antighost field b derived from the extended action \hat{I} of eqn. (2.5). In general, the equations of motion may be used in proving the invariance of an operator under a symmetry. One may be slightly surprised at the need here to use the antighost equation of motion; however, in standard approaches, this is necessary after gauge-fixing. Customarily, after gauge-fixing, g_{ij} is treated as a c -number and so δg_{ij} is set to 0; after doing so, Q_B -invariance of conformal vertex operators relies on the fact that the equation of motion for b derived from the ordinary action I is $D_i c_j + D_j c_i - g_{ij} D_k c^k = 0$.

2.4.2. Integration on moduli space As we have defined it so far, F_{ν_1, \dots, ν_n} is a differential form on the space \mathcal{J} of conformal structures on Σ . However, as in section 2.1, we can interpret F_{ν_1, \dots, ν_n} as the pullback of a differential form on the appropriate moduli space, which in the present case is the moduli space $\mathcal{M}_{g,n}$ that parametrizes a Riemann surface Σ of genus g with n punctures p_1, \dots, p_n . Let $\mathcal{D}_{p_1, \dots, p_n}$ be the group of orientation-preserving diffeomorphisms of Σ that are trivial at p_1, \dots, p_n . This group is generated by vector fields v on Σ that vanish at p_1, \dots, p_n :

$$(2.42) \quad v^i(p_1) = \dots = v^i(p_n) = 0.$$

$\mathcal{M}_{g,n}$ is a quotient:

$$(2.43) \quad \mathcal{M}_{g,n} = \mathcal{J} / \mathcal{D}_{p_1, \dots, p_n}.$$

Just as in section 2.1, for $F_{\nu_1, \dots, \nu_n}(g|\delta g)$ to be a pullback from $\mathcal{M}_{g,n}$, we need two properties:

(1) $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(g|\delta g)$ must be $\mathcal{D}_{p_1, \dots, p_n}$ -invariant. This is manifest from the definition.

(2) $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(g|\delta g)$ must also be annihilated by contraction with any of the vector fields generating $\mathcal{D}_{p_1, \dots, p_n}$. The condition is as in eqn. (2.15) but now with the external vertex operators included:

$$(2.44) \quad \int \mathcal{D}(X, b, c) \exp(-\widehat{I}) \prod_{s=1}^n \mathcal{V}_s(X, b, c; p_s) \int_{\Sigma} d^2\sigma \sqrt{g} (D_i v_j + D_j v_i) b^{ij} = 0.$$

To prove this using the change of variables (2.17), we need to know that the vertex operators $\mathcal{V}_s(p_s)$ are invariant under $c^i \rightarrow c^i + \epsilon v^i$, so that including these vertex operators does not modify the previous argument. This is true because $\mathcal{V}_s(p_s)$ is proportional to $c(p_s)$ and $\tilde{c}(p_s)$ but not their derivatives, and moreover $v^i(p_s) = 0$ as in (2.42).

Having understood that $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(g|\delta g)$ is the pullback of a differential form on $\mathcal{M}_{g,n}$ – which we denote by the same name – we can give a formal recipe for computing S -matrix elements. If the vertex operators \mathcal{V}_s all have ghost number 2, then the number of antighost insertions needed to get a non-zero path integral is $6g - 6 + 2n$, so that is the degree of $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}$. But this number is the same as the real dimension of $\mathcal{M}_{g,n}$, so formally we can integrate $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}$ over $\mathcal{M}_{g,n}$:

$$(2.45) \quad \langle \mathcal{V}_1 \dots \mathcal{V}_n \rangle_g = \int_{\mathcal{M}_{g,n}} F_{\mathcal{V}_1, \dots, \mathcal{V}_n}.$$

This integral is supposed to give the genus g contribution to the scattering amplitudes of the string states corresponding to $\mathcal{V}_1, \dots, \mathcal{V}_n$.

Just as in section 2.3, the only problem is that $\mathcal{M}_{g,n}$ is not compact. The integrals (2.45) are infrared-divergent because of tachyons and massless tadpoles, and hence in bosonic string theory the definition of the S -matrix is purely formal. The real arena of application of the ideas that we have explained is superstring theory, which admits a similar formalism, but with a sensible infrared behavior. At the most basic level, the infrared behavior of superstring theory is sensible because supersymmetric theories are tachyon-free; more delicate questions of massless tadpoles will be discussed in section 8.

2.4.3. Gauge invariance Now let us discuss the gauge-invariance of the formalism. Gauge-invariance of the S -matrix means that the integral (2.45) must vanish if we replace one of the vertex operators, say \mathcal{V}_1 , with $\{Q_B, \mathcal{W}_1\}$:

$$(2.46) \quad \int_{\mathcal{M}_{g,n}} F_{\{Q_B, \mathcal{W}_1\}, \mathcal{V}_2, \dots, \mathcal{V}_n} = 0.$$

To explore this question, we can invoke (2.36):

$$(2.47) \quad dF_{\mathcal{W}_1, \nu_2, \dots, \nu_n} + F_{\{Q_B, \mathcal{W}_1\}, \nu_2, \dots, \nu_n} = 0.$$

As explained in appendix B, we can assume here that \mathcal{W}_1 is annihilated by b_n and \tilde{b}_n , $n \geq 0$, or equivalently that it is constructed from c and \tilde{c} but not their derivatives. This ensures that $\mathcal{W}_1(p_1)$ is invariant under $c^i \rightarrow c^i + \epsilon v^i$, where v^i vanishes at the points p_s . So the same argument that we have already given shows that $F_{\mathcal{W}_1, \nu_2, \nu_3, \dots, \nu_n}$ is the pullback of a differential form on $\mathcal{M}_{g,n}$, in this case a form of degree $6g - 7 + 2n = \dim \mathcal{M}_{g,n} - 1$. So we can understand eqn. (2.47) as a relation between differential forms on $\mathcal{M}_{g,n}$, and in particular we have

$$(2.48) \quad \int_{\mathcal{M}_{g,n}} F_{\{Q_B, \mathcal{W}_1\}, \nu_2, \dots, \nu_n} = - \int_{\mathcal{M}_{g,n}} dF_{\mathcal{W}_1, \nu_2, \dots, \nu_n}.$$

If $\mathcal{M}_{g,n}$ were compact, we would now invoke Stokes’s theorem to deduce the vanishing of the right hand side of eqn. (2.48), and this would establish gauge-invariance. Since actually $\mathcal{M}_{g,n}$ is non-compact, to demonstrate the vanishing of the right hand side of eqn. (2.48), we have to integrate by parts and show that there are no surface terms arising in the infrared region at infinity. Because of the unphysical infrared singularities of bosonic string theory, the proper framework for this discussion is really superstring theory. We return to this question in section 7, after describing the appropriate superstring measure and gaining some information about its behavior in the infrared region.

2.4.4. More general vertex operators Had we represented our external states by arbitrary Q_B -invariant vertex operators,¹⁰ rather than the special ones we actually used, we would have needed a more elaborate formalism [17, 18]. This does not seem necessary for the purposes of the present paper, so we will just give a few hints of what is involved. If the \mathcal{V}_s (or the gauge parameters \mathcal{W}_s in the above analysis) are Q_B -invariant and primary, but depend on derivatives of c and \tilde{c} , we can define $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}$ (or $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$) in a conformally invariant fashion as a closed form on the space \mathcal{J} of conformal structures, but it is not a pullback from $\mathcal{J}/\mathcal{D}_{p_1, \dots, p_n}$. However, at a given mass level, the \mathcal{V}_s and \mathcal{W}_s can only depend on derivatives of c and \tilde{c} up to some finite

¹⁰Actually, even in a very general formalism, the vertex operators should be constrained to be annihilated by $b_0 - \tilde{b}_0$, for a reason that will appear in the next paragraph. It follows that they are also annihilated by $L_0 - \tilde{L}_0 = \{Q_B, b_0 - \tilde{b}_0\}$.

order N . So $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}$ (or $F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n}$) is always a pullback from $\mathcal{J}/\mathcal{D}_{p_1, \dots, p_n}^{(N+1)}$, where $\mathcal{D}_{p_1, \dots, p_n}^{(N+1)}$ consists of orientation-preserving diffeomorphisms of Σ that are trivial up to order $N + 1$ near the p_s . If the \mathcal{V}_s or \mathcal{W}_s are not primary, one must drop the assumption of conformal invariance and endow Σ with a metric (or a suitable system of local parameters) near the punctures p_s .

In either or both of these cases, one can define $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}$ (or $F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n}$) as a closed form of the appropriate degree on a space $\mathcal{Z}_{g,n}$ that is a fiber bundle over $\mathcal{M}_{g,n}$. Provided that the vertex operators are all annihilated by $b_0 - \tilde{b}_0$, one can define $\mathcal{Z}_{g,n}$ so that its fibers are contractible (see [18]). This ensures that it is possible to choose a section Υ of the fibration $\mathcal{Z}_{g,n} \rightarrow \mathcal{M}_{g,n}$ and that all such sections are homologous. One defines the scattering amplitudes by integrating $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}$ over Υ . This procedure would give a unique and gauge-invariant result, independent of the choice of Υ , if $\mathcal{M}_{g,n}$ were compact. So as usual the only real subtleties – beyond the need for a somewhat intricate formalism – involve the behavior in the infrared region at infinity.

2.4.5. Open and/or unoriented strings So far we have considered oriented closed strings only. The generalizations to open and/or unoriented strings do not change much of what we have said so far.

In a theory of oriented open and closed strings, the string worldsheet Σ is an oriented two-manifold that may have a boundary. Open-string vertex operators are inserted at boundary points. The simplest open-string vertex operators are of the form $\mathcal{U} = cU$, where U is a primary field of dimension 1 constructed from matter fields only. The worldsheet path integral may be analyzed along the above lines and – in the presence of both closed- and open-string vertex operators – gives rise to a naturally defined differential form on an appropriate moduli space of Riemann surfaces with boundary and in general with both bulk and boundary punctures. To get a result of this form, one needs the fact that $b_n \mathcal{U} = 0$, $n \geq 0$; this is the case for $\mathcal{U} = cU$.

In a theory of unoriented strings, the only change is that the worldsheet Σ is unoriented, and the vertex operators must be defined in a way that does not depend on a choice of orientation; they must be invariant under exchange of local holomorphic and antiholomorphic variables.

We will say much more about open and/or unoriented strings in section 9. Until that point, we illustrate most ideas with closed oriented strings; we consider open strings when this is illuminating or there is something distinctive to say.

2.5. Integrated vertex operators

2.5.1. Two types of vertex operator As an important illustration of some of these ideas, we will explain the relation between unintegrated and integrated vertex operators. See for example pp. 163-4 of [32].

Consider a closed-string vertex operator of the form $\mathcal{V} = \tilde{c}cV$, where V is a primary of dimension $(1, 1)$. A primary of dimension $(1, 1)$ is a two-form or measure that can be integrated over the worldsheet in a natural way. So instead of inserting \mathcal{V} at a point in Σ , we could consider inserting a factor of $\int_{\Sigma} V$, which one might prefer to write as $\int_{\Sigma} dzd\tilde{z}V_{z\tilde{z}}$. The operator \mathcal{V} , inserted at a point in Σ , is called the unintegrated form of the vertex operator, while the insertion $\int_{\Sigma} V$ (or sometimes just the operator V) is called the integrated form.

Roughly speaking, it is equivalent to insert \mathcal{V} at a point in Σ – the choice of this point is then one of the moduli over which we integrate in computing the scattering amplitude – or to insert a factor of $\int_{\Sigma} V$.

Similarly, for open strings, it is, roughly speaking, equivalent to insert an unintegrated vertex operator $\mathcal{U} = cU$ or an integrated vertex operator $\oint_{\partial\Sigma} U$. Here U is integrated over the boundary of Σ or a component thereof (if several open-string vertex operators are inserted on the same boundary component, then as usual one can fix their cyclic order).

The formalism based on unintegrated vertex operators is always correct when used in conjunction with the Deligne-Mumford compactification of the moduli space of Riemann surfaces or super Riemann surfaces¹¹ and with an appropriate infrared regulator. To make this clear is one of the main goals of the present paper.

By contrast, relying on integrated vertex operators, though extremely useful in practice, will run into trouble for certain questions involving behavior at exceptional momenta, including but not limited to questions about mass renormalization and massless tadpoles. (For some examples, see [39, 42, 43]; the relevant models were studied at the level of low energy effective field theory in [44].) The basic reason for this is illustrated in fig. 1: the Deligne-Mumford compactification is constructed in such a way that external vertex operators never collide, while collisions of vertex operators occur if one represents an external string state by an integrated insertion $\int_{\Sigma} V$ (the integral is over all of Σ including the locations of other vertex operators). Accordingly,

¹¹The Deligne-Mumford compactification for super Riemann surfaces was constructed by Deligne in the 1980s [40]. This work is unfortunately unpublished, but has been presented in a series of lectures [41]. An introduction to the Deligne-Mumford compactification can be found, for example, in section 6 of [16].

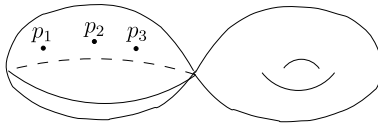


Figure 1: On the locus in moduli space at which naively speaking several punctures in a Riemann surface Σ collide, what actually happens in the Deligne-Mumford compactification is that Σ splits off a genus zero component that contains the punctures. This is illustrated here for the case of three coinciding punctures, labeled p_1, p_2, p_3 . The genus zero component in this example has four distinguished points – the p_i and the “node” at which it meets the other component of Σ – so it has one complex modulus. This modulus is lost if one represents the vertex operator insertions at the p_i (or even 2 of them) in integrated form. Because of this, the use of integrated vertex operators does not treat correctly questions for which this region of the moduli space is important. The difficulties arise when the total momentum flowing through the node is on-shell.

the map from integrated to unintegrated vertex operators, which we derive in section 2.5.2, works only over the uncompactified moduli space $\mathcal{M}_{g,n}$, where all punctures are distinct, and does not give a good treatment of the compactification.

Still, the formalism with integrated vertex operators is important as it is usually the simplest approach when it is applicable and it is the basis for many practical calculations. A sufficient criterion that, with minor extensions, accounts for most successful uses of integrated vertex operators is that one may use integrated vertex operators if the external momenta are sufficiently generic that the sum of the momenta in any set of colliding vertex operators is off-shell.¹² This ensures that one can compute by analytic continuation from a region of external momenta in which the subtleties of the compactification are irrelevant.

2.5.2. The derivation Let z and \bar{z} be local holomorphic and antiholomorphic coordinates on Σ . We could insert a vertex operator $\mathcal{V} = c\bar{c}V$ at any

¹²For example, in computing a genus 0 scattering amplitude, three vertex operators are fixed and one usually uses the integrated form of all other vertex operators. If the three fixed vertex operators have reasonably generic momenta, then the criterion just stated is satisfied. This justifies the usual computation of tree-level scattering amplitudes.

point

$$(2.49) \quad z = \mathbf{z}, \quad \tilde{z} = \tilde{\mathbf{z}}.$$

We are going to evaluate the formula (2.28) – or any of the related formulas of section 2.3 – with two of the moduli m_s of Σ being the choice of \mathbf{z} , $\tilde{\mathbf{z}}$.

In the formalism of section 2.4, we are instructed to keep fixed the point p at which \mathcal{V} is inserted and instead vary the metric of Σ . How then do we vary \mathbf{z} and $\tilde{\mathbf{z}}$ keeping everything else fixed? We do this by letting the metric g_{ij} of Σ depend on \mathbf{z} and $\tilde{\mathbf{z}}$ in a way that can be removed by a diffeomorphism, but not by a diffeomorphism that leaves fixed the point p .

We will make a diffeomorphism between two different parametrizations of Σ , which we will call (i) and (ii). In (i), we describe Σ by local complex coordinates w , \tilde{w} (which for what follows may as well be complex conjugates). We suppose that a vertex operator $\mathcal{V} = \tilde{c}eV$ is inserted at the point $w = \tilde{w} = 0$, which we call p . We assume also that the parameter w is well-defined in the disc $|w| \leq 1$, and that no vertex operators other than \mathcal{V} are inserted in this disc. Here $|w| = \sqrt{w\tilde{w}}$. In general, w is a complex parameter, not necessarily holomorphic, but it will be a holomorphic parameter near $w = 0$.

In description (ii), we use holomorphic and antiholomorphic local coordinates z and \tilde{z} (which may as well be complex conjugates). The two descriptions will be related by a diffeomorphism that depends on a parameter \mathbf{z} – which will be one of the moduli of Σ – and which we will restrict to the region $|\mathbf{z}| \leq \epsilon$ for some small ϵ . We want

$$(2.50) \quad z = \begin{cases} w + \mathbf{z} & \text{if } |w| < \epsilon \\ w & \text{if } |w| > 1 - \epsilon. \end{cases}$$

For example, we can have

$$(2.51) \quad z = w + f(|w|)\mathbf{z},$$

where $f(|w|)$ is a monotonic function that equals 1 for $|w| < \epsilon$ and 0 for $|w| > 1 - \epsilon$. We similarly express \tilde{z} in terms of w and \tilde{w} and also $\tilde{\mathbf{z}}$:

$$(2.52) \quad \tilde{z} = \tilde{w} + f(|w|)\tilde{\mathbf{z}}.$$

The coordinate transformations (2.51) and (2.52) are not holomorphic in general, though they are holomorphic when restricted to the region $|w| \leq \epsilon$. So although we assume z to be a holomorphic parameter, w is only holomorphic for $|w| < \epsilon$.

In description (ii), the vertex operator \mathcal{V} is inserted at a \mathbf{z} -dependent position, namely $z = \mathbf{z}$, $\tilde{z} = \tilde{\mathbf{z}}$. But we assume that in description (ii), the metric of Σ is independent of \mathbf{z} . In description (i), \mathcal{V} is inserted at $w = \tilde{w} = 0$, independent of \mathbf{z} , but the metric of Σ will then have to depend on \mathbf{z} and $\tilde{\mathbf{z}}$. Since this \mathbf{z} and $\tilde{\mathbf{z}}$ -dependence of the metric could be removed by using the coordinates z, \tilde{z} instead of w, \tilde{w} , it is of the “pure gauge” form

$$(2.53) \quad \begin{aligned} \partial_{\mathbf{z}} g_{IJ} &= D_I v_J + D_J v_I \\ \partial_{\tilde{\mathbf{z}}} g_{IJ} &= D_I \tilde{v}_J + D_J \tilde{v}_I, \end{aligned}$$

for some vector fields v^I and \tilde{v}^I . In these formulas, one can consider the indices I, J to take the values w, \tilde{w} . We can describe v by its components $(v^w, v^{\tilde{w}})$ in this basis. Up to sign,¹³ these components are just the derivatives of (w, \tilde{w}) with respect to \mathbf{z} at fixed z, \tilde{z} , since those derivatives determine how w, \tilde{w} must vary with \mathbf{z} to keep z, \tilde{z} fixed:

$$(2.54) \quad (v^w, v^{\tilde{w}}) = - \left(\partial_{\mathbf{z}}|_{z, \tilde{z}} w, \partial_{\mathbf{z}}|_{z, \tilde{z}} \tilde{w} \right).$$

To evaluate these derivatives in general, we would need to differentiate the function $f(|w|)$ that appears in eqns. (2.51) and (2.52). But we are only interested in what happens near $w = 0$. For $|w| < \epsilon$, we have $z = w + \mathbf{z}$, $\tilde{z} = \tilde{w} + \tilde{\mathbf{z}}$, so the evaluation of the derivatives in (2.54) just gives

$$(2.55) \quad (v^w, v^{\tilde{w}}) = (1, 0).$$

Similarly, near $w = 0$, we have

$$(2.56) \quad (\tilde{v}^w, \tilde{v}^{\tilde{w}}) = (0, 1).$$

Now we want to evaluate the measure of integration over the modular parameters \mathbf{z} and $\tilde{\mathbf{z}}$. For this, we use description (i), since it satisfies the assumptions of our general formalism that the metric of Σ varies but the insertion points of vertex operators are held fixed. The measure for integration over \mathbf{z} and $\tilde{\mathbf{z}}$ is

$$(2.57) \quad d\mathbf{z} d\tilde{\mathbf{z}} \Psi_{\mathbf{z}} \Psi_{\tilde{\mathbf{z}}},$$

¹³To understand the sign, observe that a function h that depends only on z and \tilde{z} is annihilated by $\partial_{\mathbf{z}} + \partial_{\mathbf{z}} w|_{z, \tilde{z}} \partial_w + \partial_{\mathbf{z}} \tilde{w}|_{z, \tilde{z}} \partial_{\tilde{w}}$, and so obeys $\partial_{\mathbf{z}} h = -(\partial_{\mathbf{z}} w|_{z, \tilde{z}} \partial_w + \partial_{\mathbf{z}} \tilde{w}|_{z, \tilde{z}} \partial_{\tilde{w}})h$, with a minus sign. In (2.54), we are applying the same idea for a metric that depends only on z and \tilde{z} .

where we evaluate (2.31) to get

$$(2.58) \quad \Psi_{\mathbf{z}} = \frac{1}{2\pi} \int_{\Sigma} dw d\tilde{w} \sqrt{g} b^{ij} D_i v_j, \quad \Psi_{\tilde{\mathbf{z}}} = \frac{1}{2\pi} \int_{\Sigma} dw d\tilde{w} \sqrt{g} b^{ij} D_i \tilde{v}_j.$$

We have to evaluate the product

$$(2.59) \quad \Psi_{\mathbf{z}} \Psi_{\tilde{\mathbf{z}}} \tilde{c}V(0).$$

with an insertion of $\mathcal{V} = \tilde{c}V$ at $w = \tilde{w} = 0$. If we had no insertion $\tilde{c}V(0)$, we would simply say that $\Psi_{\mathbf{z}} = \Psi_{\tilde{\mathbf{z}}} = 0$ after integrating by parts and using the equation of motion $D_I b^{IJ} = 0$. However, with an insertion of c or \tilde{c} , we actually run into a delta function contribution¹⁴ $(\sqrt{g}/2\pi) D_i b^{ij}(p') c^k(p) = \delta^{jk} \delta(p' - p)$. Evaluating $\Psi_{\mathbf{z}}$ and $\Psi_{\tilde{\mathbf{z}}}$ with the help of these delta functions, we find that we can replace $\Psi_{\mathbf{z}} \Psi_{\tilde{\mathbf{z}}} \tilde{c}V(p)$ with $v^w \tilde{v}^{\tilde{w}} V(p) = V(p)$, where in the last step we used (2.55) and (2.56). So we can replace $d\mathbf{z} d\tilde{\mathbf{z}} \Psi_{\mathbf{z}} \Psi_{\tilde{\mathbf{z}}} \tilde{c}V(p)$ by $d\mathbf{z} d\tilde{\mathbf{z}} V(p)$. (There are no such contributions for other vertex operators, since by hypothesis there are no other vertex operators inserted in the region where the metric in description (i) depends on \mathbf{z} and $\tilde{\mathbf{z}}$.)

In description (i), the integral over \mathbf{z} and $\tilde{\mathbf{z}}$ will not be very transparent because the metric of Σ depends on \mathbf{z} and $\tilde{\mathbf{z}}$. However, at this point, we can go over to description (ii), in which the metric of Σ does not depend on \mathbf{z} and $\tilde{\mathbf{z}}$. In that description, the point $w = \tilde{w} = 0$ is mapped to $(z, \tilde{z}) = (\mathbf{z}, \tilde{\mathbf{z}})$, so the vertex operator insertion is $V(\mathbf{z}, \tilde{\mathbf{z}})$ and the dependence on \mathbf{z} and $\tilde{\mathbf{z}}$ is entirely contained in the expression $d\mathbf{z} d\tilde{\mathbf{z}} V(\mathbf{z}, \tilde{\mathbf{z}})$. So the dependence of the scattering amplitude on the insertion of the vertex operator \mathcal{V} that we have been analyzing comes entirely from a factor

$$(2.60) \quad \int_{\Sigma} d\mathbf{z} d\tilde{\mathbf{z}} V(\mathbf{z}, \tilde{\mathbf{z}}).$$

We have to include this factor in addition to whatever other vertex operator insertions there may be, and in addition to whatever other moduli Σ may have. We have arrived at the description of the S -matrix in terms of an integrated vertex operator.

In this derivation, we chose local coordinates and considered only a small region $|\mathbf{z}| < \epsilon$. To justify the statement that the integral in (2.60) should extend over all of Σ , we simply cover Σ with small open sets in each of which we can proceed in the fashion just described, using convenient local

¹⁴The statement that there is such a delta function is a variant of the explanation in eqn. (2.17).

parameters. (Of course, there is a potential problem when the vertex operator \mathcal{V} meets another vertex operator. A sufficient criterion for avoiding trouble was stated at the end of section 2.5.1.)

Obviously, in this derivation we assumed that the position at which \mathcal{V} is inserted on Σ is one of the moduli of the problem. There are a few exceptional cases in which this is not true – the three-point function in genus 0, and the one-point function in genus 1. Those examples are simple enough to be treated by hand and lead to natural formulas involving correlation functions of the integrated vertex operators, even though the above derivation does not apply. For example, for the three-point function in genus 0, there are no moduli at all; one simply has to evaluate the three-point function $\langle \tilde{c}cV_1(z_1, \tilde{z}_1)\tilde{c}cV_2(z_2, \tilde{z}_2)\tilde{c}cV_3(z_3, \tilde{z}_3) \rangle$, which factors as $\langle \tilde{c}c(z_1, \tilde{z}_1)\tilde{c}c(z_2, \tilde{z}_2)\tilde{c}c(z_3, \tilde{z}_3) \rangle \langle V_1(z_1, \tilde{z}_1)V_2(z_2, \tilde{z}_2)V_3(z_3, \tilde{z}_3) \rangle$. So the result can be conveniently expressed in terms of the three-point functions of the “integrated” vertex operators V_i , even though in this example those operators are not integrated.

2.5.3. More on gauge invariance Let us reconsider gauge invariance – that is, the decoupling of BRST-trivial states – from the point of view of integrated vertex operators. The integrated vertex operator V is a matter primary field. We expect it to decouple if V is a null vector.

The cases of massless and massive particles are rather different, as we will see momentarily. A null vector that represents a pure gauge mode of a massless particle is necessarily of the form¹⁵ $V = L_{-1}\Phi_{1,0}$ or $V = \tilde{L}_{-1}\Phi_{0,1}$. We have $L_{-1}\Phi_{1,0} = \partial_z\Phi_{1,0}$, $\tilde{L}_{-1}\Phi_{0,1} = \partial_{\tilde{z}}\Phi_{0,1}$, leading to $\int_{\Sigma} V = \int_{\Sigma} \partial_z\Phi_{1,0}$ or $\int_{\Sigma} V = \int_{\Sigma} \partial_{\tilde{z}}\Phi_{0,1}$. In each case, assuming for the moment that there are no surface terms, $\int_{\Sigma} V = 0$ vanishes since V is a total derivative, or equivalently an exact form.

For massive particles, the story is less simple. At a massive level, a null vector can be, for example, of the form $V = (L_{-2} + (3/2)L_{-1}^2)\Phi_{1,-1}$. This is not a total derivative on the Riemann surface Σ , so it is not true that $\int_{\Sigma} V$ vanishes. Rather, the null vector decouples upon performing the full integral over the moduli space $\mathcal{M}_{g,n}$ of Riemann surfaces of genus g with n punctures. However, it is clumsy to make this argument in terms of integrated vertex operators. It is much more effective to go back to the unintegrated vertex operator $\mathcal{V} = \tilde{c}cV$, which is of the form $\{Q_B, \mathcal{W}\}$ if V is a null vector. Then as in section 2.4.3, the BRST machinery shows that the coupling of $\{Q_B, \mathcal{W}\}$

¹⁵ $\Phi_{n,m}$ will denote a matter primary field of antiholomorphic and holomorphic conformal dimensions n and m . The simplest null vectors are treated in appendix B, with the antiholomorphic modes suppressed.

involves an exact form on $\mathcal{M}_{g,n}$, and the decoupling of $\{Q_B, \mathcal{W}\}$ follows by integration by parts on $\mathcal{M}_{g,n}$.

In each of these cases, one needs to analyze possible surface terms at infinity in the integral over Σ or over $\mathcal{M}_{g,n}$. In closed-string theory,¹⁶ in the integral over Σ , by “infinity” one means points at which V meets another vertex operator. A simple criterion for ensuring that there can be no difficulty is that the external momenta are sufficiently generic in a sense stated in the last paragraph of section 2.5.1. This case suffices for computing the S -matrix, so for massless modes of the closed bosonic string, there are no anomalies in BRST-symmetry. Another way to reach this conclusion is to observe that a collision of two vertex operators can already occur in genus zero, so whatever effects it produces are already included in the definition of the tree-level BRST symmetry. A related statement from a spacetime point of view is that in closed-string theories, loop effects do not trigger spontaneous breaking of gauge symmetries that can be proved by integration by parts on Σ . We elaborate on this in sections 7 and 8.

Matters are different for massive modes. Gauge-invariance for massive modes has to be proved by integration by parts on $\mathcal{M}_{g,n}$. In that context, “infinity” is associated to all possible degenerations of Σ that occur in the Deligne-Mumford compactification of $\mathcal{M}_{g,n}$ (see for example section 6 of [16] for an introduction). In this case, a full treatment involves a variety of issues that we will attempt to elucidate in section 7.

The distinction between the two cases of massless and massive null vectors can be phrased as follows. There is a forgetful map $\pi : \mathcal{M}_{g,n} \rightarrow \mathcal{M}_{g,n-1}$ that forgets one of the punctures. This map is a fibration; the fiber is a copy of Σ , parametrizing the position of the puncture that one is forgetting:

$$(2.61) \quad \begin{array}{ccc} \Sigma & \rightarrow & \mathcal{M}_{g,n} \\ & & \downarrow \pi \\ & & \mathcal{M}_{g,n-1}. \end{array}$$

The use of the integrated vertex operator amounts to performing an integral over $\mathcal{M}_{g,n}$ by integrating first over the fibers of this fibration. The integral

¹⁶The restriction to closed-string theories is important here because if Σ has a boundary, then integration by parts on Σ can certainly produce a boundary term, giving a simple mechanism for gauge symmetry breaking (see section 7.4). This is not a surface term at infinity in the everyday sense, though it has that interpretation in the Deligne-Mumford compactification (see section 9.2.3). The argument given in the text does apply to open-string gauge invariances in theories of open and closed strings, since the boundary of a string worldsheet itself has no boundary.

over $\mathcal{M}_{g,n}$ that represents the coupling of a massless null vector vanishes upon integration over the fibers of the fibration (2.61), because in this case V is a total derivative on Σ . By contrast, integration over the fibers of π does not help much in understanding the coupling of a massive null vector; in that case, the only simplicity arises upon performing the full integral over $\mathcal{M}_{g,n}$.

What we have just described is a close analog of a phenomenon that occurs in superstring theory and accounts for some of the subtlety of that subject. In superstring theory, Σ becomes a super Riemann surface. As long as one considers only Neveu-Schwarz vertex operators, there is a fibration like (2.61). Integration over the fibers of this fibration suffices to establish the decoupling of massless null vectors in the NS sector. For Ramond vertex operators, however, there is no fibration analogous to (2.61); this is because a Ramond vertex operator is inserted at a singularity of the superconformal structure of Σ , and it does not make sense to move such a vertex operator while keeping fixed the moduli of Σ . Accordingly, in the Ramond sector, even for massless states – gravitinos – gauge-invariance can only be established by a full integral over the supermoduli space, not just an integral over the worldsheet. This fact was one source of difficulty in the literature of the 1980s.

An important point is that as long as one is only forgetting one puncture, the fibration (2.61) extends over the Deligne-Mumford compactifications of $\mathcal{M}_{g,n}$ and $\mathcal{M}_{g,n-1}$. For an analogous fibration that forgets two or more punctures, this would not be true. A locus in which three or more punctures collide needs to be blown up to get the Deligne-Mumford compactification (this was depicted in fig. 1). The three colliding vertex operators can be, for example, two that are expressed in integrated form and a third that is either integrated or unintegrated. Accordingly, the use of integrated vertex operators leads to difficulty in some calculations, as already remarked in section 2.5.1.

2.6. Some variations on the theme

Since an odd variable is its own delta function, instead of writing the vertex operator as $\mathcal{V} = \tilde{c}cV$, we can write

$$(2.62) \quad \mathcal{V} = \delta(\tilde{c})\delta(c)V.$$

This will be a good starting point for understanding superstring vertex operators.

The delta functions in (2.62) have an intuitive explanation. The ghost field c^i is a vector field on Σ – a generator of the diffeomorphism group of Σ . In the presence of an unintegrated vertex operator insertion at a point $p \in \Sigma$,

we want to consider as symmetries only those diffeomorphisms that leave p fixed. They are generated by vector fields that vanish at p . So we want to set the ghost fields \tilde{c}, c to zero at p . This is accomplished by the delta functions.

Another useful variation on what we have explained is as follows. A metric g on an oriented two-manifold Σ determines a complex structure J . Instead of using the path integral to define a differential form $F(g|\delta g)$ on the space \mathcal{J} of all metrics on Σ , we could have used it to define a differential form $F(J|\delta J)$ on the space \mathcal{Z} of all complex structures on Σ . In fact, this has several advantages, including manifest conformal invariance.

If we work with variations of complex structures rather than metrics, then eqn. (2.3) for the variation of the action becomes, in local complex coordinates,

$$(2.63) \quad \delta I = \frac{-i}{4\pi} \int_{\Sigma} d\tilde{z} dz \left(\delta J_z^z T_{zz} - \delta J_z^{\tilde{z}} T_{\tilde{z}\tilde{z}} \right),$$

and similarly eqn. (2.5) for the extended action becomes

$$(2.64) \quad \hat{I} = I + \frac{-i}{4\pi} \int_{\Sigma} d\tilde{z} dz \left(\delta J_z^z b_{zz} - \delta J_z^{\tilde{z}} b_{\tilde{z}\tilde{z}} \right).$$

From this starting point, the analysis proceeds in an obvious way. The complex structure of $\mathcal{M}_{g,n}$ is more evident in this description. δJ_z^z is a form of type $(1,0)$ on the space of complex structures and $\delta J_z^{\tilde{z}}$ is a form of type $(0,1)$. After dividing by diffeomorphisms (which act holomorphically on the space of complex structures), δJ_z^z and $\delta J_z^{\tilde{z}}$ descend, respectively, to $(1,0)$ - and $(0,1)$ -forms on $\mathcal{M}_{g,n}$.

For bosonic strings, it does not much matter if one works with metrics or with complex structures. In the generalization to superstrings, the description in terms of complex structures is more efficient.

3. A measure on supermoduli space

In this section, we will adapt the derivation of section 2 to superstring theory. For this, one simply replaces Riemann surfaces by super Riemann surfaces, and extends the stress tensors, ghosts, and antighosts of section 2 to supermultiplets. All of the bosonic formulas have natural superanalogs as long as one has the courage to write them down.

To appreciate the construction of the superstring measure, one has to become comfortable with interpreting the integral over certain even variables (the commuting ghosts and the differentials of odd moduli) as an algebraic

operation – a slightly generalized Gaussian integral – rather than interpreting it literally as an integral. This is natural in the general theory of integration over supermanifolds; see for example section 3.3.2 of [15]. One also has to become comfortable with certain delta function operators (aspects of which were introduced in [7] and [5]) that may be slightly off-putting at first sight, but that on further reflection make perfect sense. In the present section, we make only the minimum necessary remarks on these matters, deferring more detail on the commuting ghost system to section 10.

One difference between superstrings and bosonic strings is that technically in the superstring case, it seems simpler to express the construction in terms of complex structures (as in section 2.6) rather than in terms of metrics (as in the rest of section 2).

For brevity, we concentrate here on the heterotic string, in which only the right-moving modes have superconformal symmetry. The generalization to superstring constructions in which both left- and right-movers have superconformal symmetry should be clear and occasionally we make a few remarks.

3.1. A short account of super Riemann surfaces

We begin with a bare minimum background from super Riemann surface theory. Many topics are treated much more fully in [16].

3.1.1. Worldsheets The worldsheet Σ of a heterotic string has a local antiholomorphic coordinate \tilde{z} and local holomorphic coordinates $z|\theta$. From a holomorphic point of view, Σ is endowed with a superconformal structure, which in suitable coordinates – called superconformal coordinates – is determined by the operator

$$(3.1) \quad D_\theta = \frac{\partial}{\partial\theta} + \theta \frac{\partial}{\partial z}.$$

We are really not interested in D_θ but in the line bundle \mathcal{D} whose holomorphic sections are of the form $f(z|\theta)D_\theta$; this is a holomorphic subbundle of what we will call $T_R\Sigma$, the holomorphic tangent bundle of Σ . $T_R\Sigma$ is generated by ∂_z and ∂_θ . Similarly, we write $T_L\Sigma$ for the antiholomorphic tangent bundle, which is generated by $\partial_{\tilde{z}}$; $T_L^*\Sigma$ for the antiholomorphic cotangent bundle, generated by $d\tilde{z}$; and $T_R^*\Sigma$ for the holomorphic cotangent bundle, generated by dz and $d\theta$. We also call $T_L^*\Sigma$ and $T_R^*\Sigma$ the spaces of $(0, 1)$ -forms and $(1, 0)$ -forms on Σ .

The intuition that holomorphic and antiholomorphic variables on Σ are independent of each other can be captured by thinking of Σ as a smooth

submanifold of a complex supermanifold $\Sigma_L \times \Sigma_R$, where Σ_L is an ordinary Riemann surface, Σ_R is a super Riemann surface, and Σ_L is very close to the complex conjugate of the reduced space $\Sigma_{R,\text{red}}$ of Σ_R . (See section 5 of [15].) From that point of view, holomorphic and antiholomorphic functions on Σ are restrictions to Σ of holomorphic functions on Σ_R and Σ_L ; similarly $T_R\Sigma$ and $T_L\Sigma$ are restrictions to Σ of the holomorphic tangent bundles $T\Sigma_R$ and $T\Sigma_L$, respectively. This gives the most precise approach to a heterotic string worldsheet and we will adopt this point of view in the present paper when needed.

The simplest type of holomorphic field on Σ is a field $\Phi^{[n]}$ that is a section of \mathcal{D}^n . In superconformal coordinates, such a field has an expansion

$$(3.2) \quad \Phi^{[n]} = u + \theta v,$$

where u has conformal dimension $-n/2$ and v has conformal dimension $-n/2 + 1/2$.

A coordinate transformation is called superconformal if it multiplies D_θ by a scalar function (and hence preserves the line bundle \mathcal{D} generated by D_θ). Superconformal coordinates are only unique up to a superconformal transformation. A vector field generates a superconformal transformation – and we call it a superconformal vector field – if its commutator with D_θ is a ($z|\theta$ -dependent) multiple of D_θ . Odd and even superconformal vector fields take the form

$$(3.3) \quad \begin{aligned} \nu_f &= f(z) (\partial_\theta - \theta \partial_z) \\ V_g &= g(z) \partial_z + \frac{g'(z)}{2} \theta \partial_\theta, \end{aligned}$$

with holomorphic functions $f(z)$, $g(z)$. These functions can be combined to a superfield $\mathcal{V}^{[2]}(z|\theta) = g(z) + 2\theta f(z)$, which is a section of \mathcal{D}^2 .

The following are a basis of superconformal vector fields that are regular except possibly for a pole at $z = 0$:

$$(3.4) \quad \begin{aligned} G_r &= z^{r+1/2} (\partial_\theta - \theta \partial_z), & r \in \mathbb{Z} + 1/2 \\ L_n &= -z^{n+1} \partial_z - \frac{1}{2} (n+1) z^n \theta \partial_\theta, & n \in \mathbb{Z}. \end{aligned}$$

The pole is absent if

$$(3.5) \quad r \geq -1/2, \quad n \geq -1.$$

A short calculation shows that the vector fields (3.4) obey the super Virasoro algebra (with zero central charge) in the Neveu-Schwarz (NS) sector:

$$(3.6) \quad \begin{aligned} [L_m, L_n] &= (m - n)L_{m+n} \\ \{G_r, G_s\} &= 2L_{r+s} \\ [L_m, G_r] &= \left(\frac{m}{2} - r\right)G_{m+r}. \end{aligned}$$

If $z|\theta$ are local superconformal coordinates and \tilde{z} is an antiholomorphic coordinate that is sufficiently close to the complex conjugate of z , we call $\tilde{z};z|\theta$ a system of standard local coordinates. (We do not ask for \tilde{z} to equal the complex conjugate of z as this condition is not invariant under superconformal transformations. See section 5 of [15] for more discussion.)

We will use the term supercomplex structure to refer to the complex structure of Σ plus its holomorphic superconformal structure (the choice of the line bundle \mathcal{D} generated by D_θ). Just as on an ordinary complex manifold, the complex structure can be defined by a linear transformation \mathcal{J} of the cotangent bundle of Σ that obeys $\mathcal{J}(dz) = idz$, $\mathcal{J}(d\theta) = id\theta$, $\mathcal{J}(d\tilde{z}) = -id\tilde{z}$. These conditions can be described by saying that dz and $d\theta$ are one-forms of type $(1, 0)$, and furnish a basis of the holomorphic cotangent bundle $T_R^*\Sigma$, while $d\tilde{z}$ is of type $(0, 1)$, and furnishes a basis of the antiholomorphic cotangent bundle $T_L^*\Sigma$. \mathcal{J} obeys $\mathcal{J}^2 = -1$, just as on an ordinary complex manifold, and also is compatible with the existence of a holomorphic superconformal structure, as described above. The space of all such \mathcal{J} 's is an infinite-dimensional complex supermanifold \mathcal{J} .

3.1.2. Deformations To construct superstring perturbation theory, we have to study deformations of Σ . *A priori*, we have to consider both deformations of the complex structure \mathcal{J} of Σ and deformations of its superconformal structure, that is deformations of the embedding of \mathcal{D} in $T_R\Sigma$. We will see shortly that there are no nontrivial deformations of the superconformal structure without changing the complex structure, so we really only need to consider deformations of \mathcal{J} . Of course, we only allow deformations of \mathcal{J} that preserve the existence of a holomorphic superconformal structure. For our purposes in this paper, what we need to know can be summarized as follows (see for example sections 3.5.3-4 of [16]). The allowed variations of \mathcal{J} are determined¹⁷ by the components $\delta\mathcal{J}_z^z$ and $\delta\mathcal{J}_\theta^{\tilde{z}}$ which respectively

¹⁷ The constraint $\mathcal{J}^2 = -1$ implies the vanishing of $\delta\mathcal{J}_\theta^\theta$, $\delta\mathcal{J}_z^\theta$, $\delta\mathcal{J}_\theta^z$, $\delta\mathcal{J}_z^z$, and $\delta\mathcal{J}_z^{\tilde{z}}$. The superconformal structure determines $\delta\mathcal{J}_z^\theta$ in terms of $\delta\mathcal{J}_z^z$ and $\delta\mathcal{J}_z^{\tilde{z}}$ in terms of $\delta\mathcal{J}_\theta^{\tilde{z}}$.

represent deformations of the holomorphic and antiholomorphic structure of Σ . Their θ expansions read

$$(3.7) \quad \begin{aligned} \delta \mathcal{J}_z^z &= h_z^z + \theta \chi_z^\theta \\ \delta \mathcal{J}_\theta^z &= e_\theta^z + \theta h_\theta^z. \end{aligned}$$

In the first equation, in conventional language, h_z^z is a metric perturbation and χ_z^θ is the gravitino field. In the second equation, h_θ^z is a metric perturbation while e_θ^z can be set to zero as a gauge condition (using the θ -dependent part of q^z in eqn. (3.10) below). Geometrically, $\delta \mathcal{J}_z^z$ is a $(0, 1)$ -form on Σ with values in \mathcal{D}^2 , and $\delta \mathcal{J}_\theta^z$ is a section of \mathcal{D}^{-1} with values in $T_L \Sigma$. For any point $p \in \Sigma$, we can interpret¹⁸ $\delta \mathcal{J}_z^z(p)$ as a $(1, 0)$ -form on \mathcal{J} , the space of supercomplex structures, and $\delta \mathcal{J}_\theta^z(p)$ as a $(0, 1)$ -form on \mathcal{J} .

Now we have to take account of diffeomorphisms. Deformations of Σ that are generated by a vector field on Σ – the generator of an infinitesimal diffeomorphism of Σ – are uninteresting in a diffeomorphism-invariant worldsheet theory. A general vector field takes the form

$$(3.8) \quad q^z \partial_z + \left(q^z \partial_z + \frac{1}{2} D_\theta q^z D_\theta \right) + q^\theta D_\theta.$$

(The advantage of writing the expansion this way is explained in [16].) The $q^\theta D_\theta$ term is used to eliminate deformations of Σ in which the embedding of \mathcal{D} in $T_R \Sigma$ is changed. This works as follows. The change in D_θ generated by $q^\theta D_\theta$ is

$$(3.9) \quad \delta D_\theta = [D_\theta, q^\theta D_\theta] = (D_\theta q^\theta) D_\theta - 2q^\theta \partial_z.$$

On the right hand side, the first term proportional to D_θ does not change the line bundle \mathcal{D} generated by D_θ , but the second term $-2q^\theta \partial_z$, since q^θ is an arbitrary function, represents an arbitrary change in the embedding of \mathcal{D} in $T_R \Sigma$. ($T_R \Sigma$ is generated by D_θ and ∂_z , while \mathcal{D} is generated by D_θ , so to change the embedding of \mathcal{D} in $T_R \Sigma$, one must shift D_θ by a multiple of ∂_z .) Thus, we need only consider deformations of the complex structure \mathcal{J} of Σ .

¹⁸This is precisely analogous to what happens on an ordinary Riemann surface with complex structure J : the variation δJ_z^z is a $(1, 0)$ -form on the space of complex structures, while δJ_θ^z is a $(0, 1)$ -form.

Given this, the nontrivial deformations of Σ are deformations of \mathcal{J} modulo deformations generated by $q^{\tilde{z}}$ and q^z . Such deformations take the form

$$(3.10) \quad \begin{aligned} \delta \mathcal{J}_{\tilde{z}}^z &= \partial_{\tilde{z}} q^z \\ \delta \mathcal{J}_{\theta}^{\tilde{z}} &= D_{\theta} q^{\tilde{z}}. \end{aligned}$$

The moduli of Σ over which one ultimately has to integrate in order to compute scattering amplitudes correspond to deformations of \mathcal{J} modulo these trivial ones.

3.1.3. Action and ghosts Now consider a superconformal field theory on Σ with action I . In superstring theory, I is the sum of a matter action that we call I_X , where X refers generically to all matter fields on the worldsheet, and a ghost action I_{gh} . When we vary the supercomplex structure of Σ , I changes by an amount proportional to the supercurrent and stress tensor of Σ . In heterotic string theory, the formula reads

$$(3.11) \quad \delta I = \frac{1}{2\pi} \int_{\Sigma} \mathcal{D}(\tilde{z}, z|\theta) \left(\delta \mathcal{J}_{\tilde{z}}^z \mathcal{S}_{z\theta} + \delta \mathcal{J}_{\theta}^{\tilde{z}} T_{\tilde{z}\tilde{z}} \right).$$

This is the analog of the bosonic formula eqn. (2.3). As in section 3.3 of [16], we make the convenient abbreviation

$$(3.12) \quad \mathcal{D}(\tilde{z}, z|\theta) = -i[d\tilde{z}; dz|d\theta],$$

to get more natural-looking formulas.¹⁹ In eqn. (3.11), $\mathcal{S}_{z\theta}$ is the holomorphic superfield that contains the holomorphic supercurrent and stress tensor,

$$(3.13) \quad \mathcal{S}_{z\theta} = S_{z\theta} + \theta T_{zz},$$

and $T_{\tilde{z}\tilde{z}}$ is the antiholomorphic stress tensor. These fields obey

$$(3.14) \quad \partial_{\tilde{z}} \mathcal{S}_{z\theta} = 0 = D_{\theta} T_{\tilde{z}\tilde{z}} = \partial_z T_{\tilde{z}\tilde{z}},$$

ensuring that δI vanishes if $\delta \mathcal{J}$ is generated by an infinitesimal diffeomorphism (in other words if $\delta \mathcal{J}$ has the form of eqn. (3.10)). Geometrically, $S_{z\theta}$ is a section of \mathcal{D}^{-3} and $T_{\tilde{z}\tilde{z}}$ is a section of $(T_L^* \Sigma)^2$. These facts ensure that

¹⁹The analog on an ordinary Riemann surface is as follows. If $z = x + iy$, then $d\bar{z} \wedge dz = 2i dx \wedge dy$ is imaginary, so it is sometimes convenient to introduce the real two-form $d^2z = -id\bar{z} \wedge dz$, which is a bosonic analog of $\mathcal{D}(\tilde{z}, z|\theta)$.

the integrand in (3.11) is a $(0, 1)$ -form on Σ with values in \mathcal{D}^{-1} , so that the integral does not depend on the choice of coordinates.

Acting on possible operator insertions at $z = 0$, $\mathcal{S}_{z\theta}$ has a mode expansion

$$(3.15) \quad \mathcal{S}_{z\theta}(z|\theta) = \frac{1}{2} \sum_{r \in \mathbb{Z} + 1/2} z^{-r-3/2} G_r + \theta \sum_{n \in \mathbb{Z}} z^{-n-2} L_n.$$

The coefficients are quantum operators corresponding to the superconformal vector fields in eqn. (3.4). Similarly, the modes of T_{zz}^{\sim} are the left-moving Virasoro generators:

$$(3.16) \quad T_{zz}^{\sim} = \sum_{n \in \mathbb{Z}} \tilde{z}^{-n-2} \tilde{L}_n.$$

The last characters that we have to introduce are the supersymmetric ghosts and antighosts, or superghosts for short. The holomorphic ghosts are a superfield²⁰ C^z of $N_{\text{gh}} = 1$. C^z is an odd section of \mathcal{D}^2 and its theta expansion reads

$$(3.17) \quad C^z = c^z + \theta \gamma^\theta$$

where c^z and γ^θ have conformal dimensions -1 and $-1/2$, respectively. Similarly, the holomorphic antighosts are a superfield $B_{z\theta}$ of $N_{\text{gh}} = -1$. B is a section of \mathcal{D}^{-3} , with theta expansion

$$(3.18) \quad B_{z\theta} = \beta_{z\theta} + \theta b_{zz},$$

where β and b have conformal dimensions $3/2$ and 2 , respectively. The fields b_{zz} , c^z are the anticommuting Virasoro ghosts that are familiar from the bosonic string, while $\beta_{z\theta}$, γ^θ are commuting ghost fields that are associated to the odd generators of the super-Virasoro algebra. The $\beta\gamma$ system has unusual properties that have long played a central role in quantization of superstrings [1]. The antiholomorphic ghosts and antighosts are anticommuting fields \tilde{C}^z of $N_{\text{gh}} = 1$ and \tilde{B}_{zz}^{\sim} of $N_{\text{gh}} = -1$ that are sections respectively of $T_L \Sigma$ and $(T_L^* \Sigma)^2$. Their theta expansions read

$$(3.19) \quad \tilde{B}_{zz}^{\sim} = \tilde{b}_{zz}^{\sim} + \theta \tilde{f}_{zz\theta}^{\sim}, \quad \tilde{C}^z = \tilde{c}^z + \theta \tilde{g}_\theta^z,$$

²⁰Though we will display them for the moment, the traditional subscripts and superscripts in C^z and $B_{z\theta}$ are natural only when the super Riemann surface Σ is split. In general, it is best to simply think of C and B as sections of \mathcal{D}^2 and \mathcal{D}^{-3} , respectively. The same goes for related objects such as the supercurrent \mathcal{S} .

where \tilde{c} and \tilde{b} play the role of antiholomorphic Virasoro ghosts that are familiar in the bosonic string, while \tilde{f} and \tilde{g} are auxiliary fields that vanish on-shell. The ghost and antighost fields are governed by the action

$$(3.20) \quad I_{\text{gh}} = \frac{1}{2\pi} \int \mathcal{D}(\tilde{z}, z|\theta) \left(B_{z\theta} \partial_{\tilde{z}} C^z + \tilde{B}_{\tilde{z}\tilde{z}} D_{\theta} \tilde{C}^{\tilde{z}} \right),$$

and in particular the equations of motion for the antighosts read

$$(3.21) \quad \partial_{\tilde{z}} B_{z\theta} = 0 = D_{\theta} \tilde{B}_{\tilde{z}\tilde{z}}.$$

The subscripts and superscripts carried by the fields are intended as reminders of how they transform under reparametrizations; we often omit them to reduce clutter.

For the present paper, one of the most important properties of the ghost and antighost fields is the BRST transformation law of the antighosts:

$$(3.22) \quad \begin{aligned} [Q_B, B_{z\theta}] &= \mathcal{S}_{z\theta} \\ \{Q_B, \tilde{B}_{\tilde{z}\tilde{z}}\} &= T_{\tilde{z}\tilde{z}}. \end{aligned}$$

3.2. The extended action and the integration measure

The matter plus ghost action I of the heterotic string is BRST-invariant if \mathcal{J} is regarded as a fixed quantity. Suppose, however, that we let the BRST charge Q_B act on \mathcal{J} , producing variations $\delta\mathcal{J}_z^z$ and $\delta\mathcal{J}_{\tilde{\theta}}^{\tilde{z}}$. We understand these as fields of $N_{\text{gh}} = 1$; $\delta\mathcal{J}_z^z$ is odd and $\delta\mathcal{J}_{\tilde{\theta}}^{\tilde{z}}$ is even. To preserve $Q_B^2 = 0$, we take

$$(3.23) \quad \{Q_B, \delta\mathcal{J}_z^z\} = [Q_B, \delta\mathcal{J}_{\tilde{\theta}}^{\tilde{z}}] = 0.$$

If we extend the action of Q_B in this way, the action is no longer Q_B -invariant. But from our study of the bosonic string, we know what to do. We restore the Q_B -invariance by extending the action, adding a coupling of $\delta\mathcal{J}$ to the antighosts:

$$(3.24) \quad I \rightarrow \hat{I} = I + \frac{1}{2\pi} \int_{\Sigma} \mathcal{D}(\tilde{z}, z|\theta) \left(\delta\mathcal{J}_z^z B_{z\theta} - \delta\mathcal{J}_{\tilde{\theta}}^{\tilde{z}} \tilde{B}_{\tilde{z}\tilde{z}} \right).$$

The extended action \hat{I} is Q_B -invariant.

From here we proceed as in our study of the bosonic string. We integrate out all other variables – both matter fields X and superghosts – to define a function of \mathcal{J} and $\delta\mathcal{J}$ only:

$$(3.25) \quad F(\mathcal{J}, \delta\mathcal{J}) = \int \mathcal{D}(X, B, C, \tilde{B}, \tilde{C}) \exp(-\hat{I}).$$

Integrating over the BRST multiplets $X, B, C, \tilde{B}, \tilde{C}$ is a Q_B -invariant operation, so $F(\mathcal{J}, \delta\mathcal{J})$ is Q_B -invariant:

$$(3.26) \quad [Q_B, F(\mathcal{J}, \delta\mathcal{J})] = 0.$$

From here, we can, up to a certain point, reason as we did in the case of the bosonic string. Acting on functions of \mathcal{J} and $\delta\mathcal{J}$ only, Q_B can be regarded as the exterior derivative on the space \mathcal{J} of all supercomplex structures. Here, for any point $p \in \Sigma$, we interpret the matrix elements of \mathcal{J} acting on the cotangent space to Σ at p as functions on \mathcal{J} , while $\delta\mathcal{J}_z^z(p)$ and $\delta\mathcal{J}_\theta^z(p)$ are respectively $(1, 0)$ -forms and $(0, 1)$ -forms on \mathcal{J} . With this interpretation, we can understand $F(\mathcal{J}, \delta\mathcal{J})$ as a form on \mathcal{J} ; eqn. (3.26) says that this form is closed:

$$(3.27) \quad dF(\mathcal{J}, \delta\mathcal{J}) = 0.$$

3.2.1. A lightning review of integration on supermanifolds We did not call $F(\mathcal{J}, \delta\mathcal{J})$ a differential form; in fact, technically it is better called a pseudoform. The reader may wish to consult an introduction to integration theory on supermanifolds (such as [15]), but here is a very brief summary. Let M be a supermanifold with even coordinates $t^1 \dots t^m$ and odd coordinates $\theta^1 \dots \theta^n$. For each even or odd coordinate t^i or θ^j , we introduce corresponding variables dt^i or $d\theta^j$ with the opposite statistics (so dt^i is odd and $d\theta^j$ is even). We write generically x for all variables $t^1 \dots | \dots \theta^n$ and dx for the corresponding differentials. We say that a function $F(x, dx)$ has degree s if

$$(3.28) \quad F(x, \lambda dx) = \lambda^s F(x, dx).$$

We define the exterior derivative, mapping functions of degree s to functions of degree $s + 1$:

$$(3.29) \quad d = \sum_I dx^I \frac{\partial}{\partial x^I}, \quad d^2 = 0.$$

Whether or not the original set of variables $t^1 \dots | \dots \theta^n$ had a natural measure, there is always a natural measure for the extended set of variables x , dx , because of the way the variables come in pairs with opposite statistics. The integral of a function $F(x, dx)$ is defined as a Berezin integral over all variables x and dx , whenever this makes sense:

$$(3.30) \quad \int \mathcal{D}(x, dx) F(x, dx).$$

Here, however, we meet the main difference between ordinary integration theory and integration theory on supermanifolds. On an ordinary manifold with only even coordinates $t^1 \dots t^m$ and therefore only odd differentials $dt^1 \dots dt^m$, a function $F(x, dx)$ is inevitably polynomial in the differentials. But on a supermanifold, there are also odd coordinates $\theta^1 \dots \theta^n$ and therefore even differentials $d\theta^1 \dots d\theta^n$, so $F(x, dx)$ is not necessarily polynomial in the dx 's. If $F(x, dx)$ has polynomial dependence on the dx 's, we call it a differential form; otherwise we call it a pseudoform or just a form.

In fact, if $F(x, dx)$ is polynomial in the even variables $d\theta^1 \dots d\theta^n$, then the integral over those variables diverges. A typical example of a form that can be integrated over a supermanifold of dimension $m|n$ is

$$(3.31) \quad F(x, dx) = f(t^1 \dots | \dots \theta^n) dt^1 \dots dt^m \delta(d\theta^1) \dots \delta(d\theta^n).$$

The integrals over the $d\theta$'s can be done with the aid of the delta functions. Notice that $F(x, dx)$ has degree $m - n$. It also has another important "quantum number" which has no analog in integration theory on ordinary manifolds. This is the "picture number," defined as minus²¹ the number of even differentials with respect to which $F(x, dx)$ has delta function support. So the form $F(x, dx)$ defined in (3.31) has picture number $-n$. By a form of superdegree $m|n$, we mean a form of degree $m - n$ and picture number $-n$. A form of superdegree $m|n$ is also called an $m|n$ -form. An $m|n$ -form, such as $F(x, dx)$ as defined above, can be integrated over a supermanifold of dimension $m|n$.

The typical example (3.31) of a form that can be integrated may look unappealing, because the odd differentials dt^i and even differentials $d\theta^j$ are treated differently. We can write the definition in a possibly more pleasing way if we recall that an odd variable is its own delta function, so that we can replace dt^i with $\delta(dt^i)$. Thus, we can rewrite (3.31) as follows:

$$(3.32) \quad \begin{aligned} F(x, dx) &= f(t^1 \dots | \dots \theta^n) \delta(dt^1) \dots \delta(d\theta^n) \\ &= f(t^1 \dots | \dots \theta^n) \delta^{m|n}(dt^1 \dots | \dots d\theta^n). \end{aligned}$$

²¹The minus sign is included in the definition to agree with the choice in [1].

In general, we want to consider forms that have delta function localization at $d\theta = 0$ for some of the $d\theta$'s and polynomial dependence on the others. By “delta function localization,” we refer to a function with distributional support at $d\theta = 0$, either a delta function or a (possibly repeated) derivative of a delta function. For example, for a single odd variable θ , the form

$$(3.33) \quad \frac{\partial^r}{\partial(d\theta)^r} \delta(d\theta)$$

has degree $-r - 1$ and picture number -1 . So it is a form of superdegree $-r|1$. We allow derivatives of a delta function, not just delta functions, because this is necessary to get a theory in which all important operations can be defined (a key example is contraction with a vector field, discussed in appendix A, which can involve the derivative with respect to an even differential $d\theta$).

On a supermanifold M of any dimension, an $m|n$ -form can potentially be integrated on a submanifold $N \subset M$ of dimension $m|n$. We say “potentially” because just as in integration on a bosonic manifold, sometimes the integral may diverge.

We conclude this lightning introduction to integration theory on supermanifolds with a simple example on $\mathbb{R}^{1|2}$ with coordinates $t|\theta^1\theta^2$. One purpose of the example is to show that it is useful to consider objects like $\delta'(d\theta)$. Consider the form

$$(3.34) \quad F(x, dx) = \theta^2 dt \delta'(d\theta^1).$$

This form has degree -1 and picture number -1 , so it is a $0|1$ -form and can potentially be integrated on a submanifold $N \subset M$ of that dimension. We define N by $t = \alpha\theta^1$, $\theta^2 = \lambda\theta^1$, where α is an odd parameter and λ is an even one. So N is parametrized by θ^1 . Restricted to N , we have $F(x, dx) = -\alpha\lambda\theta^1 d\theta^1 \delta'(d\theta^1) = -\theta^1 \alpha \lambda \delta(d\theta^1)$ (where we used the fact that for an even variable y , $y\delta'(y) = -\delta(y)$). Performing the Berezin integral over θ^1 and $d\theta^1$, we get finally $\int_N \mathcal{D}(\theta^1, d\theta^1) F = -\alpha\lambda$.

3.2.2. Bosonic integration as an algebraic operation An unusual feature of superstring quantization is that [1] the path integral of the $\beta\gamma$ commuting ghost system is not really an integral in the usual sense. That is because the fields β, γ obey no reality condition and the action

$$(3.35) \quad I_{\beta\gamma} = \frac{1}{\pi} \int_{\Sigma} d^2z \beta \partial_{\bar{z}} \gamma$$

has no reality or positivity property. As a result, the “path integral” of the $\beta\gamma$ system has to be understood as a formal algebraic operation, somewhat like the Berezin integral for fermions.

Something similar happens in general integration theory over supermanifolds when the odd variables θ^i do not have a real structure.²² One has to postulate formulas like

$$(3.36) \quad \int \mathcal{D}(d\theta) \frac{\partial^r}{\partial(d\theta)^r} \delta(d\theta) = \delta_{r,0},$$

which have to be understood in a formal sense if $d\theta$ is a complex variable. We are in that situation here, because $\delta\mathcal{J}$ contains even as well as odd modes and these even modes have no real structure. (Both $\delta\mathcal{J}_{\tilde{\theta}}^z$ and $\delta\mathcal{J}_z^z$ have even modes, but the even modes in $\delta\mathcal{J}_{\tilde{\theta}}^z$ play little role because they couple to auxiliary fields in \tilde{B} that vanish by their equation of motion. The important even modes are the modes in $\delta\mathcal{J}_z^z$ that couple to β .) So the coupling $\int \delta\mathcal{J}B$ in the extended action \hat{I} has the same basic property as the original action $I_{\beta\gamma}$: it contains a bilinear coupling of complex bosons, with no reality or positivity property. By analogy with what we did for the bosonic string in section 2, we eventually will want to integrate over some modes of $\delta\mathcal{J}$ to construct superstring scattering amplitudes. Because of the lack of any reality or positivity property in the $\int \delta\mathcal{J}B$ coupling, this integral will pose exactly the same type of question as is posed by the original $\beta\gamma$ path integral.

In section 10, after gaining more experience with what sort of operator insertions are important in the $\beta\gamma$ path integral, we will give a more extensive discussion of such matters (see also [7, 46, 47]). For now, we just explain the minimum so that we can proceed. The troublesome integrals over complex bosons will always be Gaussian integrals or slight generalizations thereof. Let $u_i, v_j, i, j = 1, \dots, n$ be bosonic variables and let m be an $n \times n$ complex matrix that is nondegenerate but obeys no reality or positivity condition. We define the basic Gaussian integral

$$(3.37) \quad \int \mathcal{D}(u, v) \exp \left(- \sum_{i,j} u_i m_{ij} v_j \right) = \frac{1}{\det m}.$$

If m is positive definite, u_j is understood as the complex conjugate of v_j , and the integration measure $\mathcal{D}(u, v)$ is suitably normalized (with a factor of $1/2\pi i$ for each pair of variables), then this formula is actually a true theorem about

²²See for example section 3.3.2 of [16].

ordinary integration. In general, we just take it as a definition of what we mean by Gaussian “integration” for complex bosons, somewhat like the formal definition of the Berezin integral for fermions. As usual, we also extend the Gaussian integral to allow linear (*c*-number) sources *r* and *s* in the exponent: (3.38)

$$\int \mathcal{D}(\mathbf{u}, \mathbf{v}) \exp \left(- \sum_{i,j} \mathbf{u}_i m_{ij} \mathbf{v}_j - \sum_k (r_k \mathbf{u}_k + s_k \mathbf{v}_k) \right) = \frac{\exp(\sum_{i,j} s_i (m^{-1})_{ij} r_j)}{\det m}.$$

We also need the conventional formulas for the integral of a Gaussian times a polynomial, which involve Wick contractions using m^{-1} as a “propagator.” In a standard way, these formulas can be obtained by differentiating (3.38) with respect to *r* and *s* a desired number of times and then setting $r = s = 0$. And we want to allow an operation of integrating over some components of *r* and *s*. For now, we just observe that since the exponent in (3.38) is quadratic in the combined set of variables $\mathbf{u}, \mathbf{v}, r, s$, an extension of (3.38) in which we want to integrate over some components of *r* and *s* is just a Gaussian integral with more integration variables, so it is covered by the same definitions. We return to this in section 10.

If we set $r = 0$, we get a special case of (3.38):

$$(3.39) \quad \int \mathcal{D}(\mathbf{u}, \mathbf{v}) \exp \left(- \sum_{i,j} \mathbf{u}_i m_{ij} \mathbf{v}_j - \sum_k s_k \mathbf{v}_k \right) = \frac{1}{\det m},$$

independent of *s*. This formula is useful for understanding how we should define the following important integral:

$$(3.40) \quad G(\mathbf{v}) = \int \mathcal{D}(\mathbf{u}) \exp(-(\mathbf{u}, m\mathbf{v})).$$

If it is going to be possible to integrate over \mathbf{u} and \mathbf{v} by integrating first over \mathbf{u} and then over \mathbf{v} , we should have

$$(3.41) \quad \int \mathcal{D}(\mathbf{v}) A(\mathbf{v}) G(\mathbf{v}) = \int \mathcal{D}(\mathbf{u}, \mathbf{v}) A(\mathbf{v}) \exp(-(\mathbf{u}, m\mathbf{v})),$$

for any allowed function $A(\mathbf{v})$. Taking $A(\mathbf{v}) = \exp(-(\mathbf{s}, \mathbf{v}))$ and using (3.39), we see that we need

$$(3.42) \quad \int \mathcal{D}(\mathbf{v}) \exp(-(\mathbf{s}, \mathbf{v})) G(\mathbf{v}) = \frac{1}{\det m},$$

independent of s . This motivates the definition $G(\mathbf{v}) = \delta(m\mathbf{v})$, that is,

$$(3.43) \quad \int \mathcal{D}(\mathbf{u}) \exp(-(\mathbf{u}, m\mathbf{v})) = \delta(m\mathbf{v}),$$

along with

$$(3.44) \quad \int \mathcal{D}(\mathbf{v}) \delta(m\mathbf{v}) = \frac{1}{\det m}.$$

All of these formulas have analogs in the Berezin integral for fermions – the difference being that the factors of $\det m$ would appear in the numerator rather than the denominator – and like the Berezin integral, they should be understood as a convenient algebraic machinery for manipulating certain expressions.

3.2.3. The zero-mode coupling Given the formal calculus that we have just described, it is straightforward to generalize bosonic formulas such as (2.21) or (2.23) and determine the exact dependence of $F(\mathcal{J}, \delta\mathcal{J})$ on $\delta\mathcal{J}$. We assume that Σ has genus $g \geq 2$ so that the ghost fields C, \tilde{C} have no zero-modes. However, the antighost fields \tilde{B} and B do have zero-modes. In the following derivation, the holomorphic and antiholomorphic ghosts are decoupled from each other and can be treated separately. We consider first the B, C system, whose coupling to $\delta\mathcal{J}$ is given in eqn. (3.24):

$$(3.45) \quad \tilde{I}_{\delta\mathcal{J}B} = \frac{1}{2\pi} \int \mathcal{D}(\tilde{z}, z|\theta) \delta\mathcal{J}_z^z B_{z\theta}.$$

The number of zero-modes of the field $B_{z\theta}$ is $3g - 3|2g - 2$ (that is, $3g - 3$ even zero-modes and $2g - 2$ odd ones). We will denote the zero-modes as B_α , $\alpha = 1 \dots 3g - 3|1 \dots 2g - 2$; that is, the B_α include all even and odd zero-modes of B .

Rather as in (2.19), we expand B as a sum of zero-modes B_α and non-zero modes B'_λ with coefficients u_α and w_λ :

$$(3.46) \quad B = \sum_{\alpha=1 \dots | \dots 2g-2} u_\alpha B_\alpha + \sum_\lambda w_\lambda B'_\lambda.$$

C has an analogous expansion

$$(3.47) \quad C = \sum_\lambda \gamma_\lambda C_\lambda,$$

now with only non-zero modes C_λ with coefficients γ_λ . The ghost action (3.20) pairs the w_λ and the γ_λ by a nondegenerate bilinear pairing $\sum_\lambda m_\lambda w_\lambda \gamma_\lambda$, $m_\lambda \neq 0$. When we integrate over γ_λ , we get delta functions setting the w_λ to zero. This results from the formula

$$(3.48) \quad \int d\gamma \exp(-mw\gamma) = \delta(mw),$$

If w and γ are odd variables, this formula is a consequence of the Berezin integral (and was used in our study of bosonic strings), while if they are even variables, the formula has been explained in eqn. (3.43). Once we set the w_λ to zero, the coupling of B to $\delta\mathcal{J}$ reduces to

$$(3.49) \quad \tilde{I}_{\delta\mathcal{J}B} = \sum_{\alpha=1\dots|2g-2} u_\alpha \Psi_\alpha,$$

with

$$(3.50) \quad \Psi_\alpha = \frac{1}{2\pi} \int_\Sigma \mathcal{D}(\tilde{z}, z|\theta) \delta\mathcal{J}_z^z B_{z\theta\alpha}.$$

The integral over the zero-mode coefficients u_α therefore becomes

$$(3.51) \quad \prod_{\alpha=1\dots|2g-2} \int du_\alpha \exp(u_\alpha \Psi_\alpha) = \prod_{\alpha=1\dots|2g-2} \delta(\Psi_\alpha) = \delta^{3g-3|2g-2}(\Psi'_1 \dots | \dots \Psi''_{2g-2}).$$

(In the last version of the formula, we write Ψ_α as Ψ'_α or Ψ''_α according to whether it is even or odd.)

We can do a similar calculation for the antiholomorphic ghosts \tilde{B} and their coupling

$$(3.52) \quad \tilde{I}_{\delta\mathcal{J}\tilde{B}} = \frac{1}{2\pi} \int_\Sigma \mathcal{D}(\tilde{z}, z|\theta) \delta\mathcal{J}_\theta^{\tilde{z}} \tilde{B}_{z\tilde{z}}$$

to $\delta\mathcal{J}$. The number of zero-modes of $\tilde{B}_{z\tilde{z}}$ is $3g - 3$. We write $\tilde{B} = \sum_{\alpha=1\dots 3g-3} \tilde{u}_\alpha \tilde{B}_\alpha + \sum_\lambda \tilde{w}_\lambda \tilde{B}'_\lambda$ where \tilde{B}_α are zero-modes and \tilde{B}'_λ are non-zero modes. The integral over \tilde{C} sets the coefficients \tilde{w}_λ of non-zero modes to zero, whereupon we get

$$(3.53) \quad \tilde{I}_{\delta\mathcal{J}\tilde{B}} = \sum_{\alpha=1}^{3g-3} \tilde{u}_\alpha \tilde{\Psi}_\alpha,$$

with

$$(3.54) \quad \tilde{\Psi}_\alpha = \frac{1}{2\pi} \int_\Sigma [d\tilde{z}; dz | d\theta] \delta \mathcal{J}_\theta^z \tilde{\mathbb{B}}_{z\tilde{z}\alpha}.$$

The integral over the \tilde{u}_α gives

$$(3.55) \quad \prod_{\alpha=1}^{3g-3} \int d\tilde{u}_\alpha \exp(\tilde{u}_\alpha \tilde{\Psi}_\alpha) = \delta^{3g-3}(\tilde{\Psi}_1, \dots, \tilde{\Psi}_{3g-3}).$$

We have determined the exact dependence of $F(\mathcal{J}, \delta\mathcal{J})$ on $\delta\mathcal{J}$: it is the product of a function that depends only on \mathcal{J} and not on $\delta\mathcal{J}$ times

$$(3.56) \quad \delta^{3g-3}(\tilde{\Psi}_1 \dots \tilde{\Psi}_{3g-3}) \delta^{3g-3|2g-2}(\Psi'_1 \dots | \dots \Psi''_{2g-2}).$$

This formula (cousins of which can be found in [5, 9]) has an obvious resemblance to the formula (3.32) giving the prototype of a form of degree $m|n$. Clearly, we have learned that $F(\mathcal{J}, \delta\mathcal{J})$ is a form on \mathcal{J} of degree $6g - 6|2g - 2$. More specifically, this form has holomorphic degree $3g - 3|2g - 2$ and antiholomorphic degree $3g - 3|0$. It is hopefully clear that a similar derivation for Type II superstrings would proceed in essentially the same way and give a form whose holomorphic and antiholomorphic degree would both be $3g - 3|2g - 2$.

3.3. Where to integrate

3.3.1. Reduction to supermoduli space The superdegree of $F(\mathcal{J}, \delta\mathcal{J})$ is appropriate for a form that would be integrated over a supermanifold of dimension $6g - 6|2g - 2$. The reader who has gotten this far will undoubtedly anticipate that the supermanifold in question will be, in some sense, the moduli space of super Riemann surfaces.

The basic idea for reducing from \mathcal{J} to the moduli space of super Riemann surfaces is the same as it was in section 2.2. We introduce the group \mathcal{D} of superdiffeomorphisms of Σ . We want to show that $F(\mathcal{J}, \delta\mathcal{J})$ is a pullback from, roughly speaking, the quotient \mathcal{J}/\mathcal{D} . For this, we need the same two facts as before: (1) $F(\mathcal{J}, \delta\mathcal{J})$ should be \mathcal{D} -invariant; (2) $F(\mathcal{J}, \delta\mathcal{J})$ should vanish if contracted with one of the vector fields that generates the action of \mathcal{D} .

Property (1) is manifest from the definition of F , and property (2) is true by virtue of the same reasoning as in section 2.2. In this case, we have to

show that $F(\mathcal{J}, \delta\mathcal{J})$ is invariant if we shift $\delta\mathcal{J}$ in the fashion analogous to (2.13), or in other words, in view of eqn. (3.10), by

$$(3.57) \quad \delta\mathcal{J}_z^z \rightarrow \delta\mathcal{J}_z^z + \partial_z q^z, \quad \delta\mathcal{J}_\theta^{\tilde{z}} \rightarrow \delta\mathcal{J}_\theta^{\tilde{z}} + D_\theta q^{\tilde{z}}.$$

To show that $F(\mathcal{J}, \delta\mathcal{J})$ has this symmetry, we simply observe that the extended action \hat{I} of eqn. (3.24) has this symmetry, as one sees by integration by parts and using the classical equations of motion for B and \tilde{B} that can be deduced from the ghost action (3.20). As we discussed in relation to eqn. (2.17), instead of using the classical equations of motion for B and \tilde{B} , it is better to accompany the transformation (3.57) of $\delta\mathcal{J}$ by a shift in C and \tilde{C} such that \hat{I} is invariant. Since we integrate out C and \tilde{C} in defining $F(\mathcal{J}, \delta\mathcal{J})$, the invariance of the extended action \hat{I} under the shift (3.57) of $\delta\mathcal{J}$ together with a suitable shift in C and \tilde{C} implies that $F(\mathcal{J}, \delta\mathcal{J})$ is invariant under the shift in $\delta\mathcal{J}$. (After introducing vertex operators, we spell out some details of this argument in section 4.2.)

Thus, roughly speaking, $F(\mathcal{J}, \delta\mathcal{J})$ is a pullback from the moduli space \mathfrak{M}_g of super Riemann surfaces of genus g . Moreover, counting both holomorphic and antiholomorphic variables, $F(\mathcal{J}, \delta\mathcal{J})$ has the right superdegree to be integrated over \mathfrak{M}_g . So we can define the integral of $F(\mathcal{J}, \delta\mathcal{J})$ over this moduli space, and this will give the genus g contribution to the vacuum amplitude of the heterotic string.

As usual, some subtleties arise in the infrared region because \mathfrak{M}_g is not compact; these will occupy our attention later. But as a preliminary, one has to grapple with an important detail.

3.3.2. A hard-to-avoid detail To explain the relevant point without excess clutter, let us simplify and suppose that the moduli space \mathfrak{M} of super Riemann surfaces has dimension $1|2$ rather than $3g - 3|2g - 2$. Then locally we could pick holomorphic coordinates $m|\eta^1, \eta^2$ on \mathfrak{M} . Any such coordinate system would be valid only locally in \mathfrak{M} . In another region of \mathfrak{M} , we might use another holomorphic coordinate system $m'|\eta'^1, \eta'^2$, related to the first by

$$(3.58) \quad \begin{aligned} m' &= f(m|\eta^1, \eta^2) \\ \eta'^i &= \psi^i(m|\eta^1, \eta^2), \quad i = 1, 2, \end{aligned}$$

with some holomorphic functions $f|\psi^1, \psi^2$. \mathfrak{M} is said to be holomorphically projected if it is possible to choose the coordinates so that $f(m|\eta^1\eta^2)$ is a function of m only, and not of the η 's. The more general possibility is that

the relation is

$$(3.59) \quad m' = f_0(m) + \eta^1 \eta^2 f_2(m),$$

with functions f_0 and f_2 that depend on m only. \mathfrak{M} is said to be holomorphically split if it is possible to choose the coordinates so that f is independent of the η 's and the ψ^i are linear in the η 's. (With only two odd variables η^i , the second condition is trivial since the ψ^i are odd functions.)

It is rather special for a complex supermanifold to be holomorphically projected or split. There has been no reason to believe that the moduli space of super Riemann surfaces has this property, and recently it has been shown that this is not the case in general [45]. (Supermoduli space is holomorphically projected in low orders and this fact has been exploited in explicit computations, especially in [14].)

Now let us consider the antiholomorphic degrees of freedom of the heterotic string. If supermoduli space has dimension $1|2$, then the ordinary moduli space of bosonic Riemann surfaces has dimension $1|0$. We therefore describe the antiholomorphic structure of Σ by a local coordinate \tilde{m} . In another region of the moduli space, we would use another local coordinate \tilde{m}' related to the first by a holomorphic change of coordinates

$$(3.60) \quad \tilde{m}' = \tilde{f}(\tilde{m}).$$

What relation can we assume between \tilde{m} and m ? Naively one would like to claim that the coordinates can be chosen so that \tilde{m} is the complex conjugate of m and \tilde{m}' is the complex conjugate of m' . However, if \mathfrak{M} is not holomorphically projected, these assertions do not make sense. We cannot consistently claim that $\tilde{m} = \bar{m}$ and $\tilde{m}' = \bar{m}'$ if the relation between m' and m depends on the η 's while the relation between \tilde{m} and \tilde{m}' does not.

An elegant way to proceed²³ is as follows. We make use of the fact that the partition functions and correlation functions of the worldsheet conformal field theories that are relevant in string theory are all real-analytic. So we can analytically continue away from $\tilde{m} = \bar{m}$ as long as we do not go too far away. Similarly, if we insert a vertex operator at a point $\tilde{z}; z|\theta$; we can analytically continue away from $\tilde{z} = \bar{z}$, as long as we do not go too far. When we make this sort of analytic continuation, we can treat the moduli of left and right-movers on the worldsheet as independent complex variables.

²³See p. 95 of [50] for a very brief explanation, and section 5 of [15] for a more leisurely account, with discussion of some alternative approaches.

If we proceed in this way, the arguments of section 3.3.1 show that $F(\mathcal{J}, \delta\mathcal{J})$ is the pullback of a holomorphic form (which we denote by the same name) on a product $\mathcal{M}_L \times \mathcal{M}_R$, where \mathcal{M}_L and \mathcal{M}_R parametrize respectively the complex structures “seen” by the left-movers and right-movers.²⁴ Any of the oriented closed-string theories can be described in this way. For the heterotic string, \mathcal{M}_R is a copy of the moduli space of super Riemann surfaces, parametrizing holomorphic structures on Σ , and \mathcal{M}_L is a copy of the moduli space of ordinary Riemann surfaces, parametrizing antiholomorphic structures on Σ . Analytic continuation does not affect the statement that $F(\mathcal{J}, \delta\mathcal{J})$ is a form of superdegree $6g - 6|2g - 2$, which is the same as the complex dimension of $\mathcal{M}_L \times \mathcal{M}_R$, or the fact that F is closed, $dF(\mathcal{J}, \delta\mathcal{J}) = 0$.

In general, suppose that M is a complex supermanifold of dimension $p|q$, and $F(x, dx)$ is a closed holomorphic $p|q$ -form on M . Then F can be integrated on a subsupermanifold $\Gamma \subset M$ of real codimension $p|0$ (and thus real dimension $p|q$), with a result that only depends on the homology class of Γ . A natural source of such Γ 's is as follows. Let M_{red} be the reduced space of M , defined by setting the odd coordinates to zero. So M_{red} is an ordinary complex manifold of complex dimension p . Let Γ be a middle-dimensional cycle in M_{red} , or in other words a cycle of real codimension p . Then by relaxing the condition that the odd variables should vanish, $\Gamma \subset M_{\text{red}}$ can be thickened to a cycle $\Gamma \subset M$ of the same codimension in a way that is unique up to homology.²⁵ So we can define the integral $\int_{\Gamma} \mathcal{D}(x, dx) F(x, dx)$, and (if Γ is compact) it only depends on the homology class of Γ .

The reduced space $(\mathcal{M}_L \times \mathcal{M}_R)_{\text{red}} \subset \mathcal{M}_L \times \mathcal{M}_R$ is defined by setting to zero all of the odd coordinates. In the toy model above, the reduced space is defined by $\eta_1 = \eta_2 = 0$. The naive conditions $\tilde{m} = \bar{m}$, $\tilde{m}' = \bar{m}'$ make sense once we set the odd variables to zero, so they should be understood as defining a middle-dimensional cycle $\Gamma \subset (\mathcal{M}_L \times \mathcal{M}_R)_{\text{red}}$. Then we thicken Γ to a cycle $\Gamma \subset \mathcal{M}_L \times \mathcal{M}_R$. Though the thickening from Γ to Γ is not completely canonical, the integral

$$(3.61) \quad Z_g = \int_{\Gamma} \mathcal{D}(\mathcal{J}, \delta\mathcal{J}) F(\mathcal{J}, \delta\mathcal{J})$$

²⁴To be more exact, $F(\mathcal{J}, \delta\mathcal{J})$ is defined and holomorphic in an open set of $\mathcal{M}_L \times \mathcal{M}_R$ that includes the region near $\tilde{m} = \bar{m}$ that will be used in the following construction.

²⁵The idea is very simple: whatever equations define Γ , one includes the odd variables in those equations in an arbitrary way to get Γ . Because the odd variables are infinitesimal, there is no topology in whatever choices have to be made. See appendix D for an example of choosing Γ , and eqn. D.12 for an explanation of the fact that in the example, the possible choices are homologous.

would be independent of the choices involved if Γ were compact. As it is, what we have said about the definition of Γ needs to be supplemented with an explanation of how Γ should behave at infinity. We postpone this until we have acquired in section 6 a basic understanding of the behavior of the superstring measure in the infrared region.

Γ (and its generalizations with vertex operator insertions) is the appropriate integration cycle for heterotic string perturbation theory. The integral (3.61) is the g -loop contribution to the vacuum amplitude of the heterotic string.

One might have hoped to define a natural moduli space of heterotic string worldsheets over which one would integrate to compute scattering amplitudes. This appears to be too optimistic, for the reasons sketched above. The integration cycle Γ is what exists instead.

3.4. Integration over a slice

In the analytic continuation of section 3.3.2, $F(\mathcal{J}, \delta\mathcal{J})$ is understood as a holomorphic form on $\mathcal{M}_L \times \mathcal{M}_R$ (which is supposed to be integrated on a cycle Γ that is close to the “diagonal”). To make the following discussion concrete, it is useful to pick local holomorphic coordinates $m_1 \dots m_{3g-3} | \eta_1 \dots \eta_{2g-2}$ on \mathcal{M}_R ; we often denote them collectively as \mathbf{m}_α , $\alpha = 1 \dots | \dots 2g - 2$. And we pick local holomorphic coordinates $\tilde{m}_1 \dots \tilde{m}_{3g-3}$ on \mathcal{M}_L . We make no assumption about all these parameters except that they give a good parametrization of a region of $\mathcal{M}_L \times \mathcal{M}_R$ (sufficiently close to the “diagonal” $\tilde{m}_\alpha = \overline{m}_\alpha$) in which we want to get a better understanding of the heterotic string path integral.

The modular parameters \tilde{m}_β and \mathbf{m}_α determine the supercomplex structure \mathcal{J} of Σ up to diffeomorphism. Just as in our study of the bosonic string in section 2.3, we can make the formula for the vacuum amplitude Z_g more concrete by picking a local slice transverse to the action of the diffeomorphism group on \mathcal{J} . Picking a slice means that we pick a definite family of \mathcal{J} 's parametrized by the \tilde{m}_β and \mathbf{m}_α . We then regard \mathcal{J} as a function of the \tilde{m}_β and \mathbf{m}_α , and similarly we express $\delta\mathcal{J}$ in terms of the corresponding differentials $d\tilde{m}_\beta$ and $d\mathbf{m}_\alpha$:

$$\begin{aligned}
 \delta\mathcal{J}_z^z &= \sum_{\alpha=1\dots|\dots 2g-2} \frac{\partial\mathcal{J}_z^z}{\partial\mathbf{m}_\alpha} d\mathbf{m}_\alpha \\
 \delta\mathcal{J}_\theta^{\tilde{z}} &= \sum_{\beta=1\dots 3g-3} \frac{\partial\mathcal{J}_\theta^{\tilde{z}}}{\partial\tilde{m}_\beta} d\tilde{m}_\beta.
 \end{aligned}
 \tag{3.62}$$

The worldsheet action of the heterotic string will in general have a complicated dependence on the modular parameters \mathbf{m}_α and \tilde{m}_β . But it is always linear in the $d\mathbf{m}$'s and $d\tilde{m}$'s, since it is linear in $\delta\mathcal{J}$. So – as in section 2.3 – it is possible to “integrate out” the differentials $d\mathbf{m}_\alpha$ and $d\tilde{m}_\beta$, to get a reduced description with fewer variables.

Going back to (3.24), we see that the part of the action that is linear in the differentials is

$$(3.63) \quad I_{d\mathbf{m},d\tilde{m}} = \sum_{\alpha=1\dots|2g-2} d\mathbf{m}_\alpha B^{(\alpha)} + \sum_{\beta=1\dots 3g-3} d\tilde{m}_\beta \tilde{B}^{(\beta)},$$

with

$$(3.64) \quad \begin{aligned} B^{(\alpha)} &= \frac{1}{2\pi} \int_{\Sigma} \mathcal{D}(\tilde{z}, z|\theta) \frac{\partial \mathcal{J}_z^z}{\partial \mathbf{m}_\alpha} B_{z\theta} \\ \tilde{B}^{(\beta)} &= \frac{1}{2\pi} \int_{\Sigma} \mathcal{D}(\tilde{z}, z|\theta) \frac{\partial \mathcal{J}_\theta^z}{\partial \tilde{m}_\beta} \tilde{B}_{z\tilde{z}}. \end{aligned}$$

We can think of $B^{(\alpha)}$ and $\tilde{B}^{(\beta)}$ as the antighost modes that are conjugate to the moduli \mathbf{m}_α and \tilde{m}_β . From (3.63), the integral over the differentials gives

$$(3.65) \quad \begin{aligned} \prod_{\beta=1\dots 3g-3} \int \mathcal{D}(d\tilde{m}_\beta) \exp(-d\tilde{m}_\beta \tilde{B}^{(\beta)}) \cdot \prod_{\alpha=1\dots|2g-2} \mathcal{D}(d\mathbf{m}_\alpha) \exp(-d\mathbf{m}_\alpha B^{(\alpha)}) \\ = \prod_{\beta=1\dots 3g-3} \delta(\tilde{B}^{(\beta)}) \cdot \prod_{\alpha=1\dots|2g-2} \delta(B^{(\alpha)}). \end{aligned}$$

In short, we can integrate out the differentials in favor of a delta function for each conjugate antighost mode. We abbreviate the product of delta functions that we have just obtained as

$$(3.66) \quad \delta^{3g-3}(\tilde{B}^{(\beta)}) \delta^{3g-3|2g-2}(B^{(\alpha)}).$$

We now define a function $\Lambda(\tilde{m}, \mathbf{m})$ of the moduli \tilde{m}_β and \mathbf{m}_α by performing a path integral over all worldsheet fields of the heterotic string (both matter and superghosts) with the above delta functions included and the moduli held fixed:

$$(3.67) \quad \Lambda(\tilde{m}, \mathbf{m}) = \int \mathcal{D}(X, B, C, \tilde{B}, \tilde{C}) \exp(-I) \delta^{3g-3}(\tilde{B}^{(\beta)}) \delta^{3g-3|2g-2}(B^{(\alpha)}).$$

(Λ was first introduced in [5] and used to define what we will shortly call Ξ .) The argument given in the last paragraph of section 2.3 applies here to show that $\Lambda(\tilde{m}, \mathbf{m})$ does not depend on the choice of slice that was used in the computation. The delta functions depend on the slice, but the dependence disappears when we integrate over the fields C and \tilde{C} .

The delta function insertions in (3.67) are actually needed in order for the worldsheet path integral to give a sensible and non-zero result. An antighost zero-mode that is not removed by a delta function would cause the worldsheet path integral to vanish (in the case of a fermionic zero-mode) or to diverge (in the case of a bosonic one). As usual, when studying vacuum amplitudes, we restrict to $g \geq 2$ so that there are no C or \tilde{C} zero-modes and hence delta functions for antighost zero-modes suffice to give a sensible path integral. Later, when we include external vertex operators, we will always, for all $g \geq 0$, have a suitable set of delta functions either present in the vertex operators or coming from integrating out the $d\mathbf{m}$'s and $d\tilde{m}$'s so that the path integral is sensible and not trivially zero.

Let us use an even more condensed notation in which x represents all moduli \mathbf{m}_α and \tilde{m}_β together. With the differentials dx included, there is a natural measure $\mathcal{D}(x, dx)$ for the combined system of moduli and their differentials. (Here “natural” means that $\mathcal{D}(x, dx)$ is invariant under any reparametrization of the x 's together with the induced transformation of the dx 's.) As usual, this is so because of the way the variables come in pairs with opposite statistics. Although $\mathcal{D}(x, dx)$ is completely natural, if we factor it as the product of a measure for the x 's,

$$(3.68) \quad \mathcal{D}(x) = [d\tilde{m}_1 \dots d\tilde{m}_{3g-3}; dm_1 \dots dm_{3g-3} | d\eta_1 \dots d\eta_{2g-2}]$$

and the measure for the dx 's that we used in integrating them out in eqn. (3.65), then neither factor is natural by itself. One way to get a natural formula is to multiply $\mathcal{D}(x)$ by the function $\Lambda(x)$ that we generated by integrating over the dx 's as well as other variables:

$$(3.69) \quad \Xi(x) = \mathcal{D}(x) \Lambda(x).$$

Concretely, if we transform to a new set of holomorphic coordinates on $\mathcal{M}_L \times \mathcal{M}_R$, then $\mathcal{D}(x)$ is multiplied by the Berezinian of the change of coordinates (this Berezinian is the superanalogue of the Jacobian for a change of coordinates), while the inverse of this factor multiplies the product of delta functions in (3.66), and likewise therefore the function $\Lambda(x)$.

Geometrically, $\Xi(x)$ is a holomorphic section of the Berezinian line bundle of $\mathcal{M}_L \times \mathcal{M}_R$. This is just a fancy way to say that $\Xi(x)$ is a well-defined measure on $\mathcal{M}_L \times \mathcal{M}_R$, in the holomorphic sense. The relation between $F(\mathcal{J}, \delta\mathcal{J})$ and $\Xi(x)$ is just that

$$(3.70) \quad \Xi(x) = \int \mathcal{D}(dx) F(\mathcal{J}, \delta\mathcal{J}).$$

Concretely, on the right hand side, $F(\mathcal{J}, \delta\mathcal{J})$ was defined by an integral over heterotic string matter fields and superghosts, while on the left hand side $\Xi(x)$ was defined by integrating over the dx 's as well as the matter fields and superghosts. So we can map F to Ξ by integrating out the dx 's. In general (see for example section 3.3.3 of [15]), integrating out the differentials in this fashion gives a natural map from a pseudoform, in our case $F(\mathcal{J}, \delta\mathcal{J})$, to a section of the Berezinian, here $\Xi(x)$.

On a complex supermanifold M of complex dimension $p|q$, in general a holomorphic section of the Berezinian can be integrated on a cycle of real codimension $p|0$; the integral only depends on the homology class of the cycle. In our case, the cycle on which we wish to integrate $\Xi(x)$ is the integration cycle Γ of the heterotic string. This integral gives the genus g contribution to the heterotic string vacuum amplitude:

$$(3.71) \quad Z_g = \int_{\Gamma} \Xi(x).$$

The formula (3.71) is completely equivalent to our earlier formula (3.61) for Z_g . The two formulas arise by performing the same integrals in different orders. Eqn. (3.61) arises in a procedure in which we compute Z_g by integrating first over matter fields and superghosts to compute $F(\mathcal{J}, \delta\mathcal{J})$, after which we integrate over x and dx to get Z_g . Eqn. (3.71) corresponds to a procedure in which we first integrate over dx to get the product of delta functions in (3.66), then integrate over matter fields and superghosts to compute $\Xi(x)$, and finally integrate over x .

The advantage of eqn. (3.61) is that it enables us to maintain manifest BRST symmetry at all stages. Eqn. (3.71) has enabled us to make contact with the literature, and has other applications as well.

3.4.1. More on changes of coordinates By now we understand that the vacuum amplitude of the heterotic string can be computed by integrating the naturally-defined section $\Xi(x)$ of $Ber(\mathcal{M}_L \times \mathcal{M}_R)$ or equivalently the pseudoform $F(\mathcal{J}, \delta\mathcal{J})$ over the integration cycle Γ . Moreover, once we introduce vertex operators, the same machinery can be used to compute scattering

amplitudes. The only possible problems have to do with the behavior in the infrared region at infinity.

Here we will describe an important subtlety²⁶ that arises when one tries to choose coordinates and evaluate the integral over Γ . This subtlety is related to what was described in section 3.3.2, so we will return to the simplified model in which we suppose that \mathcal{M}_R has dimension 1|2, with local coordinates $m|\eta^1, \eta^2$, and \mathcal{M}_L has dimension 1|0 with a local coordinate \tilde{m} .

With only two odd coordinates, the general possible form of an even section Ξ of $Ber(\mathcal{M}_L \times \mathcal{M}_R)$ is

$$(3.72) \quad \Xi = [d\tilde{m}; dm|d\eta^1, d\eta^2] \left(\Upsilon_0(\tilde{m}, m) + \Upsilon_2(\tilde{m}, m)\eta^1\eta^2 \right).$$

Naively speaking, since we are planning to perform a Berezin integral over all variables including η^1 and η^2 , it looks like we only care about Υ_2 ; one would think that Υ_0 is projected out when we integrate over η^1 and η^2 .

The trouble with this way of thinking is that although Ξ is a naturally-defined object, its separation into Υ_0 and $\Upsilon_2\eta^1\eta^2$ is not natural. Consider the change of coordinates from $m|\eta^1, \eta^2$ to

$$(3.73) \quad \begin{aligned} m' &= m + a(m)\eta^1\eta^2 \\ \eta'^1 &= \eta^1 \\ \eta'^2 &= \eta^2. \end{aligned}$$

In the new coordinate system,

$$(3.74) \quad \Xi = [d\tilde{m}; dm'|d\eta'^1, d\eta'^2] \left(\Upsilon'_0(\tilde{m}, m') + \Upsilon'_2(\tilde{m}, m')\eta'^1\eta'^2 \right),$$

with

$$(3.75) \quad \Upsilon'_2(\tilde{m}, m') = \Upsilon_2(\tilde{m}, m') - \partial_{m'}(a(m')\Upsilon_0(\tilde{m}; m')).$$

We see that we need to know Υ_0 in the old coordinate system in order to compute Υ'_2 in the new coordinate system.

It is true that the term in Υ'_2 that depends on Υ_0 is a total derivative, so one might hope that it would integrate to zero. However, such arguments are not as useful in practice as one might think. Any coordinate system is only valid locally, and locally anything is a total derivative. It is only if \mathcal{M}_R is holomorphically projected, so that we can avoid ever making changes of

²⁶See [48] for an early treatment of some of these issues.

coordinates in which an even coordinate is shifted by a fermion bilinear, that one can in a fairly natural way forget Υ_0 and study only Υ_2 .

The moduli space of super Riemann surfaces is actually not holomorphically projected in general [45]. Even if it were, relying on such a projection would possibly make it difficult to understand the gauge invariance and space-time supersymmetry of superstring scattering amplitudes. Finally, it turns out that even in cases in which \mathcal{M}_R is holomorphically projected, the proper definition of the integration cycle Γ (in the infrared region at infinity that we will study in section 6) is sometimes such that a naive computation using the holomorphic projection and throwing away Υ_0 gives the wrong answer. For an example, see appendix D.

In short, to understand superstring theory without generating multiple complications, one has to work with the naturally-defined objects $\Xi(x)$ or $F(\mathcal{J}, \delta\mathcal{J})$, rather than projections of them that depend on coordinate choices.

In the superstring literature of the 1980s, one typically integrated over the odd variables to get a measure depending only on the even moduli \tilde{m}_β and m_α . This gave formulas that were only locally defined and depended on a specific procedure for integrating over the odd variables. The formulas would change by a total derivative, as in eqn. (3.75) above, if one changed the procedure. It was necessary to keep track of these total derivatives. All this greatly added to the complexity of the subject. Working on supermoduli space with the natural objects $\Xi(x)$ and $F(\mathcal{J}, \delta\mathcal{J})$ may or may not help with practical calculations, but it does make the conceptual framework more simple.

3.5. A basis of odd moduli

Now we will describe the parametrizations of $\mathcal{M}_L \times \mathcal{M}_R$ that are most usual in practice. One begins with a purely bosonic Riemann surface Σ_{red} with local moduli \tilde{m}_β and m_α , and with a choice of spin structure. The spin structure determines a square root of the canonical bundle of Σ_{red} that we call $K^{1/2}$; its inverse we call $T^{1/2}$. From Σ_{red} , we build a super Riemann surface Σ by adding an odd variable θ that is valued in $T^{1/2}$.

We want to deform Σ by turning on a gravitino field χ_z^θ . This field enters the general formula eqn. (3.7) for the (1, 0) part of the variation of \mathcal{J} :

$$(3.76) \quad \delta\mathcal{J}_z^z = h_z^z + \theta\chi_z^\theta.$$

Our main interest here is not the bosonic field h_z^z that describes the bosonic moduli m_α , but the gravitino field χ_z^θ that describes the odd moduli. At the

linearized level, it is subject to the gauge equivalence

$$(3.77) \quad \chi_z^\theta \rightarrow \chi_z^\theta + \partial_z y^\theta.$$

(This equivalence follows from the action (3.10) of a vector field $q^z \partial_z$ on $\delta \mathcal{J}_z^z$; one expands $q^z(\tilde{z}; z|\theta) = s^z(\tilde{z}; z) + \theta y^\theta(\tilde{z}; z)$ and then (3.77) is the symmetry generated by y^θ .) Modulo this gauge equivalence, χ_z^θ takes values in the sheaf cohomology group $H^1(\Sigma_{\text{red}}, T^{1/2})$, whose dimension is $2g - 2$.

So we can specify a family of fermionic deformations of Σ by picking a set of $2g - 2$ c -number χ_z^θ fields, representing a basis of $H^1(\Sigma_{\text{red}}, T^{1/2})$. We call these fields $\chi_z^{(\sigma)\theta}$, $\sigma = 1, \dots, 2g - 2$. Any generic set of $\chi^{(\sigma)}$'s will do, since the condition that they project to linearly independent elements of $H^1(\Sigma_{\text{red}}, T^{1/2})$ is satisfied generically. Then we expand the gravitino field as

$$(3.78) \quad \chi_z^\theta = \sum_{\sigma=1}^{2g-2} \eta_\sigma \chi_z^{(\sigma)\theta},$$

with odd coefficients η_σ . The η_σ can serve as the odd moduli of \mathcal{M}_R , in a suitable region.

This is actually a convenient way to give a local parametrization of the odd directions in supermoduli space. However, there are two very important pitfalls. The first is simply that we have to make sure that the $\chi^{(\sigma)}$'s do give a basis for $H^1(\Sigma_{\text{red}}, T^{1/2})$. This is untrue precisely if a linear combination of the $\chi^{(\sigma)}$'s, with c -number coefficients e_σ that are not all zero, can be gauged away by some y^θ :

$$(3.79) \quad \partial_z y^\theta = \sum_\sigma e_\sigma \chi_z^{(\sigma)\theta}.$$

The cokernel of the operator ∂_z , mapping sections of $T^{1/2}$ to $(0, 1)$ -forms with values in $T^{1/2}$, is $H^1(\Sigma_{\text{red}}, T^{1/2})$, which has dimension $2g - 2$, and generically to get a non-zero solution of an equation such as (3.79) with a non-zero source on the right hand side, we need to adjust $2g - 1$ parameters.²⁷ The $2g - 2$ parameters e_σ are not quite enough, but if we also vary one of the bosonic moduli m_α , then we have enough parameters so that it is natural for eqn.

²⁷One of these parameters is an overall scaling of the right hand side of the equation, or in other words of the e_σ ; this does not affect whether the equation has a solution. Modulo this scaling, we need to adjust $2g - 2$ parameters to get a solution, the same number as the dimension of the cohomology group.

(3.79) to have a solution for some isolated values of m_α . The fermionic gauge-fixing that is defined by the choice of slice (3.78) is then wrong at such a value of m_α , and the measure on supermoduli space, computed as in section 3.6.2 using this slice, will develop a pole. (Explicitly, the pole arises because the product of delta functions in (3.93) does not remove all of the zero-modes of the field $\beta_{z\theta}$.)

Except in very low genus, this phenomenon is practically unavoidable and in the literature there is a technical name for these poles – they are called spurious singularities. They are spurious because they result from an invalid gauge-fixing. To compute correctly in a formalism with spurious singularities can become complicated in general, as one has to keep track of total derivatives supported on the locus of spurious singularities. We will not try to explain how to do that here. Rather, our goal is to understand how to calculate in a framework in which all singularities are physically sensible and represent the effects of on-shell particles. So we want to avoid spurious singularities.

That does not mean avoiding the use of the slices we have described. It means that such a slice should be used in the following sense. We pick a suitable set of $\chi^{(\sigma)}$'s and a sufficiently small open set \mathcal{U} in $\mathcal{M}_L \times \mathcal{M}_R$ so that eqn. (3.79) has no solution when Σ_{red} is parametrized by \mathcal{U} . Then, in that region, we use eqn. (3.78) to define fermionic moduli η_σ . Together with the moduli m_α and \tilde{m}_β of Σ_{red} , this gives a local coordinate system on $\mathcal{M}_L \times \mathcal{M}_R$. In this coordinate system, we compute the invariant objects $F(\mathcal{J}, \delta\mathcal{J})$ or $\Xi(x)$ on $\mathcal{M}_L \times \mathcal{M}_R$ that have to be integrated to compute the heterotic string amplitude. Of course, with a particular basis of gravitino modes, this procedure works only in a suitable open set \mathcal{U} . But we can cover $\mathcal{M}_L \times \mathcal{M}_R$ with open sets $\mathcal{U}^{(\zeta)}$ in each of which we can pick an appropriate basis $\chi^{(\sigma;\zeta)}$ of gravitino fields such that eqn. (3.79) nowhere has a non-zero solution. This gives us in each $\mathcal{U}^{(\zeta)}$ a good set of odd moduli $\eta_{\sigma;\zeta}$. Using these local slices, we compute $F(\mathcal{J}, \delta\mathcal{J})$ or $\Xi(x)$ in each $\mathcal{U}^{(\zeta)}$; the results automatically agree on intersections $\mathcal{U}^{(\zeta)} \cap \mathcal{U}^{(\zeta')}$ since $F(\mathcal{J}, \delta\mathcal{J})$ and $\Xi(x)$ are independent of the coordinate choices.

After covering $\mathcal{M}_L \times \mathcal{M}_R$ by the open sets $\mathcal{U}^{(\zeta)}$ and defining a basis of odd moduli $\eta_{\sigma;\zeta}$, $\sigma = 1, \dots, 2g - 2$ in each open set as in the last paragraph, what we do *not* want to do is to integrate over the $\eta_{\sigma;\zeta}$ in each $\mathcal{U}^{(\zeta)}$ to get a measure on the reduced space parametrized by the bosonic moduli m_α and \tilde{m}_β only. These computations would not fit together in a simple way on intersections $\mathcal{U}^{(\zeta)} \cap \mathcal{U}^{(\zeta')}$, for the following reason. As we will see shortly, when we change basis from $\chi^{(\sigma;\zeta)}$ to $\chi^{(\sigma;\zeta')}$ in the intersection $\mathcal{U}^{(\zeta)} \cap \mathcal{U}^{(\zeta')}$, the moduli m_α undergo the sort of transformation that we studied in eqn. (3.73).

In the language of that discussion, integrating over the odd moduli $\eta_{\sigma;\zeta}$ in each $\mathcal{U}^{(\zeta)}$ would entail dropping Υ_0 and keeping Υ_2 . A correct computation has to keep track of Υ_0 . Conceptually, the simplest correct procedure is to use the local slices to compute the invariant objects $F(\mathcal{J}, \delta\mathcal{J})$ or $\Xi(x)$ (or their generalizations with external vertex operators), and then evaluate the amplitude using (3.61) or (3.71). In that way, we extract only invariant information from the local slices and we never meet spurious singularities. It is also possible but cumbersome to describe a correct procedure without the geometric interpretation that we have just explained. See for example [49].

We still have to explain why, when we change basis for the gravitino modes from $\chi^{(\sigma;\zeta)}$ to $\chi^{(\sigma;\zeta')}$, the bosonic moduli undergo a transformation such as (3.73). Any change of basis for the $\chi^{(\sigma)}$ can be made by replacing them with linear combinations of themselves and modifying them by gauge transformations. The subtlety entirely comes from the gauge transformations.²⁸ To implement a general change of basis, we need to make a gauge transformation on each $\chi^{(\sigma)}$ separately

$$(3.80) \quad \chi_z^{(\sigma)\theta} \rightarrow \chi_z^{(\sigma)\theta} + \partial_z y^{(\sigma)\theta}.$$

This amounts to the gauge transformation

$$(3.81) \quad \chi \rightarrow \chi + \partial_z y,$$

with χ as in (3.78) and

$$(3.82) \quad y = \sum_{\sigma=1}^{2g-2} \eta_{\sigma} y^{(\sigma)}.$$

The problem is that although in leading order in odd variables, a gauge transformation with parameter y acts on χ as in (3.77) while leaving h invariant, in the next order, h_z^z is transformed by

$$(3.83) \quad h_z^z \rightarrow h_z^z + y^{\theta} \chi_z^{\theta}.$$

In the present context, this is

$$(3.84) \quad h_z^z \rightarrow h_z^z + \sum_{\sigma, \sigma'} \eta_{\sigma} \eta_{\sigma'} y^{(\sigma)\theta} \chi_z^{(\sigma')\theta}.$$

²⁸The reason for this is that replacing the $\chi^{(\sigma)}$ by linear combinations of themselves does not change the space of gravitino modes over which we are integrating, only its parametrization by the η 's. See the discussion of eqn. (3.91).

This deformation of h_z^z arises as soon as there are two or more odd moduli. Importantly, even though it is induced by a gauge transformation of χ , the shift in h_z^z is not a pure gauge from a purely bosonic point of view, since it is not of the form $\partial_z w^z$ for any w^z . So this shift in h_z^z changes the bosonic geometry in a nontrivial fashion and shifts the even moduli m_α by bilinears in the η 's.

In the context of supergeometry, this mixing of even and odd variables is quite natural and one should not expect to be able to define bosonic moduli while ignoring the existence of odd variables. So the result probably should not be a surprise. It implies that except in special cases in low orders, one should not expect to be able to avoid the subtleties described in sections 3.3.2 and 3.4.1.

3.6. Integrating out odd variables

3.6.1. General procedure We will now make the considerations of section 3.5 more concrete. The worldsheet action of the heterotic string is actually linear in the gravitino field χ_z^θ , so the dependence of the action on this field is entirely contained in the linear formula (3.11). (For Type II superstrings, matters are a little more complicated; the action has a term that involves the product $\chi_z^\theta \tilde{\chi}_z^\theta$.) For the gravitino field (3.78), the gravitino coupling comes out to be

$$(3.85) \quad I_\eta = \sum_{\sigma=1}^{2g-2} \frac{\eta_\sigma}{2\pi} \int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} S_{z\theta}.$$

(This is an integral on the reduced space Σ_{red} , with θ integrated out; θ is similarly integrated out in the following formulas.)

There is also a corresponding coupling of $d\eta_\sigma$ to the holomorphic antighost field $B_{z\theta}$. We can deduce it from (3.24). It actually is related to (3.85) by the BRST symmetry $\{Q_B, \eta_\sigma\} = d\eta_\sigma$, $[Q_B, \beta_{z\theta}] = S_{z\theta}$:

$$(3.86) \quad I_{d\eta} = \sum_{\sigma=1}^{2g-2} \frac{d\eta_\sigma}{2\pi} \int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} \beta_{z\theta}.$$

We have given the reader fair warning that in general it is not illuminating to integrate out the odd variables. On the other hand, it is also important to know what happens when we do integrate them out, both to make contact with the literature and more importantly because integrating out the odd variables is useful for performing practical calculations in low orders.

According to the above formulas, for each σ , the dependence of the integrand of the path integral on the pair $\eta_\sigma, d\eta_\sigma$ is a simple factor

$$(3.87) \quad \exp\left(-\frac{1}{2\pi} \int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} (\eta_\sigma S_{z\theta} + d\eta_\sigma \beta_{z\theta})\right).$$

If we integrate out the pair η_σ and $d\eta_\sigma$, we get simply

$$(3.88) \quad \delta\left(\int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} \beta_{z\theta}\right) \cdot \int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} S_{z\theta}.$$

This is a standard formula; see [5, 7] or eqn. (3.335) in [8]. Usually the differentials $d\eta_\sigma$ are not introduced explicitly in the formalism, and one says simply that (3.88) results from integration over η_σ . Since an odd variable is its own delta function, we can also write (3.88) with the two factors treated more symmetrically:

$$(3.89) \quad \delta\left(\int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} \beta_{z\theta}\right) \cdot \delta\left(\int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} S_{z\theta}\right).$$

The result (3.88) or (3.89) is invariant under a rescaling of the gravitino mode $\chi_z^{(\sigma)\theta}$, because the two factors transform oppositely. A more fundamental explanation of why such rescaling does not matter is that the gravitino field (3.78) is unchanged under

$$(3.90) \quad \chi_z^{(\sigma)\theta} \rightarrow \lambda \chi_z^{(\sigma)\theta}, \quad \eta_\sigma \rightarrow \lambda^{-1} \eta_\sigma, \quad \lambda \in \mathbb{C}^*,$$

and the extended action is invariant if we also rescale $d\eta_\sigma$ by a factor of λ^{-1} . So when we integrate out η_σ and $d\eta_\sigma$ – an operation that is certainly invariant under a common rescaling of η_σ and $d\eta_\sigma$ – we get a result that is invariant under rescaling of the gravitino mode $\chi_z^{(\sigma)\theta}$.

An important generalization of what we have just explained is that the operation of integrating out several odd moduli η_1, \dots, η_s only depends on the linear span of the gravitino fields $\chi_z^{(1)\theta} \dots \chi_z^{(s)\theta}$. The reason for this is that in the definition of the gravitino field and in the extended action we can compensate for a linear transformation

$$(3.91) \quad \begin{pmatrix} \chi_z^{(1)} \\ \vdots \\ \chi_z^{(s)\theta} \end{pmatrix} \rightarrow M \begin{pmatrix} \chi_z^{(1)} \\ \vdots \\ \chi_z^{(s)\theta} \end{pmatrix}$$

for any invertible matrix M by

$$(3.92) \quad \begin{aligned} (\eta_1 \ \dots \ \eta_s) &\rightarrow (\eta_1 \ \dots \ \eta_s) M^{-1}, \\ (d\eta_1 \ \dots \ d\eta_s) &\rightarrow (d\eta_1 \ \dots \ d\eta_s) M^{-1}. \end{aligned}$$

The combined operation leaves fixed the gravitino field and the extended action. Integrating out the η 's and $d\eta$'s is invariant under the transformation (3.92). So the factor

$$(3.93) \quad \prod_{\sigma=1}^s \left(\delta \left(\int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} \beta_{z\theta} \right) \cdot \int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} S_{z\theta} \right)$$

that comes by integrating out the η 's and $d\eta$'s is invariant under (3.91), as one may readily verify.

3.6.2. The picture-changing operator The example of this construction that is most often considered is that each gravitino mode $\chi_z^{(\sigma)\theta}$ is a delta function supported at some point $p_\sigma \in \Sigma_{\text{red}}$:

$$(3.94) \quad \chi_z^{(\sigma)\theta} = \delta_{p_\sigma}.$$

There is a subtlety here: $\chi_z^{(\sigma)\theta}$ is supposed to be a $(0, 1)$ -form valued in $T^{1/2}$. The space of such forms with delta function support at p_σ is one-dimensional, but there is no natural way to pick a non-zero vector in this space. So there is no natural way to normalize the delta function in (3.94). However, if we are planning to integrate out the odd variables, this does not matter very much, because of an observation in section 3.6.1: the formalism is invariant under a rescaling of the gravitino modes (with a compensating rescaling of the odd modulus η_σ).

For a delta-function gravitino supported at $p \in \Sigma_{\text{red}}$, the factor (3.88) associated with integrating out an odd modulus becomes

$$(3.95) \quad \mathfrak{Y}(p) = \delta(\beta(p)) S_{z\theta}(p).$$

This is called the picture-changing operator. It was originally defined rather differently in [1]. The definition (3.95) was first presented in [7, 46], where the meaning of operators such as $\delta(\beta(p))$ was also analyzed. (We explore this subject further in section 10 of the present paper.) $\mathfrak{Y}(p)$ is a primary field of conformal dimension 0, the product of the supercurrent $S_{z\theta}$, which has dimension $3/2$, and $\delta(\beta)$, which has dimension $-3/2$.

Actually, the definition of $\mathcal{Y}(p)$ needs some clarification, because there is a singularity in the operator product of $S_{z\theta}(p)$ and $\delta(\beta(p))$. The usual approach is to fix the ambiguity (which involves an operator of the form $\delta'(\beta)\partial_z b_{zz}$, where b_{zz} is the fermionic antighost) by using BRST symmetry. Equivalently, one may regularize the definition of $\mathcal{Y}(p)$ by using a smooth gravitino wavefunction with very concentrated support and taking the limit as it approaches a delta function.

At least locally on the reduced space $(\mathcal{M}_L \times \mathcal{M}_R)_{\text{red}}$, one can use picture-changing operators to integrate over all odd moduli. For this, one picks points p_σ , $\sigma = 1, \dots, 2g - 2$, and one lets

$$(3.96) \quad \chi_{\tilde{z}}^{(\sigma)\theta} = \delta_{p_\sigma}.$$

Initially, we take the p_σ to be distinct so that the $\chi_{\tilde{z}}^{(\sigma)\theta}$'s are linearly independent and have a chance to provide a basis for $H^1(\Sigma_{\text{red}}, T^{1/2})$. With this choice of gravitino modes, integrating out the odd moduli will give a product of picture-changing operators

$$(3.97) \quad \prod_{\sigma=1}^{2g-2} \mathcal{Y}(p_\sigma).$$

It is interesting to ask when the use of picture-changing operators to integrate out all the odd moduli will lead to a spurious singularity. This will happen when there is a non-zero solution of the equation (3.79), which here reads

$$(3.98) \quad \partial_{\tilde{z}} y^\theta = \sum_{\sigma=1}^{2g-2} e_\sigma \delta_{p_\sigma}.$$

A solution y^θ of this equation is a section of $T^{1/2}$ that has simple poles at the points p_1, \dots, p_{2g-2} (with residues e_1, \dots, e_{2g-2}) and is holomorphic elsewhere. So y^θ is a holomorphic section of the line bundle $T^{1/2}(p_1 + \dots + p_{2g-2}) = T^{1/2} \otimes \mathcal{O}(\sum_{\sigma=1}^{2g-2} p_\sigma)$. This is, by definition, the line bundle whose holomorphic sections are sections of $T^{1/2}$ that may have simple poles at the p_i . It has degree $g - 1$, since $T^{1/2}$ has degree $-(g - 1)$ and allowing the poles increases the degree by $2g - 2$. A generic holomorphic line bundle of degree $g - 1$ has no holomorphic section; however, when one varies one complex parameter (a modulus of Σ_{red} , or the choices of the p_i), it is generic for such a section to arise at special values of this parameter. In particular, if

one integrates out the odd moduli of Σ via a product of picture-changing operators, one should expect to encounter spurious singularities as the bosonic moduli – the moduli of Σ_{red} – are varied. These are poles that appear where

$$(3.99) \quad H^0(\Sigma_{\text{red}}, T^{1/2}(\Sigma_{\sigma} p_{\sigma})) \neq 0.$$

So far, we have assumed that the points p_{σ} are distinct. However the criterion (3.99) for a spurious singularity makes perfect sense even if some of the p_{σ} coincide. In fact, the procedure of integrating out odd moduli by using gravitino wavefunctions with delta function support has a generalization (not a limit; the distinction will become clear) when some of the p_{σ} coincide. To explain this heuristically, let us choose a local coordinate z such that p_1 and p_2 correspond, say, to $z = 0$ and $z = \epsilon$. So two of our gravitino wavefunctions are

$$(3.100) \quad \chi_{\tilde{z}}^{(1)\theta} = \delta^2(z), \quad \chi_{\tilde{z}}^{(2)\theta} = \delta^2(z - \epsilon).$$

But the result of integrating out the odd moduli is invariant under a linear transformation (3.91) of the gravitino modes, and as a special case of this, we can replace $\chi_{\tilde{z}}^{(2)\theta}$ with

$$(3.101) \quad \hat{\chi}_{\tilde{z}}^{(2)\theta} = -\frac{1}{\epsilon} \left(\chi_{\tilde{z}}^{(2)\theta} - \chi_{\tilde{z}}^{(1)\theta} \right) = -\frac{1}{\epsilon} \left(\delta^2(z - \epsilon) - \delta^2(z) \right).$$

In this form, we can take the limit as $\epsilon \rightarrow 0$, getting

$$(3.102) \quad \hat{\chi}_{\tilde{z}}^{(2)\theta} = \partial_z \delta^2(z).$$

Thus if one gravitino mode is a delta function at a given point, it is fairly natural to take a second one to be the derivative of a delta function at the same point.

By evaluating (3.88), we find that if $\chi_{\tilde{z}}^{(\sigma)\theta} = \partial_z \delta^2(z - z_0)$ for some z_0 , then integrating out the corresponding odd modulus η_{σ} and its differential $d\eta_{\sigma}$ gives a factor

$$(3.103) \quad \delta(\partial_z \beta(z_0)) \partial_z S_{z\theta}(z_0).$$

Again some regularization of this singular product is required.

Similarly, the picture-changing formalism has a generalization when any number of the points p_1, \dots, p_{2g-2} coincide. If k of them are to coincide at some point in Σ_{red} , say $z = 0$, we take k of the gravitino wavefunctions to be $\delta^2(z), \partial_z \delta^2(z), \dots, \partial_z^{k-1} \delta^2(z)$.

The foregoing remarks must be treated with care. It is true, as we have just seen, that the picture-changing formalism has a natural analog when two or more of the points p_σ coincide. However, this analog is not simply the limit of the picture-changing formalism when, say, $p_1 \rightarrow p_2$. The reason is that there is a short-distance singularity in the operator product $\mathcal{Y}(p_1)\mathcal{Y}(p_2)$ for $p_1 \rightarrow p_2$. The above classical treatment amounted to normal-ordering this product and dropping the singular terms. The singular terms are Q_B -exact, so that their contribution to $F(\mathcal{J}, \delta\mathcal{J})$ is an exact form, but as usual such exact forms must be treated carefully. So the picture-changing formalism, defined initially for distinct points p_σ , has an analog when some of the points are taken to coincide, but it does not in general have a limit as the points approach each other.

3.6.3. Picture number We already observed in discussing eqn. (3.67) that on a genus g Riemann surface with no vertex operator insertions, we need $2g-2$ delta functions of the commuting ghost field $\beta_{z\theta}$ in order to get a sensible path integral. These insertions remove singularities that would otherwise arise from zero-modes of this field. These delta functions may be in general delta functions of arbitrary modes of the field $\beta_{z\theta}$, of the form

$$(3.104) \quad \delta \left(\int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} \beta_{z\theta} \right)$$

for any c -number gravitino field $\chi_z^{(\sigma)}$. Any generic $2g-2$ insertions of this kind will give a path integral that makes sense (in the absence of delta functions of γ , as discussed shortly).

We call an operator of the form (3.104) an operator of picture number 1. Generically, the picture number 1 operators in (3.104) are not local operators. They become local operators – although relatively unfamiliar ones – if $\chi_z^{(\sigma)\theta}$ has delta function or derivative of delta function support. In that case, we get local operators of the form $\delta(\beta(z))$, $\delta(\partial_z\beta(z))$, $\delta(\partial_z^2\beta(z))$, etc., all of picture number 1.

According to generalities of quantum field theory, once we allow a local operator $\delta(\beta(z))$, we also must allow its z -derivative $\partial_z\delta(\beta(z))$. Formally

$$(3.105) \quad \partial_z\delta(\beta(z)) = \partial_z\beta(z) \cdot \delta'(\beta(z)).$$

This indicates that a reasonable formalism will have to include not just delta function operators $\delta(\beta(z))$ but more general distributional operators $\delta'(\beta(z))$, $\delta''(\beta(z))$, and so on. In fact, an operator such as $\delta'(\beta(z))$ can remove a β zero-mode and help give a sensible path integral, just as $\delta(\beta)$ can. We will describe

this in more detail in section 10. So we assign picture number 1 to $\delta'(\beta(z))$, and similarly to $\delta''(\beta(z))$, $\delta'(\partial_z\beta(z))$, etc., just as to $\delta(\beta(z))$. Similarly we consider a nonlocal operator

$$(3.106) \quad \delta^{(k)} \left(\int_{\Sigma_{\text{red}}} d^2z \chi_z^{(\sigma)\theta} \beta_{z\theta} \right)$$

(where $\delta^{(k)}$ is the k^{th} derivative of a delta function) to have picture number 1.

All this matches integration theory on supermanifolds, as very briefly reviewed in section 3.2.1. In that subject, one assigns picture number -1 to $\delta(d\theta)$ (where θ is an odd coordinate so $d\theta$ is an even differential) and also to $\delta'(d\theta)$, $\delta''(d\theta)$, etc. (See eqn. (3.34) for a concrete example using a form $\delta'(d\theta)$.) The role of a picture number -1 operator is to reduce by 1 the number of $d\theta$'s over which one has to integrate, and any of these delta functions has that effect. The reader who wishes to understand more deeply the analogy between picture number in integration on supermanifolds and picture number in the $\beta\gamma$ system is urged to consult [11].

Now we will explain some important aspects of the $\beta\gamma$ system. Some readers may want to jump to section 10 (which does not depend on the intervening parts of this paper), where the following is explained more systematically. Readers who find what is stated in the next few paragraphs to be at least temporarily sufficient can simply proceed. In any event, the relevant facts are this. The ghost field γ is a section of a line bundle $T^{1/2}$. Suppose for simplicity that this line bundle has no holomorphic sections (which is the case for $g \geq 2$); then γ has no zero-modes. The index of the ∂_z operator acting on sections of $T^{1/2}$ is $-(2g - 2)$, so β has $2g - 2$ zero-modes and it takes $2g - 2$ insertions of $\delta(\beta)$ to get a sensible path integral. Now suppose that we include in the path integral a factor $\delta(\gamma(p))$ for some point p . Since this forces γ to vanish at p , it effectively replaces the line bundle $T^{1/2}$ by $T^{1/2}(-p)$ (whose sections are sections of $T^{1/2}$ that vanish at p). The index is now $-(2g - 1)$ and a sensible path integral requires $2g - 1$ operators of the general type $\delta(\beta)$. More generally, when delta functions of the commuting ghost γ are considered, as well as delta functions of β , one finds that a $\beta\gamma$ path integral with r delta functions of β and s delta functions of γ is sensible only if

$$(3.107) \quad r - s = 2g - 2.$$

(This criterion is necessary but only generically sufficient, since exceptionally it can happen that the delta functions do not remove all the zero-modes.) This statement is valid for all g , not just the case $g \geq 2$ that we started with.

The condition (3.107) does not depend on exactly what sort of delta functions are considered; from this point of view, we do not want to distinguish $\delta(\gamma)$ from $\delta(\partial_z\gamma)$, $\delta'(\gamma)$, etc., just as we did not make such a distinction for delta functions of β . So we assign picture number 1 to any delta function of β , regardless of details of its construction, and picture number -1 to any delta function of γ . Then the general statement is that the net picture number of all operators must add up to $2g - 2$ in order for the path integral to be sensible. Needless to say, superstring scattering amplitudes are always described by sensible path integrals.

The $\beta\gamma$ system also has an anomalous symmetry – the holomorphic ghost number $-$ under which β and γ have charges -1 and 1 , respectively. This quantum number is also carried by b and c (the $\chi_z^\theta S_{z\theta}$ term in the action does not allow separate ghost number symmetries for b , c and for β , γ , though there is a separate antiholomorphic ghost number symmetry carried by \tilde{b} , \tilde{c}). The b , c system has a ghost number anomaly $-(3g - 3)$ and the β , γ system has an anomaly $2g - 2$, so the net ghost number anomaly is $-(g - 1)$. So to get a non-zero path integral, we need to insert a product of operators with a net ghost number $-(g - 1)$. A typical example is the product of delta functions of b and β that we generated in (3.66). This was

$$(3.108) \quad \delta^{3g-3|2g-2}(B^{(\alpha)}) = \prod_{\alpha'=1}^{3g-3} \delta(b^{(\alpha')}) \prod_{\alpha''=1}^{2g-2} \delta(\beta^{(\alpha'')}).$$

Each $\delta(b)$ has ghost number -1 and each $\delta(\beta)$ has ghost number 1 . (That is because under the scaling $(b, \beta) \rightarrow \lambda^{-1}(b, \beta)$, which is the ghost number symmetry of the antighosts, $\delta(b)$ scales like b but $\delta(\beta)$ scales oppositely to β .) So the product of delta functions in (3.108) does have net ghost number $-(g - 1)$.

The assertion that the ghost number is an anomalous symmetry means that a path integral with operator insertions of the wrong ghost number is well-defined but vanishes. Picture number is not really a symmetry in that sense, anomalous or not, since a path integral with operator insertions of the wrong picture number is divergent or not well-defined, rather than being zero.

4. The Neveu-Schwarz sector

In this section, we begin our study of superstring vertex operators. For the most part, it does not matter much which supersymmetric string theory we consider, since left- and right-moving worldsheet degrees of freedom can be treated independently in many respects. For definiteness, we concentrate on

the heterotic string. In this case, there is a supersymmetric structure only for right-movers, and there are two types of vertex operator – Neveu-Schwarz (NS) and Ramond (R). We begin with the more straightforward NS case, deferring Ramond vertex operators to section 5.

4.1. Neveu-Schwarz vertex operators

A Neveu-Schwarz vertex operator is simply inserted at a point $\tilde{z}; z|\theta$ on a heterotic string worldsheet Σ . This operation increases the holomorphic dimension of the moduli space by 1|1, the extra moduli being simply the values of z and θ at which the vertex operator is inserted. Similarly, the antiholomorphic dimension of the moduli space is increased by 1|0, the extra modulus being \tilde{z} .

Let us try to understand, by analogy with section 2.4.1, what is the simplest type²⁹ of vertex operator that we can use to compute scattering amplitudes involving NS states.

Just as in section 2.4, for it to be possible to express scattering amplitudes in terms of forms on a finite-dimensional space (basically the moduli space of super Riemann surfaces with punctures), we will want the vertex operators to depend only on the ghost fields \tilde{c} , c , and γ , and not on their derivatives. An equivalent condition is that

$$(4.1) \quad b_n \mathcal{V} = \tilde{b}_n \mathcal{V} = \beta_r \mathcal{V} = 0, \quad n, r \geq 0,$$

where the b , \tilde{b} modes were defined in eqn. (2.39) and

$$(4.2) \quad \beta_r = \frac{1}{2\pi i} \oint dz z^{r+1/2} \beta_{z\theta}.$$

(The need for the condition (4.1) is explained in section 4.2.)

We certainly also want our vertex operators to be annihilated by the BRST charge Q_B . As in our discussion of the bosonic string, if $Q_B \mathcal{V} = 0$ and \mathcal{V} also obeys (4.1), then it follows that from a holomorphic point of view, \mathcal{V} is a superconformal primary of dimension 0

$$L_n \mathcal{V} = 0, \quad n \geq 0$$

²⁹Instead of the simplest choice, one can ask for the most general possible choice. As discussed for the bosonic string in section 2.4.4, in general we could use arbitrary Q_B -invariant operators that are annihilated by $b_0 - \tilde{b}_0$. To use this larger class of vertex operators, one needs a more elaborate formalism that does not assume superconformal symmetry.

$$(4.3) \quad G_r \mathcal{V} = 0, \quad r \geq 1/2,$$

and similarly that from an antiholomorphic point of view, \mathcal{V} is a conformal primary of dimension 0,

$$(4.4) \quad \tilde{L}_n \mathcal{V} = 0, \quad n \geq 0.$$

This latter condition is familiar from the bosonic string.

The conditions (4.3) have an intuitive meaning. Looking back to (3.4), we see that the generators L_n and G_r with $n, r \geq 0$ are precisely the ones that vanish at $z = \theta = 0$. So the point $z = \theta = 0$ is invariant under the symmetries generated by those operators, and if we want to insert a vertex operator at $z = \theta = 0$ that preserves all the superconformal symmetries that leave fixed this point, then that operator should obey (4.3). Similarly, (4.4) says that \mathcal{V} is invariant under antiholomorphic changes of coordinate that leave fixed the point at which it is inserted.

How should \mathcal{V} depend on the ghost fields? We can answer this as follows. Since inserting an NS vertex operator increases the antiholomorphic dimension of moduli space by $1|0$ and the holomorphic dimension by $1|1$, in a formalism similar to what we had for bosonic strings in section 2 or for the superstring vacuum amplitude in section 3, there will be an extra delta function insertion for each of \tilde{b} , b , and β . We write this schematically as

$$(4.5) \quad \delta(\tilde{b})\delta(b)\delta(\beta).$$

Let us consider first the holomorphic degrees of freedom. The operator $\delta(b)\delta(\beta)$ has ghost number 0 and picture number 1. We want \mathcal{V} to be such that when we include \mathcal{V} in the path integral and also include an operator with the quantum numbers $\delta(b)\delta(\beta)$, the path integral remains sensible and non-zero. For this, \mathcal{V} must have holomorphic ghost number 0 and picture number -1 . The only operator with these quantum numbers that does not depend on derivatives of c or γ is $c\delta(\gamma)$. We need the factor $\delta(\gamma)$ to get picture number -1 (we cannot use other delta functions such as $\delta(\partial_z \gamma)$ or $\partial_z \delta(\gamma) = \partial_z \gamma \delta'(\gamma)$, since they depend on derivatives of γ). And given that $\delta(\gamma)$ has ghost number -1 , we need a factor of c to make the holomorphic ghost number of \mathcal{V} vanish.

As a bonus, we can extend this argument to show that \mathcal{V} cannot depend on the holomorphic antighosts b and β . Ordinary functions of b and β would make the holomorphic ghost number negative, and delta functions of β would make the picture number positive.

For the antiholomorphic degrees of freedom, we can make a shorter version of the same argument. If when we insert \mathcal{V} in the path integral, we can insert

an extra factor of $\delta(\tilde{b})$ with the path integral remaining non-zero, then \mathcal{V} must have antiholomorphic ghost number 1. Given that \mathcal{V} does not depend on the derivatives of \tilde{c} , it must be precisely proportional to \tilde{c} , with no dependence on \tilde{b} .

The conclusion then is that an NS vertex operator that will lead to a simple superconformal formalism has the form

$$(4.6) \quad \mathcal{V} = \tilde{c}\delta(\gamma)V,$$

where V is constructed from matter fields only. In fact, as we will discuss, such an operator is Q_B -invariant if and only if V is a superconformal primary of antiholomorphic and holomorphic dimensions $(1, 1/2)$.

The need for the factor $\delta(\gamma)$ is perhaps daunting, though by now this is an old story, going back to [1, 7]. But we should not be faint of heart; we know from section 3 that operators of the general form $\delta(\beta)$ are unavoidable, so we should be prepared to deal with $\delta(\gamma)$ as well.

Actually, the factor $c\delta(\gamma)$ has an intuitive explanation. As in eqn. (2.62), we can write c as $\delta(c)$ so $c\delta(\gamma)$ becomes $\delta(c)\delta(\gamma)$. The ghosts represent symmetry generators, but in the presence of a vertex operator, only superconformal vector fields that vanish at the position of the vertex operator are symmetries. So c and γ should both be constrained to vanish at the position of the vertex operator, and this is achieved via the delta functions.³⁰

Finally, we have to impose the condition of BRST invariance. A vertex operator of the form $\mathcal{V} = \tilde{c}\delta(\gamma)V$ is Q_B -invariant if and only if V is a conformal and superconformal primary (for antiholomorphic and holomorphic variables, respectively) of dimension $(1, 1/2)$. Thus, V must obey

$$(4.7) \quad \begin{aligned} L_n^X V &= \frac{1}{2}\delta_{n,0}V, \quad n \geq 0 \\ G_r^X V &= 0, \quad r > 0 \\ \tilde{L}_n^X V &= \delta_{n,0}V, \quad n \geq 0, \end{aligned}$$

where the operators on the left are Virasoro and super Virasoro generators of the matter system. The ghost factor $\tilde{c}\delta(\gamma)$ is a primary of conformal dimension $(-1, -1/2)$, so V being a primary of dimension $(1, 1/2)$ ensures that \mathcal{V} is

³⁰To elaborate on this slightly, let us go back to the formula (3.3) for general superconformal vector fields. The condition that ν_f and V_g both leave fixed the point $z = \theta = 0$ is that $f(0) = g(0) = 0$. The fields γ and c are ghosts that correspond to the vector fields f and g , so we want $\gamma(0) = c(0) = 0$. This is enforced by the delta functions.

a primary of dimension $(0, 0)$. We refer to such a \mathcal{V} as a superconformal vertex operator. According to the BRST version of the no-ghost theorem, every BRST cohomology class (at non-zero momentum) contains a representative that is a superconformal vertex operator, so in particular such operators suffice for computing the S -matrix.

Let us briefly explain the role of the $\delta(\gamma)$ factor in \mathcal{V} , since this is the most unusual feature. In verifying that $Q_B \mathcal{V} = 0$ for \mathcal{V} a superconformal vertex operator, one needs among other things to verify that \mathcal{V} is annihilated by the following part of Q_B :

$$(4.8) \quad Q_B^* = \sum_{r \in \mathbb{Z} + 1/2} \gamma_r G_{-r}^X.$$

Here γ_r are the modes

$$(4.9) \quad \gamma(z) = \sum_{r \in \mathbb{Z} + 1/2} z^{-r+1/2} \gamma_r.$$

The conditions (4.7) ensure that $\gamma_r G_{-r}^X \mathcal{V} = 0$ for $r < 0$. For $r > 0$, we see from (4.9) that vanishing of γ_r for r positive means that $\gamma(z)$ vanishes at $z = 0$. In the presence of the identity operator, $\gamma(z)$ would be regular but non-vanishing at $z = 0$; to make γ vanish at $z = 0$, we need to insert an operator $\delta(\gamma(0))$. That is why $\mathcal{V}(z)$ is proportional to $\delta(\gamma(z))$.

For our applications, we need to know that the operator $\mathcal{V} = \tilde{c} \delta(\gamma) V$ is Q_B -invariant not just in the ordinary sense but also in the extended sense in which Q_B acts on \mathcal{J} . This is true by the same sort of argument given at the end of section 2.4.1, using the antighost equation of motion derived from the extended action (3.24). The same goes for the superconformal Ramond vertex operators that we will study in section 5.

4.1.1. Other picture numbers? In conventional language, we seem to have arrived at a unique choice for the picture numbers of the NS sector vertex operators: because of the factor of $\delta(\gamma)$, they have picture number -1 . (This is the picture number of the unintegrated form of the vertex operator; passing to the integrated form of the vertex operator maps the picture number to 0, as we describe later.) However, this resulted from a possibly innocent-looking assertion that we made at the beginning: we claimed that adding an NS puncture increases the holomorphic dimension of the moduli space of super Riemann surfaces by 1|1. This is the right answer for the conventional definition of the moduli space of super Riemann surfaces with NS punctures. Actually, as explained in section 4.3 of [16], one can modify the definition

of supermoduli space in a way that increases its odd dimension, and then one can calculate using NS vertex operators of any desired integer picture number $k \leq -1$. (One can choose the picture number of each NS vertex operator independently, as long as they are all no greater than -1 .) The main property of this more general formalism is that it can be reduced to the standard one by integrating out the extra odd moduli. We will not pursue the generalization, which seems to have no benefit, except to answer the question, “Is there a natural way to calculate scattering amplitudes using NS vertex operators of picture number different from -1 ?”

There seems to be no natural way to compute scattering amplitudes, in general, using (unintegrated) NS vertex operators of picture number greater than -1 . To do this, one would want a version of supermoduli space with punctures with an odd dimension less than the usual value, and this appears not to exist. In genus 0, one can integrate over odd moduli with the help of picture-changing operators, and then, after placing the picture-changing operators at the positions of the external vertex operators, one can compute scattering amplitudes with vertex operators of arbitrary picture number (with the right overall sum); see eqn. (4.31). In higher genus, such an approach will lead to the spurious singularities described in section 3, and also to difficulties in understanding massless tadpoles, as we describe starting in section 6.

In section 5, we introduce Ramond vertex operators, and there a similar story holds. The simplest procedure is based on Ramond vertex operators of picture number $-1/2$, but if one wishes, one can modify the definition of supermoduli space and compute with vertex operators of any picture number less than $-1/2$.

4.2. Scattering amplitudes for NS states

Now we would like to explain how to calculate amplitudes for the scattering of n string states, all in the NS sector, described by superconformal vertex operators $\mathcal{V}_1, \dots, \mathcal{V}_n$. The derivations in sections 2 and 3 were phrased so as to generalize straightforwardly, so we can be relatively brief.

Combining the definitions of sections 2.4 and 3.2, we define a form $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J})$ on the space \mathcal{J} of supercomplex structures:

$$(4.10) \quad F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J}) = \int \mathcal{D}(X, B, C, \tilde{B}, \tilde{C}) \exp(-\hat{I}) \prod_{i=1}^n \mathcal{V}_i(p_i),$$

with p_i being points in Σ . This definition makes sense for vertex operators that are not necessarily superconformal, though it is particularly useful in the superconformal case.

If $\mathcal{V}_1 = \{Q_B, \mathcal{W}_1\}$ and the other \mathcal{V}_i are all Q_B -invariant, then

$$(4.11) \quad F_{\{Q_B, \mathcal{W}_1\}, \mathcal{V}_2, \dots, \mathcal{V}_n} + dF_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n} = 0,$$

just as in eqn. (2.47). This relation is the basis for the proof of gauge-invariance, that is, the decoupling of BRST-trivial states from the scattering amplitudes.

As we will explain momentarily, superconformal symmetry can be used in a simple way to analyze the scattering amplitudes only if the vertex operators \mathcal{V}_i obey

$$(4.12) \quad \tilde{b}_n \mathcal{V} = b_n \mathcal{V} = \beta_r \mathcal{V} = 0, \quad n, r \geq 0.$$

To prove gauge-invariance in a superconformally invariant framework, one needs to know that if \mathcal{V} obeys (4.12) and can be written as $\{Q_B, \mathcal{W}\}$ for some \mathcal{W} , then \mathcal{W} can be chosen to obey the same constraints:

$$(4.13) \quad \tilde{b}_n \mathcal{W} = b_n \mathcal{W} = \beta_r \mathcal{W} = 0, \quad n, r \geq 0.$$

This is shown in appendix C, by adapting the corresponding bosonic argument of appendix B.

To reduce to computations on a finite-dimensional moduli space, one would like to know that $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J})$ is a pullback from the quotient $\mathcal{J}/\mathcal{D}_{p_1, \dots, p_n}$, where $\mathcal{D}_{p_1, \dots, p_n}$ is the group of diffeomorphisms of Σ that leaves fixed the punctures at p_1, \dots, p_n . For this, we need to know the usual two facts: (1) $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J})$ must be $\mathcal{D}_{p_1, \dots, p_n}$ -invariant, which is obvious from the definition; (2) $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J})$ must vanish if contracted with a vector field on \mathcal{J} that is a generator of $\mathcal{D}_{p_1, \dots, p_n}$. The explanation of this is very similar to what it was in section 2.4.2. Contraction with a vector field $q^z(\tilde{z}; z|\theta)\partial_z + \frac{1}{2}D_\theta q^z D_\theta$ on Σ will generate a shift $\delta\mathcal{J}_z^z \rightarrow \delta\mathcal{J}_z^z + \partial_z q^z$. This shifts the extended action (3.24) by

$$(4.14) \quad \hat{I} \rightarrow \hat{I} + \frac{1}{2\pi} \int_\Sigma \mathcal{D}(\tilde{z}, z|\theta) \partial_z q^z B_{z\theta}.$$

Looking back to the ghost action (3.20), we see that we can restore the invariance of the extended action by shifting C^z by $C^z \rightarrow C^z - q^z$. But we need to know that the vertex operators $\mathcal{V}_1, \dots, \mathcal{V}_n$ are invariant under $C^z \rightarrow C^z - q^z$. The only constraint on q^z is that $q^z = D_\theta q^z = 0$ at the points p_i (so that the vector field $q^z\partial_z + \frac{1}{2}D_\theta q^z D_\theta$ leaves those points fixed). There is no constraint on the derivatives of q^z with respect to z . Accordingly, the condition

for $\mathcal{V}_1(p_1), \dots, \mathcal{V}_n(p_n)$ to be invariant under $C^z \rightarrow C^z - q^z$ is that, although $\mathcal{V}_1, \dots, \mathcal{V}_n$ may depend on the ghost fields c and γ at the points p^i , they do not depend on the derivatives of those ghost fields with respect to z . In other words, the condition that we need is precisely that $b_n \mathcal{V} = \beta_r \mathcal{V} = 0$, $n, r \geq 0$. The remaining condition $\tilde{b}_n \mathcal{V} = 0$, $n \geq 0$, arises by considering in a similar way vector fields $\tilde{q}^z(\tilde{z}; z|\theta)\partial_{\tilde{z}}$. (As remarked in relation to eqn. (3.9), vector fields $q^\theta D_\theta$ have already been used to eliminate the possibility of deforming the superconformal structure of Σ independent of its complex structure, so we need not consider such vector fields now.)

The superdegree of the form $F(\mathcal{J}, \delta\mathcal{J})$ can be determined by asking what generalization of the delta function insertions of eqn. (3.66) is needed to get a sensible and non-zero path integral for the matter fields. Given the way that we determined what class of vertex operators we wanted, there is not much to say here: for each superconformal vertex operator \mathcal{V} , we need one extra $\delta(\tilde{B})$ insertion and $1|1$ additional $\delta(B)$ insertions. So with n external superconformal NS vertex operators, $F(\mathcal{J}, \delta\mathcal{J})$ has antiholomorphic superdegree $3g - 3 + n|0$ and holomorphic superdegree $3g - 3 + n|2g - 2 + n$.

So the superdegree of $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J})$ is that of a form that can be integrated over a supermanifold of dimension $6g - 6 + 2n|2g - 2 + n$. The right supermanifold is roughly speaking the quotient $\mathcal{J}/\mathcal{D}_{p_1, \dots, p_n}$, from which $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J})$ is a pullback; and roughly speaking this is the moduli space of super Riemann surfaces of genus g with n NS punctures. Actually, we meet here the same subtlety as in section 3.3.2, and we will proceed as we did there.

We use the real analyticity of the worldsheet path integral to regard the moduli \tilde{m}_β of left-movers and m_α, η_σ of right-movers as independent complex variables. Then $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J})$ can be interpreted as a form³¹ on a product $\mathcal{M}_L \times \mathcal{M}_R$, where \mathcal{M}_L and \mathcal{M}_R parametrize respectively the antiholomorphic structure and the holomorphic structure of Σ . This structure holds for all of the oriented closed string theories; for the heterotic string, \mathcal{M}_L is a copy of $\mathcal{M}_{g,n}$, the moduli space of Riemann surfaces of genus g with n punctures, while \mathcal{M}_R is a copy of $\mathfrak{M}_{g,n}$, the moduli space of super Riemann surfaces of genus g with n NS punctures. Inside the reduced space $(\mathcal{M}_L \times \mathcal{M}_R)_{\text{red}}$, there is a “diagonal” Γ characterized by saying that, when one reduces modulo the odd variables, Σ is real – its holomorphic and antiholomorphic structures are complex conjugates. With a suitable choice of coordinates, this means that $\tilde{m}_\alpha = \bar{m}_\alpha$. Letting the odd moduli vary, we “thicken” Γ slightly to a cycle $\Gamma \subset \mathcal{M}_L \times \mathcal{M}_R$ of the same codimension. Then Γ is the integration

³¹To be more precise, $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J})$ is the pullback of a section of $Ber(\mathcal{M}_L \times \mathcal{M}_R)$ that is defined and holomorphic on a neighborhood of the “diagonal” Γ that is introduced shortly.

cycle of the heterotic string with external NS vertex operators. The genus g contribution to the scattering amplitude is

$$(4.15) \quad \langle \mathcal{V}_1 \dots \mathcal{V}_n \rangle = \int_{\Gamma} F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J}).$$

We can make this formula concrete by the same methods of integrating over a slice and/or integrating out odd moduli as in section 3.

The formula (4.11) for gauge invariance becomes a statement about integration over Γ :

$$(4.16) \quad \int_{\Gamma} F_{\{Q_B, \mathcal{W}_1\}, \mathcal{V}_2, \dots, \mathcal{V}_n} = - \int_{\Gamma} dF_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n} = 0.$$

The supermanifold version of Stokes's theorem can be applied to the right hand side, so as for bosonic strings, the only potential obstruction to gauge-invariance is the possible occurrence of a surface term at infinity. We return to this after analyzing the basic structure at infinity in section 6.

The topic of this kind that perhaps requires more detail is the superstring analog of an integrated vertex operator. We turn to this next.

4.3. Integrated NS vertex operators

For Neveu-Schwarz vertex operators (but not for Ramond vertex operators), there is a superconformally invariant notion of an integrated vertex operator, analogous to what we explained in section 2.5 for the bosonic string. Let $\mathcal{V} = \tilde{c}c\delta(\gamma)V$, where V is a superconformal primary of dimension $(1, 1/2)$ constructed from the matter system only. This is the right dimension so that V can be integrated over Σ in a natural way. Thus we can define

$$(4.17) \quad K_{\mathcal{V}} = \int_{\Sigma} \mathcal{D}(\tilde{z}, z|\theta) V(\tilde{z}; z|\theta),$$

and instead of inserting \mathcal{V} at a point in Σ , we can insert a factor of $K_{\mathcal{V}}$ in the worldsheet path integral. For many purposes, it is equivalent to insert an NS vertex operator on Σ in its integrated or unintegrated form.

We can relate the two types of vertex operator by adapting the derivation of section 2.5.2. We compare two different descriptions (*i*) and (*ii*). In description (*i*), we use local coordinates $\tilde{w}; w|\zeta$ on Σ . The vertex operator $\mathcal{V} = \tilde{c}c\delta(\gamma)V$ is inserted at $\tilde{w} = w = \zeta = 0$. The complex structure \mathcal{J} depends on moduli $\tilde{\mathbf{z}}; \mathbf{z}|\boldsymbol{\theta}$ (as well as other moduli that are not relevant for this

discussion). In description (ii), Σ is described by local holomorphic coordinates $\tilde{z}; z|\theta$. The complex structure \mathcal{J} does not depend on the moduli $\tilde{\mathbf{z}}; \mathbf{z}|\theta$, but the vertex operator \mathcal{V} is inserted at $\tilde{z}; z|\theta = \tilde{\mathbf{z}}; \mathbf{z}|\theta$.

In section 2.5.2, the key point in the derivation was the identity

$$(4.18) \quad \left(\frac{1}{2\pi} \int_{\Sigma} d^2z g^z \partial_{\tilde{z}} b_{zz} \right) \cdot c^z(0) = g^z(0).$$

This identity – and the corresponding one for antiholomorphic variables – was used to remove the ghost factors $\tilde{c}c$ from the vertex operator and replace them by an integration measure $d\tilde{\mathbf{z}} d\mathbf{z}$.

For the superstring case, certainly we are going to need a superanalog of eqn. (4.18). We can rewrite (4.18) in terms of delta functions:

$$(4.19) \quad \delta \left(\frac{1}{2\pi} \int_{\Sigma} d^2z g^z \partial_{\tilde{z}} b_{zz} \right) \cdot \delta(c^z(0)) = g^z(0).$$

In this form, it is easy to guess the superanalog:

$$(4.20) \quad \delta \left(\frac{1}{2\pi} \int_{\Sigma} d^2z f^{\theta} \partial_{\tilde{z}} \beta_{z\theta} \right) \cdot \delta(\gamma^{\theta}(0)) = \frac{1}{f^{\theta}(0)}.$$

This formula will be justified in section 10 (see eqn. (10.48)).

With the help of this identity, one proceeds just as in section 2.5.2. We start in description (i), which fits the framework of section 4.2. In this framework, the integral over $\tilde{\mathbf{z}}, \mathbf{z}$, and θ is made by inserting suitable delta functions defined in eqns. (3.64) and (3.66). We write schematically

$$(4.21) \quad \delta(\tilde{B}(\tilde{\mathbf{z}})) \delta(B(\mathbf{z})) \delta(B(\theta))$$

for these delta functions. In (3.64), we have explicit formulas for them in terms of the derivatives of \mathcal{J} with respect to $\tilde{\mathbf{z}}, \mathbf{z}$, and θ . These derivatives are very simple; since the dependence of \mathcal{J} on $\tilde{\mathbf{z}}, \mathbf{z}$, and θ is induced from a diffeomorphism, we get formulas such as $\partial_{\mathbf{z}} \mathcal{J}_w^w = \partial_w v^w$, for some vector field $v^w \partial_w$, and similarly for $\tilde{\mathbf{z}}$ and θ . Inserting $\partial_{\mathbf{z}} \delta \mathcal{J}_w^w = \partial_w v^w$ in (3.64) and integrating by parts, the delta functions (4.21) have arguments that naively vanish by the equations of motion $\partial_w B_{w\zeta} = D_{\zeta} \tilde{B}_{\tilde{w}\tilde{w}} = 0$. So we get the opportunity to use the identities in (4.19) and (4.20) (and the antiholomorphic counterpart of (4.19)). These identities can be used to remove the delta functions while also removing the factors of $\tilde{c}c\delta(\gamma)$ from the vertex operator. The net effect

of this and the change of coordinates to description (ii) is to transform the insertion

$$(4.22) \quad [d\tilde{\mathbf{z}};d\mathbf{z}|d\boldsymbol{\theta}] \delta(\tilde{B}(\tilde{\mathbf{z}}))\delta(B(\mathbf{z}))\delta(B(\boldsymbol{\theta}))\tilde{c}c\delta(\gamma)V,$$

which represents the general procedure for integration over $\tilde{\mathbf{z}};\mathbf{z}|\boldsymbol{\theta}$ with an insertion of $\mathcal{V} = \tilde{c}c\delta(\gamma)V$, into simply

$$(4.23) \quad [d\tilde{\mathbf{z}};d\mathbf{z}|d\boldsymbol{\theta}] V(\tilde{\mathbf{z}};\mathbf{z}|\boldsymbol{\theta}).$$

This is the description in terms of an integrated vertex operator. The steps that we have omitted are quite similar to those in section 2.5.2.

Just as for bosonic strings, the equivalence between integrated and unintegrated vertex operators is not quite valid at infinity in moduli space (see fig. 1 of section 2.5.1). The description by unintegrated vertex operators is always valid, while the description by integrated vertex operators can lead to difficulty.

4.3.1. More on gauge invariance The remarks of section 2.5.3 have a close analog for Neveu-Schwarz vertex operators of superstring theory. The analogy is so close that we will be brief.

If a superconformal primary field V of the matter system of dimension $(1, 1/2)$ is also a descendant (in which case it is a null vector), then $\mathcal{V} = \tilde{c}c\delta(\gamma)V$ is BRST-trivial and should decouple. In the case of massless string states, this has a natural explanation by integration by parts on Σ . If (in the heterotic string) V is a massless vertex operator that is also null, then $V = G_{-1/2}W$ where W is a primary of dimension $(1, 0)$, or $V = \tilde{L}_{-1}W'$, where W' is of conformal dimension $(0, 1/2)$. In these cases, the integrated vertex operator insertion is

$$(4.24) \quad \int_{\Sigma} [d\tilde{\mathbf{z}};d\mathbf{z}|d\boldsymbol{\theta}] V = \begin{cases} \int_{\Sigma} [d\tilde{\mathbf{z}};d\mathbf{z}|d\boldsymbol{\theta}] D_{\theta}W \\ \int_{\Sigma} [d\tilde{\mathbf{z}};d\mathbf{z}|d\boldsymbol{\theta}] \partial_{\tilde{z}}W', \end{cases}$$

and vanishes by integration by parts. (To understand how to integrate by parts in the case of $D_{\theta}W$, see section 2.4.1 of [16].)

If instead V is a null vector corresponding to a massive string state, then just as in section 2.5.3, \mathcal{V} is not a total derivative on Σ . Its decoupling has to be proved by using (4.16) and integrating by parts on Γ , not just by integrating by parts on Σ .

The first few examples of Neveu-Schwarz gauge parameters are described in appendix C.

4.4. Tree-level scattering

Having understood the integrated version of an NS vertex operator, it is straightforward to compute tree amplitudes.

Let us first recall how this is done for the bosonic string. We begin with n vertex operators $\mathcal{V}_i = \tilde{c}V_i$, inserted at prescribed points in a genus 0 Riemann surface Σ_0 . We can identify Σ_0 as the complex z -plane together with a point at infinity. We use the $SL(2, \mathbb{C})$ symmetry of Σ_0 to map three of the \mathcal{V}_i , say the first three, to arbitrary points $z_i \in \mathbb{C}$. For the moduli of Σ_0 , we can take the positions z_4, \dots, z_n of the other vertex operators. As in section 2.5, to integrate over those positions, we simply replace the unintegrated vertex operators \mathcal{V}_i by their integrated counterparts $\int d^2z_i V_i(\tilde{z}_i, z_i)$. The scattering amplitude is then

$$(4.25) \quad \left\langle \prod_{i=1}^3 \tilde{c}V_i(\tilde{z}_i, z_i) \prod_{j=4}^n \int d^2z_j V(\tilde{z}_j, z_j) \right\rangle.$$

To extend this to superstring theory with NS vertex operators, the main subtlety arises for the three-point function. We start with three unintegrated vertex operators. In the heterotic string, they take the form $\mathcal{V}_i = \tilde{c}\delta(\gamma)V_i$. The moduli space $\mathcal{M}_{0,3}$ of an ordinary Riemann surface of genus 0 with three punctures is a point, but the supermoduli space $\mathfrak{M}_{0,3}$ of a super Riemann surface of genus 0 with three NS punctures has dimension $0|1$. So in computing the heterotic string NS three-point function in genus zero, there are no antiholomorphic moduli and $0|1$ holomorphic ones. So there are no $\delta(\tilde{b})$ or $\delta(b)$ insertions, and one $\delta(\beta)$ insertion. Via eqn. (4.20), we use the $\delta(\beta)$ insertion to remove the $\delta(\gamma)$ from any one of the three vertex operators, say \mathcal{V}_3 . We still have to integrate over one odd modulus, which we take to be the odd coordinate θ_3 of \mathcal{V}_3 . The integral is easily evaluated:

$$(4.26) \quad \int d\theta_3 \tilde{c}V_3(\tilde{z}_3; z_3 | \theta_3) = \tilde{c}D_\theta V_3(\tilde{z}_3; z_3 | 0).$$

The NS three-point function in genus 0 is therefore

$$(4.27) \quad \left\langle \tilde{c}c\delta(\gamma)V_1(\tilde{z}_1, z_1 | 0) \tilde{c}c\delta(\gamma)V_2(\tilde{z}_2, z_2 | 0) \tilde{c}cD_\theta V_3(\tilde{z}_3, z_3 | 0) \right\rangle.$$

Here the z_i and \tilde{z}_i may be chosen arbitrarily.

When we add an additional NS vertex insertion, we gain $1|0$ antiholomorphic moduli and $1|1$ holomorphic moduli. The extra moduli are simply

the coordinates $\tilde{z}; z|\theta$ at which an additional NS vertex operator is inserted. For each added vertex operator $\mathcal{V}_j = \tilde{c}c\delta(\gamma)V_j$, we gain a full complement of delta functions $\delta(\tilde{b})\delta(b)\delta(\beta)$. As in section 4.3, we can use the delta functions to convert the unintegrated vertex operator \mathcal{V}_j into its integrated form $\int[d\tilde{z}_j; dz_j|d\theta_j]V_j(\tilde{z}_j; z_j|\theta_j)$, whereupon the scattering amplitude becomes

$$(4.28) \quad \left\langle \tilde{c}c\delta(\gamma)V_1(\tilde{z}_1, z_1|0) \tilde{c}c\delta(\gamma)V_2(\tilde{z}_2, z_2|0) \tilde{c}cD_\theta V_3(\tilde{z}_3, z_3|0) \right. \\ \left. \times \prod_{j=4}^n \int [d\tilde{z}_j; dz_j|d\theta_j]V_j(\tilde{z}_j; z_j|\theta_j) \right\rangle.$$

In this procedure for calculating the tree-level scattering, two vertex operators, namely \mathcal{V}_1 and \mathcal{V}_2 , have picture number -1 , while the others have picture number 0 . In section 4.5, we describe a more general formula for the tree-level S -matrix.

To arrive at (4.28), we have followed the usual procedure of integrating over some moduli by the use of integrated vertex operators. (Since we are in genus 0, we were able to integrate over all moduli this way.) This procedure treats the interior of $\mathfrak{M}_{0,n}$ correctly, but actually does not treat correctly the compactification. One reason for this has nothing to do with worldsheet supersymmetry and was described in fig. 1 of section 2.5.1. With super Riemann surfaces, there is actually a second, somewhat analogous problem. At infinity in $\mathfrak{M}_{0,n}$, Σ splits up into the union of two genus 0 components Σ_ℓ and Σ_r joined at a common singularity (fig. 2). When this happens, a correct procedure requires that each of Σ_ℓ and Σ_r should have two unintegrated vertex operators – counting the singularity as an unintegrated vertex operator. (We are here imposing the condition (3.107) on both Σ_ℓ and Σ_r . If the path integrals on Σ_ℓ and Σ_r are not separately sensible, we cannot expect the path integral on Σ to factor naturally when Σ degenerates. For more on such factorization, and also for the interpretation of the singularity as an unintegrated vertex operator, see section 6.) In eqn. (4.28), this condition is satisfied if some of the integrated vertex operators collide with one of the unintegrated ones, but not if they collide only with each other.

Even though (4.28) does not treat the compactification of $\mathfrak{M}_{0,n}$ correctly, it does give the right tree-level S -matrix. In general, a method of integration that treats the compactification incorrectly will differ from a correct treatment by surface terms at infinity.³² In the case of a genus 0 scattering amplitude, as long as the external momenta are sufficiently generic (so that the

³²This follows from the analysis in section 3.4.1. Any two methods of integrating

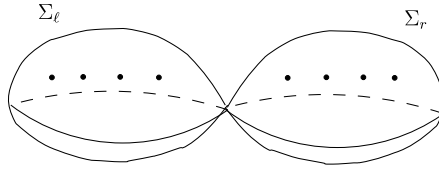


Figure 2: At infinity in moduli space, a genus 0 surface Σ splits into two genus zero components Σ_ℓ and Σ_r , joined at a common singularity. The external vertex operators are distributed between the two sides in an arbitrary fashion. In the case of external NS vertex operators, to treat the compactification of moduli space correctly, both Σ_ℓ and Σ_r should have two unintegrated vertex operators – counting the singularity as one. Regardless of which two of the original vertex operators we take in unintegrated form, this is not always the case, since they might be both on Σ_ℓ or both on Σ_r . So the formalism based on integrated vertex operators does not treat correctly the compactification of the moduli space. In the more general formalism of section 4.5 based on vertex operators of general picture numbers, one has the same problem. One would like the condition (4.32) to hold on each side; but no choice of picture numbers ensures this.

momentum flowing through the singularity is off-shell), the relevant integral is highly convergent at infinity (or more precisely the calculation is made by analytic continuation from a region of external momenta where the integral is convergent) and there is no surface term. In higher genus, a formalism that treats the region at infinity incorrectly will lead to difficulty in understanding massless tadpoles and mass renormalization, but these are not problems in genus 0.

4.5. Integration over odd moduli

The procedure introduced in sections 3.5 and 3.6 for integrating over odd moduli can straightforwardly be extended to include NS vertex operators. We start again with a reduced Riemann surface Σ_{red} with spin bundle $K^{1/2}$; by adding an odd coordinate θ , we build a super Riemann surface Σ . We insert NS vertex operators at n points $q_1, \dots, q_n \in \Sigma_{\text{red}}$ (and thus at $\theta = 0$).

over odd variables differ by a total derivative, so in particular a treatment that is correct except at infinity differs from a treatment that is correct everywhere by a surface term at infinity.

In this framework, odd moduli are incorporated by turning on a gravitino field. We expand the gravitino field as in (3.78),

$$(4.29) \quad \chi_z^\theta = \sum_{\sigma=1}^{2g-2+n} \eta_\sigma \chi_z^{(\sigma)\theta}.$$

We can still make a gauge transformation $\chi_z^\theta \rightarrow \chi_z^\theta + \partial_z y^\theta$. But now y^θ should vanish at q_1, \dots, q_n , so the gravitino modes should be understood as elements of $H^1(\Sigma_{\text{red}}, T^{1/2} \otimes \mathcal{O}(-\sum_{i=1}^n q_i))$. This space has dimension $2g - 2 + n$ (which therefore is the odd dimension of the supermoduli space $\mathfrak{M}_{g,n}$ of genus g surfaces with n NS punctures), and that is the number of gravitino modes that we need. A spurious singularity will arise if the $\chi_z^{(\sigma)\theta}$ are not linearly independent as elements of $H^1(\Sigma_{\text{red}}, T^{1/2} \otimes \mathcal{O}(-\sum_i q_i))$.

The general procedure for integrating out the odd moduli is precisely as described in section 3.6.1, and leads to the insertions described in eqns. (3.88) or (3.89). If we wish, then as in section 3.6.2, we can take the gravitino modes to have delta function support at points $p_\sigma \in \Sigma_{\text{red}}$, $\sigma = 1, \dots, 2g - 2 + n$. The necessary insertions for integrating out odd moduli are then simply a product of picture-changing operators, $\prod_{\sigma=1}^{2g-2+n} \mathfrak{Y}(p_\sigma)$, where $\mathfrak{Y}(p_\sigma)$ was defined in (3.95). The condition (3.99) for a spurious singularity becomes

$$(4.30) \quad H^0\left(\Sigma_{\text{red}}, T^{1/2} \otimes \mathcal{O}\left(\sum_{\sigma=1}^{2g-2+n} p_\sigma - \sum_{i=1}^n q_i\right)\right) \neq 0.$$

The line bundle $\mathcal{L} = T^{1/2} \otimes \mathcal{O}(\sum_{\sigma=1}^{2g-2+n} p_\sigma - \sum_{i=1}^n q_i)$ has degree $g - 1$, and just as in the discussion of eqns. (3.98) and (3.99), this implies that $H^0(\Sigma_{\text{red}}, \mathcal{L})$ generically vanishes, but can generically become non-zero as one varies one complex parameter. For example, the parameter in question could be the choice of one of the points q_i at which a vertex operator is inserted, or one of the points p_σ at which a picture-changing operator is inserted. When $H^0(\Sigma_{\text{red}}, \mathcal{L}) \neq 0$, the gauge-fixing procedure is incorrect, and the integration measure computed with this procedure acquires a pole.

There is one important situation in which spurious poles do not arise. This is the case that Σ has genus zero. For $g = 0$, the line bundle \mathcal{L} has degree -1 ; a line bundle of this degree is unique up to isomorphism, and its sheaf cohomology vanishes.

Even for $g = 0$, we still have to decide where to insert the picture-changing operators. If the p_σ do not vary holomorphically with the q_i , the resulting formulas will be inelegant, though free of spurious singularities (and correct

as long as the p_σ do not vary too wildly near infinity in $\mathfrak{M}_{0,n}$). The usual procedure [1] to insure holomorphy and minimize extraneous choices is to take the p_σ to coincide with some of the q_i . Concretely, one picks nonnegative integers r_i , $i = 1, \dots, n$ with $\sum_i r_i = n - 2$, and for each i one takes r_i of the p_σ to coincide³³ with q_i . The vertex operators $\mathcal{V}_i = \tilde{c}c\delta(\gamma)V_i$ are thereby transformed into vertex operators $\mathcal{V}_i^{(s_i)} = \tilde{c}c\widehat{V}_i^{(s_i)}$ of picture number $s_i = r_i - 1$. (One can show that $\widehat{V}_i^{(0)} = D_\theta V_i$. Explicit formulas for the $\widehat{V}_i^{(s_i)}$ quickly become complicated for $s_i > 0$.) One still has to integrate over the q_i . This can be done just as in bosonic string theory by removing the factors $\tilde{c}c$ from $n - 3$ of the vertex operators and integrating over their positions. The scattering amplitude is thus

$$(4.31) \quad \left\langle \prod_{i=1}^3 \tilde{c}c\mathcal{V}_i^{(s_i)}(\tilde{z}_i, z_i) \prod_{j=4}^n \int d^2z_j \widehat{V}_j^{(s_j)}(\tilde{z}_j, z_j) \right\rangle.$$

We have fixed in an arbitrary way the positions of the first three vertex operators and integrated over the others.

In this fashion, we may compute tree amplitudes using NS vertex operators of any pictures³⁴ $s_i \geq -1$, constrained only by

$$(4.32) \quad \sum_i s_i = -2.$$

What we have just described is a generalization of the procedure of section 4.4, in which two unintegrated vertex operators had picture -1 and all others had picture 0 .

An important detail is the following. The absence of spurious singularities means that the procedure for gauge-fixing and integration over fermionic moduli is correct in the interior of $\mathfrak{M}_{0,n}$. It does not guarantee that this procedure is correct at infinity. In fact, regardless of what one chooses for the s_i , the formula (4.31) does not treat correctly the compactification of $\mathfrak{M}_{0,n}$. This assertion is a generalization of what was explained at the end of section 4.4. To treat properly the compactification, when Σ breaks into a union of two components Σ_ℓ and Σ_r , joined at a singularity (as in fig. 2 of section 4.4), one needs a condition analogous to (4.32) on each of Σ_ℓ and Σ_r separately,

³³We have understood in section 3.6.2 that there is no trouble letting several of the p_σ coincide.

³⁴One may also remove the restriction to $s_i \geq -1$ by modifying the definition of $\mathfrak{M}_{0,n}$ so that more than one odd modulus is associated to each puncture, as in section 4.3 of [16].

with the singularity considered to contribute $s = -1$ on each branch. (This statement amounts to imposing eqn. (3.107) on Σ_ℓ and Σ_r separately, and is a special case of what we will explain in section 6.4.6.) Moreover, this should be the case irrespective of how the points q_1, \dots, q_n are distributed between Σ_ℓ and Σ_r . No choice of the s_i obeys these conditions, so regardless of what pictures one chooses, the picture-changing formalism always treats incorrectly the compactification of the supermoduli space.

Nevertheless, for the same reason as in section 4.4, the picture-changing formalism computes tree amplitudes correctly.

5. The Ramond sector

5.1. Ramond insertion points and superconformal algebra

A Neveu-Schwarz vertex operator is inserted at a generic point on a super Riemann surface Σ . By contrast, a Ramond vertex operator is inserted at a singularity of the superconformal structure of Σ . (See for example section 4.1 of [16].) In the presence of a Ramond vertex operator, Σ is still from a holomorphic point of view a complex supermanifold of dimension $1|1$, with local complex coordinates $z|\theta$. Moreover, the holomorphic tangent bundle $T_R\Sigma$ of Σ still has a distinguished subbundle \mathcal{D} . But it is no longer true that it is possible everywhere to pick local superconformal coordinates $\tau|\zeta$ so that \mathcal{D} has a section of the form $D_\zeta = \partial_\zeta + \zeta\partial_\tau$. Rather the local form near a Ramond puncture at $z = 0$ is that \mathcal{D} is generated by

$$(5.1) \quad D_\theta^* = \partial_\theta + \theta z \partial_z.$$

For $z \neq 0$, we can reduce to the superconformal form by setting

$$(5.2) \quad \tau|\zeta = (\log z)|\theta,$$

whence $D_\theta^* = D_\zeta$. But at $z = 0$, no change of coordinates will put D_θ^* (or any other section $f(z|\theta)D_\theta^*$ of \mathcal{D}) in the superconformal form $\partial_\zeta + \zeta\partial_\tau$. The obstruction is that

$$(5.3) \quad (D_\theta^*)^2 = z\partial_z,$$

which vanishes at $z = 0$. By contrast, if a change of coordinates could put D_θ^* in the form $f(\tau|\zeta)(\partial_\zeta + \zeta\partial_\tau)$ where inevitably the function $f(\tau|\zeta)$ would be everywhere non-zero (since D_θ^* is), then D_θ^* and $(D_\theta^*)^2$ would everywhere be linearly independent.

Where D_θ^* and $(D_\theta^*)^2$ are linearly dependent, we say that the superconformal structure of Σ is singular. In our example, this occurs precisely for $z = 0$. The locus $z = 0$ defines not a point but a divisor $\mathcal{F} \subset \Sigma$, of dimension $0|1$. We call a divisor of this kind a Ramond divisor or Ramond puncture.

Now let us discuss superconformal symmetry in the presence of a Ramond puncture. A superconformal vector field W is a vector field that preserves the subbundle $\mathcal{D} \subset T_R \Sigma$; concretely, this means that the commutator of W with D_θ^* is a $(z|\theta$ -dependent) multiple of D_θ^* . A little calculation reveals that odd and even superconformal vector fields take the form

$$(5.4) \quad \begin{aligned} \nu_f &= f(z)(\partial_\theta - \theta z \partial_z) \\ V_g &= z \left(g(z) \partial_z + \frac{g'(z)}{2} \theta \partial_\theta \right). \end{aligned}$$

In particular, a basis of superconformal vector fields that are holomorphic except possibly for a pole at $z = 0$ is given by

$$(5.5) \quad \begin{aligned} G_r &= z^r (\partial_\theta - \theta z \partial_z) \\ L_n &= -z^{n+1} \partial_z - \frac{nz^n}{2} \theta \partial_\theta, \end{aligned}$$

with $n, r \in \mathbb{Z}$. A short calculation reveals that

$$(5.6) \quad \begin{aligned} [L_m, L_n] &= (m - n)L_{m+n} \\ \{G_r, G_s\} &= 2L_{r+s} \\ [L_m, G_r] &= \left(\frac{m}{2} - r \right) G_{m+r}. \end{aligned}$$

This is the super Virasoro algebra (3.6), but now in the Ramond sector, since the odd generators G_r have integral grading.

The subalgebra of the super Virasoro algebra that leaves fixed the divisor \mathcal{F} at $z = 0$ is generated by $G_r, L_n, r, n \geq 0$. In contrast to the other vector fields just mentioned, G_0 , when restricted to \mathcal{F} , is not zero:

$$(5.7) \quad G_0|_{z=0} = \partial_\theta.$$

Rather than vanishing when restricted to \mathcal{F} , G_0 generates a nontrivial automorphism of \mathcal{F} :

$$(5.8) \quad \theta \rightarrow \theta + \alpha,$$

with α an odd constant. Obviously, this symmetry does not leave \mathcal{F} fixed pointwise.

What are the simplest possible vertex operators? For the bosonic string and the NS sector of superstrings, the simplest vertex operators are what we have called conformal or superconformal vertex operators. They obey three conditions, beyond BRST-invariance:

1. They are invariant under conformal or superconformal transformations that leave fixed the point at which the vertex operator is inserted.
2. They also obey a more subtle condition of invariance under shifting the ghost fields by a conformal or superconformal vector field that leaves fixed the vertex operator insertions. (For NS vertex operators in superstring theory, this condition is stated in eqn. (4.12).)
3. They have the right ghost and (in the superstring case) picture numbers so that their insertion in the path integral gives a sensible and non-zero result.

Condition 1 lets us deduce from the worldsheet path integral with insertions of BRST-invariant vertex operators $\mathcal{V}_1, \dots, \mathcal{V}_n$ a closed form $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J})$ in a superconformally invariant fashion. Condition 2 enables one to prove that $F_{\mathcal{V}_1, \dots, \mathcal{V}_n}$ is a pullback from the appropriate finite-dimensional moduli space. Condition 3 then suffices to single out a very special class of vertex operators. (Condition 1 is a consequence of Condition 2 plus BRST symmetry, so one can consider Condition 2 to be more fundamental.)

For the Ramond sector, we want to follow the same pattern. But first we have to decide if a Ramond vertex operator should be inserted at a point in the divisor \mathcal{F} or should be associated to the whole divisor. Both choices are viable (see section 4.3 of [16]), but the simpler formalism – and the one that is compatible with using the standard definition of the moduli space of super Riemann surfaces with punctures – associates a Ramond vertex operator to the whole divisor \mathcal{F} , not to a particular point in it.

In this case, the constraints that we want are that the vertex operator \mathcal{V} should be invariant under all the super Virasoro generators that leave fixed the divisor \mathcal{F} :

$$(5.9) \quad G_r \mathcal{V} = L_n \mathcal{V} = 0, \quad r, n \geq 0.$$

Since G_0 does not leave fixed any point on \mathcal{F} , the G_0 constraint implies that \mathcal{V} is associated to the whole divisor \mathcal{F} , and not to a point in it.

As in our previous experience, the simplest vertex operators that can be used to compute the S -matrix are given by a standard ghost vertex operator

times a matter vertex operator that will be a superconformal primary of the appropriate dimension. More precisely, we will have

$$(5.10) \quad \mathcal{V} = \tilde{c}c\Theta V,$$

where \tilde{c} and c are the usual anticommuting ghosts, Θ is a special vertex operator of the $\beta\gamma$ system (first constructed in [1] and described below), and V is a vertex operator of the matter system. The matter vertex operator V will be a superconformal primary:

$$(5.11) \quad \begin{aligned} L_n^X V &= \frac{5}{8}\delta_{n,0}V, \quad n \geq 0 \\ G_r^X V &= 0, \quad r \geq 0. \end{aligned}$$

Here L_n^X and G_r^X are super-Virasoro generators of the matter system. In particular, V has holomorphic dimension $5/8$, which is the appropriate value because Θ has dimension $3/8$. (The value $5/8$ is the analog of dimension 1 for bosonic strings or dimension $1/2$ for NS vertex operators of superstring theory. Antiholomorphically, V will be a primary of dimension 1 and Θ will be the identity operator, of dimension 0.) The no-ghost theorem ensures that vertex operators of this kind, with Θ as described shortly, suffice for computing the S -matrix.

To obtain from BRST invariance the conditions (5.11), including the G_0 constraint, \mathcal{V} will have to be a vertex operator of picture number $-1/2$. Indeed, the mode expansion of the $\beta\gamma$ fields reads

$$(5.12) \quad \beta(z) = \sum_r z^{-r-3/2}\beta_r, \quad \gamma(z) = \sum_r z^{-r+1/2}\gamma_r,$$

where r takes integer values in the Ramond sector. (The meaning of these formulas for the Ramond sector in the context of super Riemann surfaces is described in section 5.3.) The BRST operator Q_B contains a term

$$(5.13) \quad Q_B^* = \sum_{r \in \mathbb{Z}} \gamma_{-r} G_r^X.$$

If we want Q_B -invariance of \mathcal{V} to give precisely the G_r^X constraints in (5.11), then Θ should be annihilated by γ_r precisely if $r > 0$. We will thus have

$$(5.14) \quad \gamma_r \Theta = 0, \quad r > 0.$$

On the other hand, Condition 2 above – stating that Θ should be invariant under shifts of γ by an odd superconformal vector field – means for the Ramond sector that

$$(5.15) \quad \beta_r \Theta = 0, \quad r \geq 0.$$

The constraints (5.14) and (5.15) together uniquely determine Θ . In conventional language, they mean that Θ corresponds to the $\beta\gamma$ vacuum of picture number $-1/2$. When we want to emphasize this, we denote it as $\Theta_{-1/2}$.

As in our study of bosonic strings or the NS sector of superstrings, one can organize this discussion more systematically as follows. Condition 2 requires eqn. (5.15). States of the $\beta\gamma$ system that obey (5.15) but not (5.14) can be obtained from $\Theta_{-1/2}$ by acting with operators β_{-m} or $\delta(\beta_{-m})$, $m \geq 1$; As we explain in the next paragraph, the part of \mathcal{V} constructed from the $\beta\gamma$ system should have ghost number and picture number $-1/2$; $\Theta_{-1/2}$ has these properties and acting with β_{-m} or $\delta(\beta_{-m})$ will either reduce its ghost number or increase its picture number. (That the ghost number of $\Theta_{-1/2}$ is $-1/2$ is a standard result [1], which we will explain in section 5.3.) So the $\beta\gamma$ part of the vertex operator must be $\Theta_{-1/2}$.

Now let $\mathfrak{M}_{g, n_{\text{NS}}, n_{\text{R}}}$ be the moduli space of super Riemann surfaces Σ of genus g with n_{NS} punctures and n_{R} Ramond punctures (that is, n_{R} divisors \mathcal{F}_i , $i = 1, \dots, n_{\text{R}}$ along which the superconformal structure has the sort of singularity described in eqn. (5.3)). For topological reasons, n_{R} is always even. As explained, for example, in [16], the dimension of $\mathfrak{M}_{g, n_{\text{NS}}, n_{\text{R}}}$ is

$$(5.16) \quad \dim \mathfrak{M}_{g, n_{\text{NS}}, n_{\text{R}}} = 3g - 3 + n_{\text{NS}} + n_{\text{R}} | 2g - 2 + n_{\text{NS}} + \frac{1}{2} n_{\text{R}}.$$

In particular, this dimension receives a contribution $1|\frac{1}{2}$ for each Ramond puncture. That is why the simplest formalism uses Ramond vertex operators of picture number $-1/2$. Adding n_{R} Ramond punctures increases the odd dimension of supermoduli space by $n_{\text{R}}/2$, so in the computation of a scattering amplitude, there will be an additional $n_{\text{R}}/2$ insertions of the general type $\delta(\beta)$, each of picture number 1. But as just sketched, a natural superconformal formalism (with Ramond insertions associated to divisors rather than points) requires each Ramond vertex operator to have picture number $-1/2$. The balance is preserved nicely: adding n_{R} Ramond punctures adds $n_{\text{R}}/2$ insertions $\delta(\beta)$, each of picture number 1, and n_{R} Ramond vertex operators, each of picture number $-1/2$. Since the operator $\delta(\beta)$ has ghost number 1, it follows also that Θ should have ghost number $-1/2$ and hence must obey (5.14) as well as (5.15).

5.1.1. The alternative The alternative to what we have described is to define vertex operators that are inserted at a point on a Ramond divisor, as explained in section 4.3 of [16]. In this case, the definition of supermoduli space and the dimension formula (5.16) are modified so that the contribution of a Ramond puncture to the odd dimension is $3/2$. Moreover, if a Ramond vertex operator is supposed to be inserted at a point, one must drop the G_0 constraint in (5.11), and accordingly one replaces (5.14) with

$$(5.17) \quad \begin{aligned} \gamma_r \Theta &= 0, & r \geq 0 \\ \beta_r \Theta &= 0, & r > 0. \end{aligned}$$

In this case, in conventional language, Θ has picture number $-3/2$, again matching the contribution of the Ramond puncture to the odd dimension of moduli space. (Its ghost number is $-3/2$, also balancing that of the relevant $\delta(\beta)$ insertions.)

More generally, as explained in [16], one can calculate with Ramond vertex operators with any picture number that is no greater than $-1/2$, if one suitably modifies the definition of supermoduli space. We will stick with the simplest procedure, based on the standard definition of the supermoduli space and Ramond vertex operators of picture number $-1/2$. There seems to be no systematic procedure to compute with Ramond vertex operators of picture number greater than $-1/2$, though this can certainly be done in genus 0, as explained in [1] and in section 5.4 below.

5.1.2. A pair of Ramond punctures Since it is an important point, we want to look in another way at the claim that superconformal symmetry in the neighborhood of a Ramond puncture leads naturally to the use of Ramond vertex operators of picture number $-1/2$.

First let us describe a general example with an arbitrary number of Ramond divisors. We take $\mathbb{C}^{1|1}$ with coordinates $z|\theta$ and superconformal structure defined by

$$(5.18) \quad D_\theta^* = \partial_\theta + \theta w(z) \partial_z,$$

with

$$(5.19) \quad w(z) = \prod_{i=1}^{n_R} (z - z_i).$$

Thus there are Ramond punctures at $z = z_i$, $i = 1, \dots, n_R$. Superconformal vector fields are now

$$(5.20) \quad \begin{aligned} \nu_f &= f(z)(\partial_\theta - \theta w(z)\partial_z) \\ V_g &= w(z) \left(g(z)\partial_z + \frac{g'(z)}{2}\theta\partial_\theta \right) \end{aligned}$$

Let us specialize to the case $n_R = 2$ and place the Ramond punctures at $z = 0$ and $z = a$. So the superconformal structure is defined by

$$(5.21) \quad D_\theta^* = \partial_\theta + \theta z(z - a)\partial_z.$$

Setting $f(z) = z^s$, $g(z) = z^n$, some interesting examples of superconformal vector fields that are everywhere holomorphic are

$$(5.22) \quad \begin{aligned} \nu_s &= z^s(\partial_\theta - \theta z(z - a)\partial_z) \\ V_n &= z(z - a) \left(z^n\partial_z + \frac{nz^{n-1}}{2}\theta\partial_\theta \right), \end{aligned}$$

with $s, n \geq 0$.

We want to see what this Lie algebra of superconformal symmetries looks like when viewed from near ∞ on the complex z -plane. As a shortcut, since we only going to be considering the behavior for large z , let us just set $a = 0$. The superconformal structure can be put in a standard form by setting $\theta = \theta^*/z$; indeed, $z|\theta^*$ are superconformal coordinates, as D_θ^* is a multiple of $D_{\theta^*} = \partial_{\theta^*} + \theta^*\partial_z$. Transforming to the coordinates $z|\theta^*$ and comparing to eqn. (3.4) for the superconformal symmetries of $\mathbb{C}^{1|1}$, we see that V_n corresponds to L_{n+1} and ν_s corresponds to $G_{s+1/2}$.

Accordingly, the unbroken symmetries in the presence of two Ramond punctures, as seen from large z , are generated by L_n , $n \geq 1$ and G_r , with r half-integral and $\geq 1/2$. (Happily, these form a super Lie algebra!) Missing are the following symmetry generators of $\mathbb{C}^{1|1}$: L_{-1} , L_0 , and $G_{-1/2}$. Each insertion of a Ramond vertex operator has removed one bosonic symmetry; a pair of Ramond vertex operator insertions has removed one fermionic symmetry.

In terms of ghost fields c and γ that correspond to superconformal symmetries, a standard NS vertex operator is proportional to $c\delta(\gamma)$ (or equivalently $\delta(c)\delta(\gamma)$). By setting c and γ to zero at $z = 0$, an insertion of this operator removes the symmetries L_{-1} and $G_{-1/2}$. In a similar spirit, the operator that removes L_{-1} , L_0 , and $G_{-1/2}$ is $c\partial c\delta(\gamma)$ (or $\delta(c)\delta(\partial c)\delta(\gamma)$). So if a Ramond

sector vertex operator depends on the ghosts as $c\Theta$, where Θ is some vertex operator constructed from the $\beta\gamma$ system, we would like the OPE's to read

$$(5.23) \quad c\Theta(a) c\Theta(0) \sim c\partial c\delta(\gamma), \quad a \rightarrow 0.$$

(The symbol \sim means that the operator on the right is the leading operator appearing in the OPE; there is no claim about the a -dependence. In the present example, the higher order terms are BRST-trivial.) The non-trivial part of the OPE is

$$(5.24) \quad \Theta(a) \Theta(0) \sim \delta(\gamma), \quad a \rightarrow 0.$$

Since $\delta(\gamma)$ has picture number -1 , Θ must have picture number $-1/2$. Thus we see in a slightly different way from before that if the vertex operators are supposed to naturally reflect the superconformal symmetry of the worldsheet, then the Ramond vertex operators should have picture number $-1/2$.

5.2. Matter vertex operators in the Ramond sector

In general, what can be integrated naturally³⁵ over a heterotic string world sheet Σ without a choice of coordinates is a $(0, 1)$ -form with values in \mathcal{D}^{-1} . For example, if $\tilde{\tau};\tau|\zeta$ are standard coordinates (meaning that $\tau|\zeta$ are holomorphic superconformal coordinates and $\tilde{\tau}$ is antiholomorphic and close to the complex conjugate of τ) and $X(\tilde{\tau};\tau|\zeta)$ is a scalar superfield, then $\partial_{\tilde{\tau}}X$ is a $(0, 1)$ -form and $D_{\zeta}X$ is a section of \mathcal{D}^{-1} . So $\partial_{\tilde{\tau}}X D_{\zeta}X$ is a $(0, 1)$ -form with values in \mathcal{D}^{-1} . This makes it possible to write an action for a collection of scalar fields X^I , $I = 1 \dots 10$, describing a map from Σ to spacetime, that is to the target space of the string theory. If G_{IJ} is the metric of spacetime, then we can write the supersymmetric action

$$(5.25) \quad I_X = \frac{1}{2\pi i} \int_{\Sigma} [d\tilde{\tau};d\tau|d\zeta] G_{IJ}(X^K) \partial_{\tilde{\tau}}X^I D_{\zeta}X^J,$$

which is globally-defined, independent of the choice of standard coordinates.

In the presence of a Ramond puncture supported on a divisor \mathcal{F} , matters are a little different. Let $\mathcal{O}(-\mathcal{F})$ be the holomorphic line bundle whose sections are holomorphic functions that vanish along \mathcal{F} . What can be integrated in a natural way is a $(0, 1)$ -form valued in $\mathcal{D} \otimes \mathcal{O}(-\mathcal{F})$. The Lagrangian density

³⁵This and related facts cited below are explained much more fully in [16].

associated to the action I_X is a $(0, 1)$ -form valued in \mathcal{D} , not $\mathcal{D} \otimes \mathcal{O}(-\mathcal{F})$, and the geometrical interpretation of the fields X^I in terms of maps to spacetime does not allow us to “twist” them and change this statement. So the integrand of I_X is a section of the wrong line bundle. It will have a pole along Ramond divisors, since a section of \mathcal{D} can be regarded as a section of $\mathcal{D} \otimes \mathcal{O}(-\mathcal{F})$ that has a pole along \mathcal{F} .

Let us go back to our local model of a Ramond divisor with coordinates $\tilde{z}; z|\theta$ and superconformal structure defined by $D_\theta^* = \partial_\theta + \theta z \partial_z$. We cannot choose superconformal coordinates in a neighborhood of the Ramond divisor at $z = 0$, precisely because the superconformal structure is degenerate there. But we can certainly find superconformal coordinates away from $z = 0$:

$$\begin{aligned}
 \tilde{\tau} &= \log \tilde{z} \\
 \tau &= \log z \\
 \zeta &= \theta.
 \end{aligned}
 \tag{5.26}$$

Using the relation $D_\theta^* = D_\zeta = \partial_\zeta + \zeta \partial_\tau$, we can transform the action I_X of eqn. (5.25) from the standard coordinate system $\tilde{\tau}; \tau|\zeta$ to the coordinates $\tilde{z}; z|\theta$ that behave well along \mathcal{F} . The action acquires the expected pole from the change of coordinates:

$$I_X = \frac{1}{2\pi i} \int_\Sigma [d\tilde{z}; dz|d\theta] \frac{1}{z} G_{IJ}(X^K) \partial_{\tilde{z}} X^I D_\theta^* X^J.
 \tag{5.27}$$

To understand the implications, it is convenient to expand $X^I = x^I + \theta \psi^I$, and perform the θ integral.³⁶

$$I_X = \frac{1}{2\pi} \int_{\Sigma_{\text{red}}} d^2z \left(G_{IJ} \partial_{\tilde{z}} x^I \partial_z x^J + \frac{1}{z} G_{IJ} \psi^I \frac{D}{D\tilde{z}} \psi^J \right).
 \tag{5.28}$$

We see that the pole has disappeared from the bosonic part of the action, and affects only the fermions.

To understand the fermion action, it actually helps to return to the description by τ and $\tilde{\tau}$. (For present purposes, one may identify \tilde{z} and $\tilde{\tau}$ with \bar{z} and $\bar{\tau}$.) Since $\tau = \log z$ where z is single-valued, τ is subject to the identification

$$\tau \cong \tau + 2\pi i.
 \tag{5.29}$$

³⁶As usual, the connection that enters the covariant derivative $D/D\tilde{z}$ is the pullback to the worldsheet of the Riemannian connection in spacetime.

So in terms of τ , the reduced worldsheet Σ_{red} is a cylinder of circumference 2π . The point $z = 0$ has been projected to $\text{Re}\tau = -\infty$.

In terms of τ , the fermion action is

$$(5.30) \quad I_\psi = \frac{1}{2\pi} \int_{\Sigma_{\text{red}}} d^2\tau G_{IJ}(x^K) \psi^I \frac{D}{D\bar{\tau}} \psi^J,$$

so it is just the standard Dirac action for a fermion of positive chirality, with no unusual factors. Since they are single-valued functions of z , the ψ^I are periodic under $\tau \rightarrow \tau + 2\pi i$. Thus we have landed in what is usually called the Ramond sector [51]. From here, the quantization is standard. Each of the ψ^I , $I = 1, \dots, 10$ has a zero-mode along the circle that is parametrized by $\text{Im}\tau$. Quantizing these zero-modes gives a multiplet of states in the spinor representation of $SO(10)$. (In constructions with spacetime supersymmetry, the GSO projection [52] projects out one of the two chiralities of this spinor.) The corresponding vertex operators have dimension $5/8$; this number is $10 \cdot 1/16$ where 10 is the number of components of the field ψ^I (in ten-dimensional superstring theory) and $1/16$ is the dimension of the spin field for a single fermion. The relation to the spin field of the Ising model was elucidated in [1].

We have arrived at a standard answer, but not quite in the most familiar way. There were no square roots or double-valued fields. We associated the insertion of a Ramond vertex operator (in the matter sector, so far) to a divisor \mathcal{F} along which the heterotic string worldsheet Σ was itself perfectly smooth and all fields and coordinates were single-valued, but the superconformal structure was degenerate, because $(D_\theta^*)^2$ vanishes along that divisor. The relation of our derivation to the standard one is that locally one can put the superconformal structure of Σ in a standard form by setting $\theta = \theta'/z^{1/2}$, so that $D_\theta^* = z^{1/2}(\partial_{\theta'} + \theta' \partial_z)$. The coordinates $z|\theta'$ are superconformal, but of course θ' is double-valued. In the new coordinates, we expand $X = x + \theta' \psi'$, with $\psi' = \psi/z^{1/2}$. Now there is no pole in the action, but ψ' has a square root branch point at $z = 0$.

5.3. Ghost vertex operators in the Ramond sector

With or without Ramond punctures, superconformal vector fields are sections of the line bundle \mathcal{D}^2 . In general, in quantization of gauge theory, the ghosts transform as symmetry generators with reversed statistics. So the holomorphic ghost field C is an odd section of \mathcal{D}^2 , whether or not Ramond punctures are present.

The action density for the holomorphic ghosts is $\mathcal{L}_{BC} = B\partial_z C$, where B is the antighost field. In the absence of Ramond punctures, the Berezinian of Σ in the holomorphic sense is \mathcal{D}^{-1} , so for it to be possible to integrate \mathcal{L}_{BC} , B must be a section of \mathcal{D}^{-3} . The action is then

$$(5.31) \quad I_{BC} = \frac{1}{2\pi i} \int_{\Sigma} [d\tilde{z}; dz|d\theta] B\partial_z C.$$

This formula is actually valid in an arbitrary coordinate system, not necessarily superconformal. Invariance under a change of the antiholomorphic coordinate \tilde{z} is manifest. Under an arbitrary change of the holomorphic coordinates from $z|\theta$ to, say, $z^*|\theta^*$, the holomorphic measure $[dz|d\theta]$ transforms as

$$(5.32) \quad [dz|d\theta] = [dz^*|d\theta^*] \text{Ber}^{-1} \begin{pmatrix} \partial_z z^* & \partial_z \theta^* \\ \partial_{\theta} z^* & \partial_{\theta} \theta^* \end{pmatrix}.$$

And similarly C and B transform as appropriate powers of the same Berezinian:

$$(5.33) \quad \begin{aligned} C &= C^* \cdot \text{Ber}^{-2} \begin{pmatrix} \partial_z z^* & \partial_z \theta^* \\ \partial_{\theta} z^* & \partial_{\theta} \theta^* \end{pmatrix} \\ B &= B^* \cdot \text{Ber}^3 \begin{pmatrix} \partial_z z^* & \partial_z \theta^* \\ \partial_{\theta} z^* & \partial_{\theta} \theta^* \end{pmatrix}. \end{aligned}$$

So the holomorphic ghost action in the absence of Ramond punctures can be written in the form (5.31) in terms of arbitrary holomorphic local coordinates $z|\theta$ and antiholomorphic local coordinate \tilde{z} .

There is one immediate change in the presence of Ramond punctures. The Berezinian of Σ in the holomorphic sense is now not \mathcal{D}^{-1} but $\mathcal{D}^{-1} \otimes \mathcal{O}(-\mathcal{F})$, where \mathcal{F} is the divisor of Ramond punctures (so $\mathcal{F} = \sum_{i=1}^{n_R} \mathcal{F}_i$ if there are Ramond punctures supported on divisors $\mathcal{F}_1, \dots, \mathcal{F}_{n_R}$). So now, in order for the action I_{BC} to make sense, we must interpret B as a section of $\mathcal{D}^{-3} \otimes \mathcal{O}(-\mathcal{F})$. The fact that we can twist B in this way to accommodate the Ramond punctures means that, in contrast to the matter action I_X , the ghost action I_{BC} will have no pole along \mathcal{F} . However, the definition of the ghost fields will be modified along \mathcal{F} .

To compare to standard conformal field theory formulas, we want to reduce I_{BC} to an action defined on an ordinary Riemann surface. For this, we take Σ to be a split super Riemann surface with Ramond punctures at points $z_i \in \Sigma_{\text{red}}$. For Σ_{red} of genus 0, the relevant superconformal structure was

described explicitly in eqn. (5.18). To make the reduction, we need to know the restriction of \mathcal{D} to Σ_{red} . In the absence of Ramond punctures, the restriction of \mathcal{D} to Σ_{red} is $T^{1/2}$, a square root of the holomorphic tangent bundle $T \rightarrow \Sigma_{\text{red}}$. (A choice of square root involves a choice of spin structure on Σ_{red} , and this is therefore built into the construction of a super Riemann surface with reduced space Σ_{red} .) In the presence of Ramond punctures at points z_1, \dots, z_{n_R} , however, the restriction of \mathcal{D} to Σ_{red} is not a square root of T but a square root of $T \otimes \mathcal{O}(-q_1 - \dots - q_{n_R})$. See for example section 4.2.4 of [16]. Thus, constructing a super Riemann surface Σ with Ramond punctures at the points q_1, \dots, q_{n_R} requires a choice of a line bundle $\mathcal{R} \rightarrow \Sigma_{\text{red}}$ with an isomorphism

$$(5.34) \quad \mathcal{R}^2 \cong T \otimes \mathcal{O}(-q_1 - \dots - q_{n_R}).$$

For our purposes here, we are interested in the local behavior near a single Ramond puncture, which we may call q , at $z = 0$. So

$$(5.35) \quad \mathcal{R}^2 \cong T \otimes \mathcal{O}(-q).$$

The dual relation reads

$$(5.36) \quad \mathcal{R}^{-2} \cong K \otimes \mathcal{O}(q),$$

where K – the dual of T – is the canonical bundle of Σ_{red} . \mathcal{R} is the restriction of \mathcal{D} to Σ_{red} .

We expand the ghost field C in powers of θ :

$$(5.37) \quad C(z|\theta) = \widehat{c}(z) + \theta \widehat{\gamma}(z).$$

The reason for the hats is that, as will become clear, \widehat{c} and $\widehat{\gamma}$ do not quite coincide with the ghost fields c and γ as conventionally defined. \widehat{c} is a section of $\mathcal{R}^2 \cong T \otimes \mathcal{O}(-q)$. In other words, we can view \widehat{c} as a section of T that vanishes at $z = 0$. Since θ is a section of \mathcal{R} , we can understand $\widehat{\gamma}$ as a section of \mathcal{R} .

We make a similar expansion for B :

$$(5.38) \quad B = \widehat{\beta} + \theta \widehat{b}.$$

Here, as B is a section of $\mathcal{R}^{-3} \otimes \mathcal{O}(-q)$, it follows that $\widehat{\beta}$ is a section of $\mathcal{R}^{-3} \otimes \mathcal{O}(-q)$, and that \widehat{b} is a section of $\mathcal{R}^{-4} \otimes \mathcal{O}(-q)$. In view of (5.36), it is equivalent to say that \widehat{b} is a section of $K^2 \otimes \mathcal{O}(q)$ – in other words \widehat{b} is a

quadratic differential that may have a pole at $z = 0$ – and $\widehat{\beta}$ is a section of $K \otimes \mathcal{R}^{-1}$.

In terms of these variables, the ghost action becomes

$$(5.39) \quad I_{BC} = \frac{1}{2\pi} \int_{\Sigma_{\text{red}}} d^2z \left(\widehat{b} \partial_{\bar{z}} \widehat{c} + \widehat{\beta} \partial_{\bar{z}} \widehat{\gamma} \right).$$

This description of the fermionic ghosts \widehat{b} and \widehat{c} and the action describing them is almost standard. The only novelty is that \widehat{c} is constrained to vanish at $z = 0$ while \widehat{b} is allowed to have a pole there. We discuss the significance of this in section 5.3.2.

More unusual is the description that we have reached for the commuting ghosts $\widehat{\beta}$ and $\widehat{\gamma}$. There are no square root branch points in sight. $\widehat{\gamma}$ is a section of a line bundle \mathcal{R} , but perfectly single-valued, and similarly $\widehat{\beta}$ is a perfectly single-valued section of $K \otimes \mathcal{R}^{-1}$. The $\widehat{\beta}\widehat{\gamma}$ system is just a system of chiral bosons coupled to a line bundle – or loosely speaking, to an abelian gauge field. What then is the relation to the usual description [1] by spin fields?

5.3.1. The usual description of the $\beta\gamma$ system As in section 5.2, we can transform to the usual description of the $\beta\gamma$ system by introducing a double-valued coordinate $\theta' = \theta\sqrt{z}$, so that $z|\theta'$ are superconformal coordinates but θ' has a branch point at $z = 0$. To understand how to transform $\widehat{\gamma}$ and $\widehat{\beta}$ under the change of coordinates, we recall that $\widehat{\gamma}$ is a section of $\mathcal{R} \cong \mathcal{D}|_{\Sigma_{\text{red}}}$, so that $\widehat{\gamma}\partial_{\theta}$ should be invariant under the redefinition of θ . So we transform to the new coordinates by $\widehat{\gamma}\partial_{\theta} = \gamma\partial_{\theta'}$, or $\widehat{\gamma}\sqrt{z} = \gamma$. In transforming $\widehat{\beta}$, we view it as a section of $\mathcal{R}^{-1} \otimes K$, so we set $\widehat{\beta}/\sqrt{z} = \beta$. The conventional ghost fields γ and β in superconformal coordinates $z|\theta'$ are often denoted as $\gamma^{\theta'}$ and $\beta_{z\theta'}$; this sort of notation makes the transformation we just made more obvious.

The fields $\widehat{\gamma}$ and $\widehat{\beta}$ are regular at $z = 0$ – as they are simply chiral bosons valued in a line bundle. So the behavior of γ and β near $z = 0$ is

$$(5.40) \quad \gamma \sim z^{1/2}, \quad \beta \sim z^{-1/2}, \quad z \rightarrow 0.$$

In conventional language, eqn. (5.40) means that γ and β are coupled to a vertex operator $\Theta_{-1/2}$ at $z = 0$ that is a spin operator of picture number $-1/2$. This follows from the Ramond sector mode expansions

$$(5.41) \quad \gamma(z) = \sum_{r \in \mathbb{Z}} z^{-r+1/2} \gamma_r, \quad \beta(z) = \sum_{r \in \mathbb{Z}} z^{-r-3/2} \beta_r.$$

To get the behavior (5.40), we want

$$(5.42) \quad \begin{aligned} \gamma_r \Theta_{-1/2} &= 0, \quad r > 0 \\ \beta_r \Theta_{-1/2} &= 0, \quad r \geq 0, \end{aligned}$$

and these are the picture number $-1/2$ conditions that we found in another way in eqns. (5.14) and (5.15).

For completeness, we will explain why the operator $\Theta_{-1/2}$ has ghost number $-1/2$ and dimension $3/8$. The original explanation [1] involved a transformation from fields $\beta_{z\theta'}$, $\gamma^{\theta'}$ to a new set of fields ϕ , η , ξ . As these variables do not have a very transparent interpretation in terms of the geometry of super Riemann surfaces,³⁷ we will instead follow a method [53] that has been used in the theory of orbifolds. The short distance behavior of the $\beta(z) \cdot \gamma(w)$ operator product for $z \rightarrow w$ is

$$(5.43) \quad \beta_{z\theta'}(z) \gamma^{\theta'}(w) \sim -\frac{1}{z-w}.$$

The $\beta \cdot \gamma$ two-point function in the presence of the operator $\Theta_{-1/2}$ inserted at $z = 0$ with no other operator insertions³⁸ is

$$(5.44) \quad \langle \beta_{z\theta'}(z) \gamma^{\theta'}(w) \rangle_{\Theta_{-1/2}} = -\frac{1}{z-w} \sqrt{\frac{w}{z}}.$$

This formula is determined by the fact that it has the right behavior for $z \rightarrow w$, for $z \rightarrow 0$, and for $w \rightarrow 0$, and also has no other singularities and the slowest possible growth at infinity (the last condition reflects the fact that we take the spin field at infinity to be a primary). The ghost number current of the $\beta\gamma$ system is $J_{\beta\gamma}(w) = - : \beta\gamma(w) :$ or more explicitly

$$(5.45) \quad J_{\beta\gamma}(w) = \lim_{z \rightarrow w} \left(-\beta(z)\gamma(w) - \frac{1}{z-w} \right),$$

³⁷The reader may object that also the double-valued coordinate θ' is not entirely natural in super Riemann surface theory. Indeed, one may prefer to rephrase the computation we describe here in the coordinates $z|\theta$ that behave well at $z = 0$.

³⁸To be more exact, we assume that there are no more operator insertions except a second spin field at $z = \infty$. This second spin field has picture number $-3/2$, as one may learn by transforming $z \rightarrow z' = 1/z$ and examining the behavior of $\beta_{z\theta'}$ and $\gamma^{\theta'}$ for $z' \rightarrow 0$. More generally, as explained in section 3.6.3, a sensible $\beta\gamma$ path integral in genus 0 always involves a product of operator insertions of total picture number -2 .

where here normal-ordering is carried out by subtracting the vacuum expectation value $\langle -\beta(z)\gamma(w) \rangle_{\text{vac}} = 1/(z-w)$. A short calculation reveals that in the presence of the operator $\Theta_{-1/2}$, $J_{\beta\gamma}(w)$ has a simple pole at $w=0$ with residue $-1/2$,

$$(5.46) \quad \langle J_{\beta\gamma}(w) \rangle_{\Theta_{-1/2}} = -\frac{1}{2w},$$

so $\Theta_{-1/2}$ has ghost number $-1/2$. Similarly, the stress tensor of the $\beta\gamma$ system is

$$(5.47) \quad T_{\beta\gamma} =: \partial_z \beta_{z\theta'} \cdot \gamma^{\theta'} : - \frac{3}{2} \partial_z (: \beta_{z\theta'} \gamma^{\theta'} :).$$

A short calculation gives

$$(5.48) \quad \langle T_{\beta\gamma}(w) \rangle_{\Theta_{-1/2}} = \frac{3}{8w^2},$$

so that $\Theta_{-1/2}$ has dimension $3/8$.

It is straightforward to generalize this analysis to determine the conformal dimension and the ghost number of the operator Θ_{-t} that represents the $\beta\gamma$ ground state with picture number $-t$, for any integer or half-integer t . In this case, we want $\gamma(w)$ to behave as w^t and $\beta(z)$ to behave as z^{-t} near $z=0$. The analog of (5.44) is

$$(5.49) \quad \langle \beta_{z\theta'}(z) \gamma^{\theta'}(w) \rangle_{\Theta_{-t}} = -\frac{1}{z-w} \left(\frac{w}{z} \right)^t.$$

A calculation along the above lines gives

$$(5.50) \quad \begin{aligned} \langle J_{\beta\gamma}(w) \rangle_{\Theta_{-t}} &= -\frac{t}{w} \\ \langle T_{\beta\gamma}(w) \rangle_{\Theta_{-t}} &= -\frac{t(t-2)}{2w^2}. \end{aligned}$$

Thus Θ_{-t} has ghost number $-t$ and dimension $-t(t-2)/2$. For integer t , this operator can be constructed from a product of delta functions, as in eqn. (10.75) for $t \geq 0$, and its ghost number and dimension can be computed classically.

5.3.2. The definition of the ghosts The description of the anticommuting ghosts \hat{b} and \hat{c} that we arrived at in eqn. (5.39) differs in precisely one

way from the usual description. In the conventional formulation, c is a vector field – a section of T – and b is a quadratic differential – a section of K^2 . Instead, in our derivation, \hat{c} was a section of $T \otimes \mathcal{O}(-q)$, that is a vector field that is constrained to vanish at the Ramond puncture, while \hat{b} was a section of $K^2 \otimes \mathcal{O}(q)$, that is a quadratic differential that is allowed to have a pole at the Ramond puncture.

The relation between the two descriptions is that in the standard approach, vanishing of c at the insertion point of a vertex operator is not part of the definition of c , but the vertex operator contains a factor of c , or equivalently of $\delta(c)$, that enforces this vanishing. Once this factor is included, b can have a pole at the location of the vertex operator.

Instead of including in a bosonic string conformal vertex operator a factor $\tilde{c}c$, as one usually does, we could declare \tilde{c} and c to be fields (sections of $T_L\Sigma$ and $T_R\Sigma$, respectively) that vanish at the positions of all vertex operator insertions. Then we would say that the simplest conformal vertex operators are constructed from matter fields only.³⁹ Defining \tilde{c} and c to vanish at vertex operator insertions is natural because \tilde{c} and c are associated to gauge symmetries, and the symmetries of bosonic string theory are worldsheet diffeomorphisms that leave fixed the vertex operator insertions.

Similarly, for the NS sector of superstrings, instead of including a factor $\tilde{c}\delta(\gamma)$ in the vertex operator, we could equivalently say that \tilde{c} , c , and γ are all constrained to vanish at points at which vertex operators are inserted. This is natural for the same reason as in the last paragraph.

We did not adopt this viewpoint in the present paper, in part because it might have made the formulas look unfamiliar. However, for the Ramond sector, this viewpoint has been more or less forced upon us. What is usually regarded as a spin operator Θ of the $\beta\gamma$ system really appeared in the derivation of eqn. (5.39) as a modification of the definitions of the fields β and γ . γ , rather than being a section of $T^{1/2}$, was a section of a more general line bundle \mathcal{R} , and similarly β was a section of $K \otimes \mathcal{R}^{-1}$. The vertex operator was just this instruction about how to modify the definition of the fields. The

³⁹ A constraint that \tilde{c} and c should vanish at a given point p could be enforced by incorporating additional variables – Lagrange multipliers – in the path integral, as in eqn. (10.21). If we do this, the path integral measure, instead of being conformally invariant as in the usual description, transforms under conformal transformations like the operator $\tilde{c}c(p)$, in other words like a field of dimension $(-1, -1)$ at the point p ; this reflects the scaling behavior of the measure of the Lagrange multipliers. Hence insertion of a $(1, 1)$ matter vertex operator V at the point p is necessary to restore conformal invariance of the path integral. Analogous statements hold for the NS and R sectors of superstrings.

reason that the Ramond sector gives a sharper message about how the formalism should be developed is that a Ramond puncture is an intrinsic part of the geometry of a super Riemann surface, in contrast to an NS puncture (or a puncture on a purely bosonic worldsheet), which can be viewed if one wishes as something extra that is tacked on to a preexisting super Riemann surface. At any rate, part of the message of the Ramond sector seems to be that we should view the traditional factors $\tilde{c}c$, $\tilde{c}c\delta(\gamma)$, and $\tilde{c}c\Theta$ in the vertex operators as shorthand ways of saying how the definition of the ghosts is modified by the presence of the vertex operators.

For more on some matters considered in this section, see section 10.3.4.

5.4. Ramond amplitudes

Now we will discuss scattering amplitudes including Ramond states.

Once one constructs BRST-invariant Ramond vertex operators, many steps follow in the familiar fashion. Given any assortment of Ramond and/or Neveu-Schwarz BRST-invariant vertex operators $\mathcal{V}_1, \dots, \mathcal{V}_s$, the worldsheet integral computes for us a closed form $F_{\mathcal{V}_1, \dots, \mathcal{V}_s}(\mathcal{J}, \delta\mathcal{J})$ on the space of super-complex structures. If the vertex operators obey the usual conditions

$$(5.51) \quad b_n \mathcal{V}_1 = \beta_r \mathcal{V}_1 = 0, \quad n, r \geq 0,$$

that make possible a superconformally invariant formalism, then $F_{\mathcal{V}_1, \dots, \mathcal{V}_s}(\mathcal{J}, \delta\mathcal{J})$ is as usual a pullback from an appropriate product $\mathcal{M}_L \times \mathcal{M}_R$ of left and right moduli spaces. For example, for the heterotic string, \mathcal{M}_L is the moduli space of genus g Riemann surfaces with s punctures, and \mathcal{M}_R is the moduli space of genus g super Riemann surfaces with an appropriate number of NS and Ramond punctures. In the familiar way, one defines an integration cycle $\Gamma \subset \mathcal{M}_L \times \mathcal{M}_R$ and one defines the genus g contribution to the scattering amplitude as $\int_{\Gamma} F_{\mathcal{V}_1, \dots, \mathcal{V}_s}(\mathcal{J}, \delta\mathcal{J})$.

There are also no surprises concerning gauge-invariance. If for example $\mathcal{V}_1 = \{Q_B, \mathcal{W}_1\}$, then one has the standard relation

$$(5.52) \quad F_{\{Q_B, \mathcal{W}_1\}, \mathcal{V}_2, \dots, \mathcal{V}_s} + dF_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_s} = 0.$$

Here it is important to know that if \mathcal{V}_1 obeys the conditions (5.51), and $\mathcal{V}_1 = \{Q_B, \mathcal{W}_1\}$ for some \mathcal{W}_1 , then we can choose \mathcal{W}_1 to also obey the same conditions (see appendix C.2). This enables one to pull $F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_s}$ back to the finite-dimensional space $\mathcal{M}_L \times \mathcal{M}_R$, and to deduce from (5.52) a relation between finite-dimensional integrals:

$$(5.53) \quad \int_{\Gamma} F_{\{Q_B, \mathcal{W}_1\}, \mathcal{V}_2, \dots, \mathcal{V}_s} + \int_{\Gamma} dF_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_s} = 0.$$

This relation shows as usual that gauge-invariance holds if there are no surface terms when one integrates by parts over Γ .

One important difference between Ramond vertex operators and the other cases is that there does *not* exist, in a natural sense, an integrated version of a Ramond vertex operator. The fundamental reason for this is that the forgetful map (2.61) does not have an analog for Ramond punctures. For NS punctures there is such a map:

$$(5.54) \quad \begin{array}{ccc} \Sigma & \rightarrow & \mathfrak{M}_{g, n_{\text{NS}}, n_{\text{R}}} \\ & & \downarrow \pi \\ & & \mathfrak{M}_{g, n_{\text{NS}}-1, n_{\text{R}}} \end{array}$$

Here $\mathfrak{M}_{g, n_{\text{NS}}, n_{\text{R}}}$ is the moduli space of genus g super Riemann surfaces with n_{NS} NS punctures and n_{R} Ramond punctures. Since a NS puncture is simply a chosen point in a pre-existing super Riemann surface, we can forget such a puncture if we wish, and this gives the fibration (5.54). An integral over $\mathfrak{M}_{g, n_{\text{NS}}, n_{\text{R}}}$ can be reduced to an integral over $\mathfrak{M}_{g, n_{\text{NS}}-1, n_{\text{R}}}$ by integrating first over the fibers of this fibration, with the help of the integrated NS vertex operator. There is no analogous forgetful map for Ramond punctures, since a Ramond puncture is part of the superconformal structure of Σ ; there is no way to forget such a puncture while keeping fixed the rest of a super Riemann surface, and there is no sensible notion of two super Riemann surfaces being the same except with Ramond punctures in different places. So there is no superconformal notion of an integrated Ramond vertex operator. Any definition of an integrated Ramond vertex operator depends on a method of integrating over odd moduli, and has the limitations of such methods.

However, more or less everything else we have said, for instance in sections 3.4-3.6, about how to compute $F(\mathcal{J}, \delta\mathcal{J})$ is substantially unaffected by Ramond punctures. A few minor differences are as follows. We originally introduced the gravitino field χ_z^θ in section 3.5 as a section of $T^{1/2}$, a square root of the holomorphic tangent bundle T of the reduced space Σ_{red} . A choice of $T^{1/2}$ is equivalent to a choice of spin structure on Σ_{red} . The reason that χ_z^θ takes values in $T^{1/2}$ is that the restriction of \mathcal{D} to Σ_{red} is isomorphic to $T^{1/2}$, in the absence of Ramond punctures. However, in the presence of Ramond punctures at points $q_1, \dots, q_{n_{\text{R}}}$, the restriction of \mathcal{D} to Σ_{red} is not a square root of T but a square root of $T \otimes \mathcal{O}(-q_1 - \dots - q_{n_{\text{R}}})$. (This fact played a prominent role in section 5.3.) In other words, the restriction of \mathcal{D} , which we denote \mathcal{R} , possesses an isomorphism

$$(5.55) \quad \mathcal{R}^2 \cong T \otimes \mathcal{O}(-q_1 - \dots - q_{n_{\text{R}}}).$$

The degree of \mathcal{R} is

$$(5.56) \quad \deg \mathcal{R} = 1 - g - \frac{1}{2}n_{\mathbf{R}}.$$

With Ramond punctures, the gravitino field χ_z^θ is a $(0, 1)$ -form with values in \mathcal{R} . If we include also NS punctures at points $p_1, \dots, p_{n_{\text{NS}}}$, then χ_z^θ becomes a $(0, 1)$ -form valued in $\widehat{\mathcal{R}} = \mathcal{R} \otimes \mathcal{O}(-\sum_{i=1}^{n_{\text{NS}}} p_i)$. The degree of this line bundle is

$$(5.57) \quad \deg \widehat{\mathcal{R}} = 1 - g - n_{\text{NS}} - \frac{1}{2}n_{\mathbf{R}}.$$

A gauge transformation of the gravitino is $\chi_z^\theta \rightarrow \chi_z^\theta + \partial_z y^\theta$, where now y^θ is valued in $\widehat{\mathcal{R}}$. Modulo gauge transformations, χ_z^θ is an element of $H^1(\Sigma_{\text{red}}, \widehat{\mathcal{R}})$, whose dimension is $\Delta = 2g - 2 + n_{\text{NS}} + \frac{1}{2}n_{\mathbf{R}}$, which accordingly is the odd dimension of $\mathfrak{M}_{g, n_{\text{NS}}, n_{\mathbf{R}}}$.

In the picture-changing formalism, we select points r_1, \dots, r_Δ and write

$$(5.58) \quad \chi_z^\theta = \sum_{\sigma=1}^{\Delta} \eta_\sigma \delta_{r_\sigma}$$

with anticommuting parameters η_σ . A spurious singularity occurs where the gravitino modes in this expression do not furnish a basis of $H^1(\Sigma_{\text{red}}, \widehat{\mathcal{R}})$. The condition for this is now that

$$(5.59) \quad H^0(\Sigma, \widehat{\mathcal{R}} \otimes \mathcal{O}(\sum_{\sigma=1}^{\Delta} r_\sigma)) \neq 0.$$

The line bundle $\widehat{\mathcal{R}} \otimes \mathcal{O}(\sum_{\sigma=1}^{\Delta} r_\sigma)$ has degree $g - 1$, as always. Generically, the condition (5.59) will be satisfied as one varies one parameter, leading to a spurious pole. However, in genus 0, a line bundle of degree -1 has no holomorphic section, so the gauge-fixing prescription associated to the picture-changing formalism is always correct in the interior of moduli space. The subtleties that occur at infinity in moduli space are not important for tree-level amplitudes at reasonably generic external momenta, so there is no problem to compute scattering amplitudes of Ramond and Neveu-Schwarz states at tree-level in the picture-changing formalism.

5.4.1. The simplest examples The simplest examples of scattering amplitudes are those for which the odd dimension Δ of the moduli space vanishes,

so that no picture-changing insertions are necessary. This happens only for $g = 0$, for which $\Delta = -2 + n_{\text{NS}} + \frac{1}{2}n_{\text{R}}$. So $\Delta = 0$ for $n_{\text{NS}} = 1$, $n_{\text{R}} = 2$, and for $n_{\text{NS}} = 0$, $n_{\text{R}} = 4$.

For $n_{\text{NS}} = 1$, $n_{\text{R}} = 2$, the even and odd dimensions of the moduli space both vanish. The scattering amplitude is simply computed as a three-point function on a genus 0 super Riemann surface of one NS vertex operator and two Ramond vertex operators:

$$(5.60) \quad \left\langle \tilde{c}\delta(\gamma)V_1(\tilde{z}_1; z_1|0)\tilde{c}\Theta_{-1/2}V'_2(\tilde{z}_2; z_2)\tilde{c}\Theta_{-1/2}V'_3(\tilde{z}_3; z_3) \right\rangle.$$

Here V_1 is a matter vertex operator in the NS sector, while V'_2 and V'_3 are Ramond sector matter vertex operators. The points z_1 , z_2 , and z_3 are arbitrary and there are no moduli to integrate over. The formula is a shorthand for the following recipe. Pick points z_1 , z_2 , z_3 in a purely bosonic \mathbb{CP}^1 and use the formula (5.18) to construct a super Riemann surface Σ with Ramond divisors at z_2 and z_3 , and no others. Pick a matter vertex operator $V_1(\tilde{z}; z|\theta)$ in the NS sector and evaluate it at⁴⁰ $\tilde{z}_1; z_1|0$. Insert matter vertex operators V'_2 and V'_3 at the Ramond divisors at $z = z_2$ and $z = z_3$, respectively.⁴¹ Compute the path integral with these insertions to get the amplitude (5.60).

For $n_{\text{NS}} = 0$, $n_{\text{R}} = 4$, since there are no odd moduli, the moduli space is just the bosonic moduli space that parametrizes four points in \mathbb{CP}^1 , up to the action of $SL(2, \mathbb{C})$. We can use $SL(2, \mathbb{C})$ to specify three points in an arbitrary way; then we integrate over the fourth point. The scattering amplitude is

$$(5.61) \quad \left\langle \prod_{i=1}^3 \tilde{c}\Theta_{-1/2}V'_i(\tilde{z}_i; z_i) \int d^2z_4 \Theta_{-1/2}V'_4(\tilde{z}_4; z_4) \right\rangle,$$

where V'_i , $i = 1, \dots, 4$ are Ramond sector matter vertex operators. This formula is again a shorthand for a recipe that involves constructing a super

⁴⁰The automorphism group of a genus 0 super Riemann surface with two Ramond punctures has dimension 1|1 and enables one to map a given NS puncture to a chosen point such as $\tilde{z}_1; z_1|0$ in a unique way. See for example section 5.1.4 of [16].

⁴¹These vertex operators are associated to Ramond divisors, so there is no choice in where they are to be inserted once the geometry of Σ is fixed. When we denote them as $V'_i(\tilde{z}_i; z_i)$, $i = 2, 3$, this is just meant as a reminder of which Ramond operator is inserted at which Ramond divisor. In any case, as these operators are associated to divisors, not points, they do not depend on θ , only on \tilde{z} and z . There is no misprint in the fact that a θ -dependence of V'_2 and V'_3 is not indicated in (5.60).

Riemann surface with the specified Ramond divisors and inserting the indicated matter vertex operators at those divisors.

Still in $g = 0$, let us consider amplitudes with $\Delta > 0$. If $n_R \leq 4$, one has $\Delta \leq n_{NS}$. If one were to forget the NS insertions, there would be no odd moduli at all, so one can integrate over all odd moduli by replacing some of the NS vertex operators with their integrated versions. If $n_R > 4$, one additionally needs $\frac{1}{2}n_R - 2$ picture-changing operators. To minimize arbitrary choices and preserve holomorphy, a relatively simple procedure is to pick $\frac{1}{2}n_R - 2$ of the Ramond vertex operators and attach a single picture-changing operator to each one. Thus one calculates tree amplitudes using Ramond vertex operators of picture number $+1/2$ as well as $-1/2$, as originally described in [1]. One may also compute tree amplitudes with more general choices of the pictures.

5.5. Duality

Let \mathcal{H} be the space of all string states in a given string theory model or – by the state-operator correspondence – the space of all vertex operators. The two-point function in genus 0 gives a nondegenerate pairing $\omega : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ or equivalently an identification – which we also call ω – between \mathcal{H} and its dual space \mathcal{H}^* . For brevity, we will describe this for open strings. As usual, the discussion for a chiral sector of closed strings is almost the same. We take the string worldsheet (or its reduced space, in the case of superstrings) to be the upper half of the complex z -plane.

The pairing ω is defined by a two-point function of vertex operators inserted at, say, $z = 0$ and $z = 1$. For vertex operators \mathcal{V} , \mathcal{W} , we set

$$(5.62) \quad \omega(\mathcal{V}, \mathcal{W}) = \langle \mathcal{V}(0) \mathcal{W}(1) \rangle.$$

This formula defines a nondegenerate bilinear form on the space of all vertex operators. This pairing is BRST-invariant in the sense that

$$(5.63) \quad \omega(Q_B \mathcal{U}, \mathcal{V}) + (-1)^{|\mathcal{U}|} \omega(\mathcal{U}, Q_B \mathcal{V}) = 0.$$

(Here $|\mathcal{U}|$ is 0 or 1 depending on whether the state \mathcal{U} is bosonic or fermionic.) As a result, there is an induced pairing on the BRST cohomology: if \mathcal{V} and \mathcal{W} are in the kernel of Q_B , then $\omega(\mathcal{V}, \mathcal{W})$ is invariant under $\mathcal{V} \rightarrow \mathcal{V} + Q_B \mathcal{U}$, or under a similar transformation of \mathcal{W} . This pairing on the BRST cohomology is nondegenerate. This assertion follows from the BRST version of the no-ghost theorem, which gives convenient lightcone representatives of the BRST cohomology classes, making the nondegeneracy of the pairing manifest. An

instructive example of passing from the pairing among all vertex operators to the pairing between physical states can be found in section 5.5.1.

In the case of bosonic strings, ω has ghost number -3 ; this means that if \mathcal{U} and \mathcal{V} have definite ghost numbers, the sum of their ghost numbers must be 3 in order to have $\omega(\mathcal{U}, \mathcal{V}) \neq 0$. For example, if U and V are primary fields in the matter sector, we can have

$$(5.64) \quad \langle c\partial cU(0) cV(1) \rangle \neq 0.$$

The insertions of $c\partial c(0)$ and $c(1)$ have the following intuitive interpretation. To get a non-zero path integral, we must remove the zero-modes of the ghost field c , by eliminating the $SL(2, \mathbb{R})$ symmetry of Σ . We can do this by restricting the diffeomorphism group of Σ to its subgroup generated by vector fields that vanish, together with their first derivative, at $z = 0$ and that also vanish at $z = 1$. To achieve this, we must place these constraints on the ghost field c , so we insert $\delta(c)\delta(\partial c) = c\partial c$ at $z = 0$, and $\delta(c) = c$ at $z = 1$.

It is convenient to write \mathcal{H}_n for the subspace of \mathcal{H} consisting of states of ghost number n . Then ω is a nondegenerate pairing $\mathcal{H}_n \times \mathcal{H}_{3-n} \rightarrow \mathbb{C}$, or equivalently an isomorphism between \mathcal{H}_{3-n} and the dual of \mathcal{H}_n :

$$(5.65) \quad \omega : \mathcal{H}_n^* \cong \mathcal{H}_{3-n}.$$

For superstring theory, one should specify the picture numbers as well as the ghost numbers of the states. We write $\mathcal{H}_{n;k}$ for the space of vertex operators of ghost number n and picture number k . The inner product $\omega(\mathcal{U}, \mathcal{V})$ is again defined as in (5.62); for it to be non-zero, the ghost numbers must add to 1 and the picture numbers to -2 . So ω is a nondegenerate pairing $\mathcal{H}_{n;k} \otimes \mathcal{H}_{1-n;-2-k} \rightarrow \mathbb{C}$, or equivalently an isomorphism

$$(5.66) \quad \omega : \mathcal{H}_{n;k}^* \cong \mathcal{H}_{1-n;-2-k}.$$

In this paper, we almost always restrict ourselves to the canonical picture numbers, namely -1 for NS states and $-1/2$ for Ramond states. For the NS sector, there is no problem with this restriction: we get a nondegenerate pairing (5.62) or an isomorphism (5.66) with $k = -2 - k = -1$. Given matter primaries U and V with a non-zero two-point function in the matter sector, a typical non-zero pairing in the full theory with the ghosts is

$$(5.67) \quad \langle c\partial c\delta(\gamma)U(0) c\delta(\gamma)V(1) \rangle \neq 0.$$

For the Ramond sector, the picture number is a half-integer, and if one state has the canonical picture number $-1/2$, then the second will have to have picture number $-3/2$. This corresponds to an isomorphism

$$(5.68) \quad \omega : \mathcal{H}_{n,-3/2}^* \cong \mathcal{H}_{1-n,-1/2}.$$

The fact that the dual of a state in the canonical picture is a state of picture number $-3/2$ can be given the following interpretation. The superconformal structure of a worldsheet whose reduced space is the upper half-plane and that has Ramond punctures at $z = 0$ and $z = 1$ is generated by the odd vector field

$$(5.69) \quad D_\theta^* = \partial_\theta + z(z-1)\theta\partial_z.$$

This worldsheet has an automorphism group of dimension $0|1$, generated by the odd superconformal vector field

$$(5.70) \quad \nu = \partial_\theta - z(z-1)\theta\partial_z,$$

which is regular (and nonvanishing) at $z = \infty$. To get a sensible worldsheet path integral (without a zero-mode of the commuting ghost field γ), we have to remove this automorphism group. For this, we have to insert at one of the two Ramond divisors – either at $z = 0$ or at $z = 1$ – a vertex operator that is associated to a point on the divisor in question, not to the whole divisor. In other words, as explained in section 5.1.1, one of the two vertex operators has to have picture number $-3/2$ rather than the canonical value $-1/2$.

In section 6.5, we will find that this occurrence in the duality of an operator of non-canonical picture number is related to the fact that the Ramond sector propagator is proportional to G_0/L_0 , in contrast to $1/L_0$ in the NS sector (or for bosonic strings). The field theory limit of this is that fermions obey first order wave equations (such as the Dirac equation for spin $1/2$) in contrast to the second order wave equations obeyed by bosons.

5.5.1. Massless fermions in ten dimensions We will make the Ramond sector duality more explicit for the basic case of massless fermion vertex operators in ten-dimensional Minkowski spacetime $\mathbb{R}^{1,9}$. The basic fermion vertex operator constructed from the matter fields is [1] a spin field Σ that transforms in the spinor representation of $SO(1,9)$. Σ can have either $SO(1,9)$ chirality; for its components of positive or negative chirality, we write Σ_α and Σ^β , respectively, with $\alpha, \beta = 1, \dots, 16$. Σ_α and Σ^β transform oppositely under the GSO projection. We also write $\Sigma_\alpha(p)$, $\Sigma^\beta(p)$ for the corresponding

operators at spacetime momentum p (for example, $\Sigma_\alpha(p) = \Sigma_\alpha \exp(ip \cdot X)$, where X^I , $I = 0 \dots 9$ are bosonic fields describing the motion of the string in $\mathbb{R}^{1,9}$).

Let $\Theta_{-1/2}$ and $\Theta_{-3/2}$ be the ground states of the $\beta\gamma$ system at picture number $-1/2$ and $-3/2$, respectively. These operators transform oppositely under the GSO projection. (For example, this follows from the fact that the part of the picture-changing operator that acts on the $\beta\gamma$ system, namely $\delta(\beta)$, is GSO-odd, since⁴² it is odd under $\beta \rightarrow -\beta$.) So if $\Theta_{-1/2}\Sigma_\alpha(p)$ is GSO-even, then $\Theta_{-3/2}\Sigma^\beta(p)$ is likewise GSO-even.

Let u^α and v_β be commuting c -number spinors of the indicated chirality. The pairing in the Ramond sector at the lowest mass level is

$$(5.71) \quad \langle c\partial c\Theta_{-1/2}u^\alpha\Sigma_\alpha(p) \cdot c\Theta_{-3/2}v_\beta\Sigma^\beta(q) \rangle = u^\alpha v_\alpha (2\pi)^{10} \delta^{10}(p+q).$$

Clearly, this pairing is nondegenerate in the space of all vertex operators of this type. It remains nondegenerate if we pass to the BRST cohomology. To do this, we must restrict p and q by $p^2 = q^2 = 0$, which follows from Q_B -invariance. Q_B -invariance also imposes on u^α the constraint

$$(5.72) \quad p^I \Gamma_{I\alpha\beta} u^\beta = 0,$$

where Γ_I , $I = 0, \dots, 9$ are the gamma matrices. This constraint is simply the Dirac equation written in momentum space, and arises because the condition $Q_B(c\Theta_{-1/2}u^\alpha\Sigma_\alpha(p)) = 0$ implies $G_0(u^\alpha\Sigma_\alpha(p)) = 0$, where G_0 acts on the massless level as the Dirac operator of field theory. On v_β , Q_B -invariance imposes no such constraint,⁴³ but instead when we pass to the Q_B cohomology, there is an equivalence relation

$$(5.73) \quad v_\beta \cong v_\beta + p^I \Gamma_{I\beta\gamma} w^\gamma,$$

for any w^γ . Evidently, with the constraint (5.72) and the equivalence relation (5.73), the pairing (5.71) remains nondegenerate.

⁴²In treating bosonic integration as an algebraic operation in the sense introduced in section 3.2.2, one uses $\delta(\lambda\beta) = \lambda^{-1}\delta(\beta)$, not $\delta(\lambda\beta) = |\lambda|^{-1}\delta(\beta)$. See for example eqn. (3.37) of [15].

⁴³The G_0 constraint on $u^\alpha\Sigma_\alpha(p)$ arises because $Q_B = \gamma_0 G_0 + \dots$ and $\gamma_0\Theta_{-1/2} \neq 0$. Because $\gamma_0\Theta_{-3/2} = 0$, the condition $Q_B(c\Theta_{-3/2}v_\beta\Sigma^\beta(q)) = 0$ does not lead to a G_0 constraint on $v_\beta\Sigma^\beta(q)$. Instead, there is a gauge-equivalence on $v_\beta\Sigma^\beta(q)$, generated by a gauge transformation with gauge parameter $c\beta_0\Theta_{-3/2}w^\gamma\Sigma_\gamma(q)$.

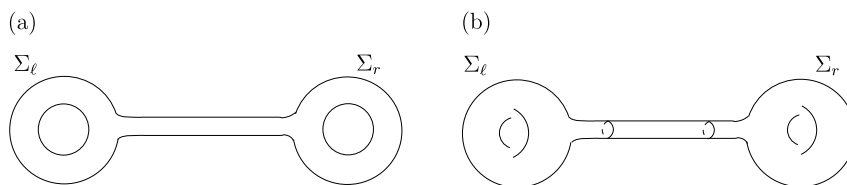


Figure 3: (a) A long strip connecting two parts of an open-string worldsheet. This is meant to be a “flat,” purely two-dimensional picture. (b) A long tube connecting two parts of a closed-string worldsheet.

6. The Propagator

In field theory, the Feynman propagator can be represented as an integral over elapsed proper time. For example, for a scalar particle of mass m , the propagator in Euclidean signature is

$$(6.1) \quad \frac{1}{p^2 + m^2} = \int_0^\infty ds \exp(-s(p^2 + m^2)),$$

where s is the Schwinger proper time parameter. The two limits of the s integral are both very important. The ultraviolet region is $s \cong 0$. The ultraviolet divergences of Feynman diagrams arise when the proper time parameters associated to the propagators in a loop are all simultaneously small. On the other hand, $s \rightarrow \infty$ is the infrared region. The pole of the propagator for $p^2 + m^2 = 0$ comes from the behavior of the integral for $s \rightarrow \infty$; the pole results from an almost on-shell particle propagating in spacetime for a long proper time. (The almost on-shell particle also propagates a long distance in spacetime along the lightcone, as one sees from the position space form of the propagator.) More generally, when one evaluates a Feynman diagram, the singularities that are associated to unitarity all arise from poles of propagators, so they all arise when some or all proper time parameters become large. So we can think of large s as the infrared or on-shell region.

In string theory, there is an immediate analog of the large s region. An open or closed string propagating for a long proper time is described by a long strip or a long tube, respectively. For examples of string worldsheets containing such a long strip or tube, see fig. 3. But one of the most fundamental facts about perturbative string theory is that there is no ultraviolet region. The roots of this statement go back nearly forty years to the original study of modular invariance in string theory [26]. In one description of a string worldsheet, it may appear that a proper time parameter becomes small in some

limit. But the same worldsheet always has an alternative description in which the proper time parameters are all bounded safely away from zero.

The most precise statement that there is no ultraviolet region in string perturbation theory is the existence of the Deligne-Mumford compactification of the moduli space of Riemann surfaces (or of super Riemann surfaces). Moduli space can be compactified by adding only limit points that correspond to $s \rightarrow \infty$; there is no need for additional limit points that would correspond to $s \rightarrow 0$ or anything else. This means that there is no potential for ultraviolet divergences in string theory; the delicate questions all involve the behavior in the infrared. It also means that in any integration by parts on moduli space – such as we have contemplated at many points in this paper, beginning with our discussion of eqn. (2.48) – the subtle issues involve the region $s \rightarrow \infty$ where one of the string states goes on-shell.

In the Deligne-Mumford compactification, one assigns a limit to a sequence of string worldsheets in which the length or a tube or strip diverges. In doing this, one exploits worldsheet conformal invariance. A string worldsheet with a long tube is conformally equivalent to a worldsheet (fig. 4) with a very narrow neck. In the case of an ordinary Riemann surface, such a worldsheet is described locally by an equation⁴⁴

$$(6.2) \quad xy = q,$$

where x is a local complex parameter on one side of the narrow neck, y is a local complex parameter on the other side, and q is a complex modulus that controls the width of the neck. The relation of this description to the “long tube” description is made simply by the change of variables or conformal mapping

$$(6.3) \quad x = e^\rho, \quad y = qe^{-\rho},$$

with

$$(6.4) \quad \rho = u + i\varphi, \quad u, \varphi \in \mathbb{R}.$$

Here φ is an angular variable of period 2π . If the coordinates x, y are valid for $|x|, |y| < 1$, then the description by ρ is good for $0 > u > -\ln(1/|q|)$. So

⁴⁴For the analogous case of a long strip, roughly speaking one takes q real and identifies x, y with \bar{x}, \bar{y} , whereupon the gluing formula of eqn. (6.2) is applicable to gluing of open-string worldsheets through a narrow neck. For more detail, see section 7.4 of [16].

for $q \rightarrow 0$, the ρ coordinate describes a tube of circumference 2π and length $\ln(1/|q|)$. In the “long tube” description, the Riemann surface seems to diverge for $q \rightarrow 0$, but the description by x and y has a limit for $q = 0$. The equation simply becomes $xy = 0$, which describes two branches, one characterized by $x = 0$ for any y , and one characterized by $y = 0$ for any x , and meeting at a singularity at $x = y = 0$. This is a rather special singularity, called a node or ordinary double point, where two branches meet.⁴⁵ The Deligne-Mumford compactification of $\mathcal{M}_{g,n}$ is achieved by allowing this type of singularity and no other. The existence of this compactification is a precise statement of the fact that in string perturbation theory there is an infrared region and no ultraviolet region. Similarly, the existence of an analogous Deligne-Mumford compactification of the moduli space of super Riemann surfaces means that there is no ultraviolet region in superstring perturbation theory. For an introduction to the Deligne-Mumford compactification, including its extensions for super Riemann surfaces and for open strings, see sections 6 and 7.4 of [16].

We will proceed as follows. In sections 6.1 and 6.2, we compute the behavior of the string measure when a Riemann surface develops a long strip or tube. This entails calculating what one may call the string propagator; in fact, with a standard gauge-fixing, string field theory leads precisely to the propagators that we will calculate (for example, see [54, 30]). The goal is to show that in all cases (open and closed bosonic strings or superstrings, including all sectors of superstrings), the string propagator has the same singular behavior – the same on-shell poles – that one would expect in a field theory with the same particles and couplings. Apart from giving a nice illustration of the machinery that we have developed up to this point, this is an essential step in proving that perturbative string scattering amplitudes have the same infrared singularities (and more generally the same singularities due to on-shell intermediate particles) that one would expect in a field theory with the same particles and couplings. The main additional step required is an analysis of the on-shell factorization of the string amplitudes, which we discuss in sections 6.4 and 6.5. Understanding what produces this factorization is also a good starting point to complete our description of the integration cycle of superstring perturbation theory; this is the topic of section 6.6. All this will prepare the ground for section 7, in which we discuss the troublesome exceptional cases of on-shell behavior in string perturbation theory: mass renormalization and massless tadpoles. These are the only possible sources of BRST anomalies for closed oriented strings, though for open and/or unoriented strings, there are additional anomalies, which we will discuss in section 9.

⁴⁵We have already discussed the singular limiting configurations for certain cases in which one branch has genus 0; see fig. 1 of section 2.5.1 and fig. 2 of section 4.4.

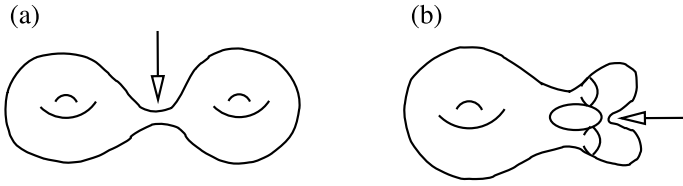


Figure 4: As explained in the text, a narrow neck in a Riemann surface – marked here by an arrow – is conformally equivalent to a long tube. The singularity or “degeneration” that occurs when the neck collapses is said to be separating in (a), and non-separating in (b).

6.1. Open-string propagators

6.1.1. Bosonic open strings We begin with the propagator of the bosonic open string.

We describe a strip S of length s and width π by real coordinates φ , u with

$$(6.5) \quad 0 \leq u \leq s, \quad 0 \leq \varphi \leq \pi.$$

We take the metric on the strip to be

$$(6.6) \quad ds^2 = du^2 + d\varphi^2.$$

Viewing u as the “time” direction, a string propagating on the strip has a conserved Hamiltonian

$$(6.7) \quad L_0 = -\frac{1}{2\pi} \int_0^\pi d\varphi T_{uu},$$

where T_{ij} is the worldsheet stress tensor. The odd-looking minus sign comes from writing the formula in Euclidean signature. Similarly the antighost field b_{ij} has a zero-mode

$$(6.8) \quad b_0 = -\frac{1}{2\pi} \int_0^\pi d\varphi b_{uu}.$$

(The ghost field c^i also has a zero-mode on the strip, but this will be less important.) It is useful to define

$$(6.9) \quad z = -i\varrho = \varphi - iu.$$

In quantization of an open string, the stress tensor and the antighost field are both free fields with left- and right-moving components. As usual, a free field on the strip is very similar to a chiral free field (purely left-moving or right-moving) on a tube obtained by gluing together two copies of the strip along their boundaries. A quick way to exhibit this is to extend φ to a real variable and extend the definition of the fields by

$$(6.10) \quad \begin{aligned} T_{zz}(u, \varphi) &= T_{\widetilde{zz}}(u, -\varphi) = T_{zz}(u, \varphi + 2\pi), \\ b_{zz}(u, \varphi) &= b_{\widetilde{zz}}(u, -\varphi) = b_{zz}(u, \varphi + 2\pi). \end{aligned}$$

The zero-modes are then

$$(6.11) \quad \begin{aligned} L_0 &= \frac{1}{2\pi} \int_0^{2\pi} d\varphi T_{zz} \\ b_0 &= \frac{1}{2\pi} \int_0^{2\pi} d\varphi b_{zz} \end{aligned}$$

The strip S has a single real modulus, the parameter $s \geq 0$. We want to learn how to integrate over s , using the general recipe of section 2.3. In this procedure, we are supposed to work on a fixed worldsheet with a metric that depends on the modulus, in contrast to the description we have given above, with a metric (6.6) that does not depend on s while the definition (6.5) of the strip S does depend on s . To convert to a description of the desired sort, we simply introduce a new “time” coordinate $t = u/s$, so that the definition of the strip becomes

$$(6.12) \quad 0 \leq t \leq 1, \quad 0 \leq \varphi \leq \pi,$$

and the metric is

$$(6.13) \quad ds^2 = s^2 dt^2 + d\varphi^2.$$

According to eqn. (2.31), in integrating over s , we must insert in the path integral a factor of

$$(6.14) \quad \begin{aligned} \Psi_s &= \frac{1}{4\pi} \int_S dt d\varphi \frac{\partial(\sqrt{g}g_{ij})}{\partial s} b^{ij} = \frac{1}{2\pi} \int_0^1 dt \int_0^\pi d\varphi s^2 b^{tt} \\ &= \frac{1}{2\pi} \int_0^1 dt \int_0^\pi d\varphi b_{uu} = \int_0^1 dt b_0 = b_0. \end{aligned}$$

In the last step, we used the fact that b_0 is time-independent, so that $\int_0^1 dt b_0 = b_0$.

We conclude that integration over the strip means integration over s with an insertion of b_0 . The path integral on the strip without any insertion is described by the operator $\exp(-sL_0)$. So integrating over s and including an insertion of b_0 , we learn that the open-string propagator is

$$(6.15) \quad \int_0^\infty ds b_0 \exp(-sL_0) = \frac{b_0}{L_0}.$$

It is also convenient to express the propagator as an integral over $q = e^{-s}$. Here q , which clearly is real and ranges from 0 to 1 for $0 \leq s < \infty$, is the open-string analog of the gluing parameter that we introduced for closed bosonic strings in (6.2). In terms of q , the open-string propagator is

$$(6.16) \quad b_0 \int_0^1 \frac{dq}{q} q^{L_0}.$$

The operator L_0 is

$$(6.17) \quad L_0 = \frac{\alpha'}{4} p^2 + N,$$

where p is the momentum and N contains the contributions of ghost and matter oscillators. For uncompactified bosonic strings, N has eigenvalues $-1, 0, 1, 2, \dots$. The condition for an open string to be on-shell is $L_0 = 0$, so the mass squared operator of the string is $M^2 = 4N/\alpha'$. Thus we can write the propagator as

$$(6.18) \quad \frac{4}{\alpha'} \frac{b_0}{p^2 + M^2}.$$

Since $b_0^2 = 0$, the operator b_0 projects the string propagation onto states annihilated by b_0 , and means (in a sense that we will state more precisely in section 6.5) that only states annihilated by b_0 propagate. The factor $4/\alpha'$ is a normalization factor that could be eliminated, if we wish, by rescaling the string coupling constant and the external vertex operators.⁴⁶ Modulo the

⁴⁶In ordinary field theory, one usually normalizes the fields so that the Feynman propagator is $1/(p^2 + m^2)$, but if one wishes, one can multiply one's fields by an arbitrary constant κ and then the propagator becomes $\kappa^2/(p^2 + m^2)$. Changing the normalization of the fields will rescale the coupling parameters and the wavefunctions of external particles. The situation is precisely the same in string theory; multiplying the propagator by a constant is equivalent to changing the string coupling constant and the normalization of vertex operators.

normalization factor and the projection on states annihilated by b_0 , the open-string propagator is just what one would guess from field theory.

The precise use of this propagator is as follows. Suppose that as in fig. 3(a), the strip is attached to Riemann surfaces Σ_ℓ and Σ_r at its left and right ends. We want to keep Σ_ℓ and Σ_r fixed and integrate over the length s of the strip.⁴⁷ Keeping Σ_ℓ and Σ_r fixed, the path integral on those surfaces generates quantum states ψ_ℓ and ψ_r . The path integral on the full surface Σ is obtained by taking ψ_ℓ and ψ_r as initial and final states for the propagation on the strip. The propagation on the strip is described by the propagator (6.18), so the path integral on Σ after integrating over s is

$$(6.19) \quad Z_{\Sigma;s} = \left\langle \psi_\ell \left| \frac{b_0}{L_0} \right| \psi_r \right\rangle.$$

A complete calculation will of course include also integrating over the moduli of Σ_ℓ and Σ_r ; in fact integrating just over s is only sensible in the region where s is large. What the computation that we have just done is good for is to isolate the singularities that arise when an open string in a particular channel goes on shell. As we have just seen, these singularities are precisely the simple poles at $p^2 + m^2 = 0$ that one would expect in field theory.

One can analyze in a similar fashion the contribution of a region in which any number of open-string states simultaneously go on-shell. We get a propagator b_0/L_0 in each channel, leading to a standard $1/(p^2 + m^2)$ pole in every channel. Presently, we will obtain similar results for open superstrings and for all the closed-string theories. Since the usual infrared and on-shell singularities of Feynman amplitudes come from the poles of propagators, this makes it more or less obvious that string theory will have the same infrared and on-shell singularities as field theory.

To be more precise, this is more or less obvious once one incorporates an important refinement, the Feynman $i\epsilon$, which will be described in section 6.3. We go in more detail about the residue of the pole of the propagator in section 6.4.

6.1.2. The NS propagator Now we consider the NS sector of open superstrings.

On the worldsheet of an open superstring, in addition to bosonic coordinates u and φ , there are holomorphic and antiholomorphic odd coordinates

⁴⁷To be more fastidious, we include antighost insertions on Σ_ℓ and Σ_r , to make the path integral non-zero, but do not integrate over their moduli.

θ and $\tilde{\theta}$. At the endpoints $\varphi = 0, \pi$ of a strip S , they are glued together, with some choices of signs. In the NS sector, the signs are opposite at the two ends.

$$(6.20) \quad \tilde{\theta} = \begin{cases} \theta & \text{if } \varphi = 0 \\ -\theta & \text{if } \varphi = \pi. \end{cases}$$

The minus sign can be moved from one end of the strip to the other by a change of variables $\tilde{\theta} \rightarrow -\tilde{\theta}$.

Right-moving massless free fields along the string depend on θ (as well as z) and left-moving ones depend on $\tilde{\theta}$ (and \tilde{z}). By extending the range of φ beyond the interval $0 \leq \varphi \leq \pi$, one can combine these two types of modes to a chiral field that depends only on φ and θ , subject to

$$(6.21) \quad \varphi \rightarrow \varphi + 2\pi, \quad \theta \rightarrow -\theta.$$

The minus sign, which comes about because the signs in (6.20) are opposite, is the reason that quantization of open strings with the boundary conditions (6.20) leads to what is usually called the NS sector.

Now let us discuss the propagator in the NS sector. In fact, the minus sign in the boundary conditions makes things simple. The strip in the NS sector has no odd moduli. An odd modulus would arise as always from a mode of the gravitino field that cannot be gauged away. To be more exact, for open superstrings we have left and right gravitino fields χ_z^θ and $\chi_{\tilde{z}}^{\tilde{\theta}}$ that are glued together on the boundary by analogy with (6.20):

$$(6.22) \quad \chi_{\tilde{z}}^{\tilde{\theta}} = \begin{cases} \chi_z^\theta & \text{if } \varphi = 0 \\ -\chi_z^\theta & \text{if } \varphi = \pi. \end{cases}$$

They fit together to a single gravitino field χ_z^θ that is defined for all φ and obeys

$$(6.23) \quad \chi_z^\theta(u, \varphi + 2\pi) = -\chi_z^\theta(u, \varphi).$$

The minus sign ensures that χ_z^θ can be gauged away by

$$(6.24) \quad \chi_z^\theta \rightarrow \chi_z^\theta + \partial_{\tilde{z}} y^\theta,$$

as one may prove via a Fourier expansion with respect to φ , using the absence of a zero-mode. Thus the strip has no odd moduli. Similarly, the minus sign in (6.21) means that the mode expansion of the commuting ghost fields β

and γ involves modes β_r, γ_r with $r \in \mathbb{Z} + 1/2$; in particular, there are no zero-modes.

Accordingly, the derivation of the propagator is almost straightforward. We only have to integrate over one even modulus s , the “length” of the strip. The same calculation as in section 6.1.1 shows that the integral gives $\int_0^\infty ds b_0 \exp(-sL_0) = b_0/L_0$.

To this we only have to add one detail and answer one question. What we need to add is the GSO projection [52]. The GSO projection, which removes the tachyon from superstring theory, reflects the fact that the worldsheet path integral of superstring theory includes a sum over worldsheet spin structures. In general, this sum cannot be carried out independently of the integral over supermoduli space, because the spin structure is part of the structure of a super Riemann surface and there is no way to sum over the spin structures of a super Riemann surface except as part of the integral over all of the even and odd moduli. However, in the region $s \rightarrow \infty$ that gives rise to the on-shell singularities of a superstring amplitude, a very partial sum over spin structures makes sense: this is the sum over pairs of spin structures that differ only by a minus sign “twist” in the u direction. For a detailed explanation of this, see section 6.2.3 of [16]; the basic idea is also indicated in section 6.4.5 below. The sum over pairs of spin structures that differ only by this particular twist gives the GSO projection Π_{GSO} that removes half the states of the NS sector, including the tachyon. With this included, the open superstring propagator in the NS sector is

$$(6.25) \quad \frac{b_0}{L_0} \Pi_{\text{GSO}}.$$

We also need to answer a question: What are the states that are propagated by the propagator (6.25)? To be more precise, what is the picture number of these states?

6.1.3. Picture number and the propagator Before getting to picture number, let us remember that for an ordinary anomaly-free or anomalous symmetry such as ghost number, the answer to a question like the last one is that states of all possible quantum numbers can propagate down the strip. To illustrate this in a precise context, let us glue together the two ends of the strip at $u = 0$ and s to make an annulus (fig. 5). In the gluing, we insert some operator \mathcal{O} . The path integral on the annulus with this insertion is

$$(6.26) \quad \int_0^\infty ds \text{Tr } \mathcal{O} b_0 \exp(-sL_0) \Pi_{\text{GSO}}.$$

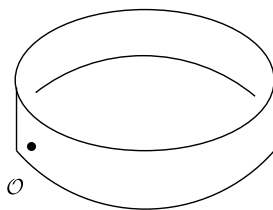


Figure 5: An annulus, built by gluing together the ends of a long strip, with an operator insertion \mathcal{O} .

The trace is a sum over all states, including states with all possible values of the ghost number and other quantum numbers.

We have already explained at the end of section 3.6.3 one reason that picture number cannot be interpreted as a symmetry, anomalous or not. Here we are about to come upon another such reason. In explaining this, we first consider the case that \mathcal{O} is independent of β and γ , so that β and γ are decoupled from the other fields in (6.26) and the $\beta\gamma$ path integral that we want is simply the partition function. We claim that the $\beta\gamma$ partition function on an annulus in the NS sector is a sum only over states of picture number -1 , with no sum over the picture number of the states propagating around the annulus.

We will explain this in two ways. First, this is the only answer that makes any sense; otherwise the annulus path integral would diverge, since L_0 would be unbounded below. The ghost vacuum with picture number $|q\rangle$ is defined by the conditions

$$(6.27) \quad \begin{aligned} \beta_r |q\rangle &= 0, & r > -q - 3/2 \\ \gamma_r |q\rangle &= 0, & r \geq q + 3/2. \end{aligned}$$

For any value of q other than -1 , the state $|q\rangle$ is annihilated by some β_r or γ_r with $r < 0$, and then the states $\beta_{-r}^k |q\rangle$ or $\gamma_{-r}^k |q\rangle$, $k = 0, 1, 2, \dots$ are linearly independent and have L_0 unbounded below. So for $q \neq -1$, the trace in (6.26) that is supposed to give the annulus path integral does not make any sense, at least for generic \mathcal{O} . For $q = -1$, there is no problem, since (6.27) says that $|q\rangle$ is annihilated by what one would want to call the annihilation operators,

$$(6.28) \quad \beta_r |q\rangle = \gamma_r |q\rangle = 0, \quad r > 0.$$

For a deeper explanation, we should ask how we define the $\beta\gamma$ path integral so that it does make sense. We explore this question more fully in sec-

tion 10. However, the basic idea is that the $\beta\gamma$ path integral without operator insertions is understood as a bosonic Gaussian integral that equals $1/\det \tilde{\partial}_{\beta\gamma}$, where $\tilde{\partial}_{\beta\gamma}$ is the kinetic operator of the $\beta\gamma$ system. Moreover $\det \tilde{\partial}_{\beta\gamma}$ (defined with any standard type of regularization) is a definite function of s and therefore the question, “What is the picture number of the states that propagate around the annulus?” is going to have a definite answer; we are not free to make any choice. As a shortcut to determine the answer, let us note that if we replace β and γ with fields β^* and γ^* of opposite statistics (and thus fermionic) but otherwise with the same Lagrangian and boundary conditions, this simply replaces the path integral of the $\beta\gamma$ system by its inverse. So the $\beta^*\gamma^*$ path integral is $\det \tilde{\partial}_{\beta\gamma}$. On the other hand, we certainly know how to interpret the $\beta^*\gamma^*$ path integral on an annulus as a sum over quantum states. On an annulus of circumference s in the time direction, taking the fields β^* and γ^* to be periodic in the s direction (this gives an insertion of $(-1)^F$, the operator that counts fermions mod 2), and setting $q = e^{-s}$, the $\beta^*\gamma^*$ path integral can be evaluated in a standard fashion as a trace in the fermion Fock space:

$$(6.29) \quad \det \tilde{\partial}_{\beta\gamma} = \text{Tr} (-1)^F \exp(-sL_{0;\beta^*\gamma^*}) = q^f \prod_{r=\frac{1}{2}, \frac{3}{2}, \dots} (1 - q^r).$$

(There is no analog of picture number for fermions, so there is no choice to be made in writing this formula.) Here $L_{0;\beta^*\gamma^*}$ is the Hamiltonian of the $\beta^*\gamma^*$ system and f is the ground state energy of the $\beta^*\gamma^*$ system in the NS sector. The $\beta\gamma$ path integral on the annulus is therefore the inverse of this, or

$$(6.30) \quad \frac{1}{\det \tilde{\partial}_{\beta\gamma}} = q^{-f} \prod_{r=\frac{1}{2}, \frac{3}{2}, \dots} \frac{1}{1 - q^r}.$$

As expected, we see that the energy levels of the $\beta\gamma$ system are bounded below; the $\beta\gamma$ path integral on the annulus is q^{-f} times a series in positive powers of q . Moreover, we see a standard partition function for a Fock space of bosons, confirming that it does make sense to define a bosonic Gaussian integral as the inverse of a corresponding fermionic Gaussian integral. Finally, we see that the ground state energy of the $\beta\gamma$ system is precisely minus that of the $\beta^*\gamma^*$ system. According to [1], this is true precisely at picture number -1 .

Now let us restore the operator \mathcal{O} in eqn. (6.26). Part of what we have said is valid in general. The $\beta\gamma$ path integral defined as a generalized Gaussian integral will always lead to a definite answer, with no freedom in the choice of

picture number. Moreover, if \mathcal{O} is simply a polynomial in β and γ and their derivatives, its inclusion does not really affect the above reasoning. (The $\beta\gamma$ path integral with insertions of elementary fields is analyzed in section 10.2.1.)

However, in general, it is possible to find an \mathcal{O} such that the states that propagate around the annulus have non-canonical picture number. A simple example with this property is $\mathcal{O} = \delta(\beta_{-1/2})\delta(\gamma_{1/2})$. We have taken a product of picture-raising and picture-lowering operators, so that the path integral with insertion of \mathcal{O} remains sensible. The operator \mathcal{O} projects onto states that are annihilated by $\beta_{-1/2}$. The Fock vacuum $|q\rangle$ with picture number $q = -1$ is not annihilated by $\beta_{-1/2}$, but the corresponding Fock vacuum with $q = 0$ does have this property.

It can be shown using methods of section 10 that the $\beta\gamma$ path integral on an annulus with insertion of \mathcal{O} computes

$$(6.31) \quad \text{Tr } \mathcal{O} q^{L_0}$$

where the trace is taken in the $\beta\gamma$ Fock space with $q = 0$ (we denote the Fock vacuum in this Fock space as $|0\rangle$). This claim may seem to present a paradox: for $q = 0$, L_0 is not bounded below, so how can the trace converge? In fact, L_0 is unbounded below for $q = 0$ because the oscillator $\gamma_{1/2}$ lowers L_0 by $1/2$. But this causes no problem in the trace $\text{Tr } \mathcal{O} q^{L_0}$, since the potentially dangerous states $\gamma_{1/2}^p |0\rangle$ with $p > 0$ (and any other states constructed with the negative energy creation operator $\gamma_{1/2}$) are annihilated by $\mathcal{O} = \delta(\beta_{-1/2})\delta(\gamma_{1/2})$, because of the second factor. Such an apparently “lucky” rescue occurs in any sensible $\beta\gamma$ path integral that receives contributions from NS or R states whose picture number is such that L_0 is unbounded below.

So it is not true in general that in an arbitrary sensible $\beta\gamma$ path integral, the $\beta\gamma$ propagator in the NS sector only propagates states of picture number -1 . But this is true in the context of superstring perturbation theory. See section 6.4.6 for further discussion.

6.1.4. The Ramond propagator The Ramond sector of open superstrings is obtained by dropping some minus signs in the gluing relations. Thus, eqn. (6.20) for the identification of the fermionic coordinates at the end of the strip becomes

$$(6.32) \quad \tilde{\theta} = \begin{cases} \theta & \text{if } \varphi = 0 \\ \theta & \text{if } \varphi = \pi, \end{cases}$$

with the same sign at both ends. The left- and right-moving supercurrents can accordingly be combined to a holomorphic supercurrent $S_{z\theta}(u, \varphi)$ that

is invariant under $\varphi \rightarrow \varphi + 2\pi$. As a result, the odd generators G_r of the super-Virasoro algebra are graded by integers, and in particular, the algebra contains a zero-mode that commutes with the Hamiltonian L_0 :

$$(6.33) \quad G_0 = \frac{1}{2\pi} \int_0^{2\pi} d\varphi S_{z\theta}(0, \varphi).$$

Similarly we can combine left- and right-moving antighost fields to a single holomorphic field $\beta_{z\theta}(u, \varphi)$ that is also invariant under $\varphi \rightarrow \varphi + 2\pi$, so that it has a zero-mode:

$$(6.34) \quad \beta_0 = \frac{1}{2\pi} \int_0^{2\pi} d\varphi \beta_{z\theta}(u, \varphi).$$

Finally, the gravitino field $\chi_z^\theta(u, \varphi)$ is now a periodic variable,

$$(6.35) \quad \chi_z^\theta(u, \varphi + 2\pi) = \chi_z^\theta(u, \varphi).$$

The consequence of this last statement is that the gravitino field in the strip has a zero-mode that should be treated as a modulus. This can be represented by the constant gravitino field

$$(6.36) \quad \chi_z^\theta = \eta,$$

where η is a constant anticommuting parameter. To gauge this mode away by

$$(6.37) \quad \chi_z^\theta \rightarrow \chi_z^\theta + \partial_z y^\theta,$$

where y^θ should be invariant under $\varphi \rightarrow \varphi + 2\pi$, we would have to let y^θ grow in the u direction. Since u ranges from 0 to s , a gauge transformation that grows with u does not behave well for $s \rightarrow \infty$, where we are trying to extract on-shell poles. This explains at least heuristically why one should treat the constant mode (6.36) as a modulus and not try to gauge it away. A precise explanation involves the Deligne-Mumford compactification; see the discussion of eqn. (6.76) below.

It is straightforward⁴⁸ to integrate over η using the procedure of eqn. (3.88) or (3.89). In fact, we can do this using the bosonic coordinates u, φ

⁴⁸Since there is only one odd modulus associated to the strip, none of the subtleties of integrating over odd moduli come into play. As explained in sections 3.3.2, 3.4.1, and 3.5, these subtleties become relevant when there are two or more odd moduli.

without even transforming to $t = u/s$. Replacing $\chi_z^{(\sigma)\theta}$ by 1, the factor (3.87) in the path integral becomes

$$(6.38) \quad \exp\left(-\frac{1}{2\pi} \int_0^{2\pi} d\varphi \int_0^s du (\eta S_{z\theta} + d\eta \beta_{z\theta})\right) = \exp(-s(\eta G_0 + d\eta \beta_0)).$$

The integral over η and $d\eta$ gives

$$(6.39) \quad G_0 \delta(\beta_0).$$

This is the only part of the evaluation of the propagator that is special to the Ramond sector. The integral over s gives the usual result b_0/L_0 , and the sum over the possible twists of the fermions in the u direction gives the GSO projection Π_{GSO} . So the Ramond sector propagator for open strings is

$$(6.40) \quad \frac{b_0 \delta(\beta_0) \Pi_{\text{GSO}} G_0}{L_0}.$$

Clearly, only states that are annihilated by b_0 and β_0 and are invariant under Π_{GSO} propagate down the strip. To understand the result (6.40) more fully, we recall that in the Ramond sector, $G_0^2 = L_0$, so we can write the propagator as

$$(6.41) \quad \frac{b_0 \delta(\beta_0) \Pi_{\text{GSO}}}{G_0}.$$

The operator G_0 is the string theory analog of the Dirac operator of field theory, and so $1/G_0$ (with some states projected out by the numerator) is the natural string theory analog of the usual Dirac propagator. For example, for massless Ramond states, G_0 reduces to $\frac{1}{2}(\alpha')^{1/2} \Gamma \cdot p$ (where Γ^I are spacetime gamma matrices and p_I is the momentum), which apart from the factor $\frac{1}{2}(\alpha')^{1/2}$ is the massless Dirac operator in momentum space as usually normalized. In general, $G_0 = \frac{1}{2}(\alpha')^{1/2} \Gamma \cdot p + N_{\text{R}}$, where N_{R} is the oscillator contribution to G_0 and $2(\alpha')^{-1/2} N_{\text{R}}$ is the mass operator for open strings in the Ramond sector.

As in section 6.1.2, we should now ask what is the picture number of the Ramond states whose propagation is described by this propagator. For a simple example, let us glue together the two ends of the strip with insertion of some operator \mathcal{O} to make an annulus. Thus the $\beta\gamma$ path integral, if it can be interpreted in terms of a sum over states, computes $\text{Tr } \mathcal{O} \delta(\beta_0) q^{L_0}$. We cannot simply take $\mathcal{O} = 1$, since the criterion of eqn. (3.107) for a sensible $\beta\gamma$ path integral would not be satisfied. We need an operator of picture number -1 ,

such as $\mathcal{O} = \delta(\gamma_0)$. So we want to calculate a $\beta\gamma$ path integral on the strip with insertion of $\delta(\gamma_0)\delta(\beta_0)$. What is the picture number of the states that contribute to this path integral? The obvious candidates are $q = -1/2$ and $-3/2$, since those are the values of q at which L_0 is bounded below; this is clear from the definition (6.27) of the ghost vacuum $|q\rangle$ with picture number q . If we treat any of the β_n or γ_n with $n > 0$ as a creation operator, the trace $\text{Tr} \delta(\gamma_0)\delta(\beta_0)q^{L_0}$ will certainly diverge. And the explicit factor of $\delta(\beta_0)$ in the propagator shows that the states that are propagated are annihilated by β_0 . Putting these facts together, we find that in this minimal example, the states whose propagation is described by the Ramond propagator have canonical picture number $q = -1/2$. As in the NS case, one can construct sensible $\beta\gamma$ path integrals that receive contributions from states of noncanonical picture number, but these are not the path integrals that arise in superstring perturbation theory.

6.2. Closed-string propagators

Now we turn to closed-string propagators.

6.2.1. Closed bosonic strings A long strip describing an almost on-shell open string has a single real modulus s , also usefully parametrized by $q = \exp(-s)$. For closed strings, s and q become complexified. For closed strings, the gluing parameter q appears in formula (6.2) that describes the gluing of two branches of a Riemann surface. But q is naturally a complex parameter, one of the moduli of the Riemann surface. So for closed strings, s must combine with a second real parameter to make a complex modulus.

Concretely, this second real modulus is obtained by “cutting” the long tube T in fig. 3(b) to separate the two ends, and then rotating one piece relative to the other by an angle α before gluing them back together. Assuming that this rotation cannot be extended as a symmetry of either Σ_ℓ or Σ_r , α is a modulus of Σ . In the Deligne-Mumford compactification of $\mathcal{M}_{g,n}$, one only considers decompositions such that Σ_ℓ and Σ_r (with their punctures deleted) both have negative Euler characteristic, and then it is automatically true that the symmetry does not extend and α is a modulus. This is also almost always true for open and/or unoriented string worldsheets, though the exceptions turn out to be important.⁴⁹

⁴⁹ The exceptional cases are that Σ_r (or Σ_ℓ) is a disc or a copy of $\mathbb{R}P^2$ with only one puncture (the node). These examples will be important in studying anomalies in section 9.

In one description of this situation, we can describe the tube T by a flat metric that does not depend on α

$$(6.42) \quad ds^2 = du^2 + d\varphi^2, \quad 0 \leq u \leq s, \quad 0 \leq \varphi \leq 2\pi.$$

φ is now an angular variable. For closed strings, we have separate holomorphic and antiholomorphic Virasoro algebras, each with its own zero-mode:

$$(6.43) \quad \begin{aligned} L_0 &= -\frac{1}{2\pi} \int_0^{2\pi} d\varphi T_{zz} \\ \tilde{L}_0 &= -\frac{1}{2\pi} \int_0^{2\pi} d\varphi T_{\tilde{z}\tilde{z}}. \end{aligned}$$

Similarly, there are separate holomorphic and antiholomorphic antighost zero-modes. For closed bosonic strings, these are

$$(6.44) \quad \begin{aligned} b_0 &= -\frac{1}{2\pi} \int_0^{2\pi} d\varphi b_{zz} \\ \tilde{b}_0 &= -\frac{1}{2\pi} \int_0^{2\pi} d\varphi b_{\tilde{z}\tilde{z}}. \end{aligned}$$

To the tube T , we want to associate the real modulus s that already appears in the metric (6.42), and another real modulus α . If Σ is obtained by gluing T at its ends onto surfaces Σ_ℓ and Σ_r , then the extra modulus can be obtained by gluing T onto Σ_ℓ in a way that is independent of α , but rotating it by an angle α before gluing it onto Σ_r .

This is a simple description, but it has two drawbacks. It does not explain in what sense α should be associated to T , rather than to the whole surface Σ . And it does not lend itself to computing the measure for the string path integral via formulas such as eqn. (2.31). For that purpose, we want to describe Σ as a fixed two-manifold, independent of α , but with a metric that depends on α . We can easily get such a description by replacing φ with a new angular coordinate

$$(6.45) \quad \hat{\varphi} = \varphi - \alpha f(u),$$

where $f(u)$ is any smooth function with $f(0) = 0$, $f(s) = 1$. If T is described by the coordinates $u, \hat{\varphi}$, then its metric becomes

$$(6.46) \quad ds^2 = du^2 + d(\hat{\varphi} + \alpha f(u))^2.$$

Now the definition of the space T and the gluing recipe are both independent of α , which appears only in the metric of T .

It is now straightforward to apply the recipe of eqn. (2.31). The integral over α must be accompanied by an insertion of

$$(6.47) \quad \Psi_\alpha = \frac{1}{4\pi} \int_0^{2\pi} d\varphi \int_0^s du \frac{\partial(\sqrt{g}g_{ij})}{\partial\alpha} b^{ij}.$$

We can evaluate Ψ_α in a simple way, since the α dependence of the metric is equivalent to a change of coordinates. It follows from this that

$$(6.48) \quad \frac{\partial g_{ij}}{\partial\alpha} = D_i v_j + D_j v_i$$

where v is a vector field. In fact,

$$(6.49) \quad v = f(u) \frac{\partial}{\partial\varphi},$$

or in other words $v^u = 0$, $v^\varphi = f(u)$. (Thus, $\exp(\alpha v)$ is the transformation from coordinates s , $\hat{\varphi}$ back to s , φ .) It follows that

$$(6.50) \quad \Psi_\alpha = \frac{1}{4\pi} \int_0^{2\pi} d\varphi \int_0^s du \sqrt{g} b^{ij} (D_i v_j + D_j v_i).$$

Integrating by parts and using the equation of motion $D_i b^{ij} = 0$, we find that we can evaluate Ψ_α as a surface term at $u = s$:

$$(6.51) \quad \Psi_\alpha = -\frac{1}{2\pi} \int_0^{2\pi} d\varphi b_{u\varphi}(\varphi, s) = b_0 - \tilde{b}_0.$$

The computation of Ψ_s , the ghost insertion that accompanies the integration over s , proceeds just as in eqn. (6.14), except that for closed strings, in the final step, we get $-\frac{1}{2\pi} \int_0^{2\pi} d\varphi b_{uu} = b_0 + \tilde{b}_0$. So $\Psi_s \Psi_\alpha = 2\tilde{b}_0 b_0$. The operator that propagates a closed string through an imaginary time s and rotates it by an angle α is

$$(6.52) \quad \exp\left(-s(L_0 + \tilde{L}_0)\right) \exp\left(-i\alpha(L_0 - \tilde{L}_0)\right).$$

The closed bosonic string propagator is then

$$2\Psi_s \Psi_\alpha \int_0^\infty ds \int_0^{2\pi} d\alpha \exp\left(-s(L_0 + \tilde{L}_0)\right) \exp\left(-i\alpha(L_0 - \tilde{L}_0)\right)$$

$$(6.53) \quad = 4\pi\tilde{b}_0b_0\delta_{L_0-\tilde{L}_0} \int_0^\infty ds \exp(-s(L_0 + \tilde{L}_0)) = \frac{2\pi\tilde{b}_0b_0\delta_{L_0-\tilde{L}_0}}{L_0}.$$

Here the integral over α has given a factor $\delta_{L_0-\tilde{L}_0}$ that ensures that only states annihilated by $L_0 - \tilde{L}_0$ propagate through the tube. The integral over s gives a factor $1/L_0$ (or equivalently $1/\tilde{L}_0$), which contains the expected closed-string poles.

The closed-string gluing parameter is $q = \exp(-(s + i\alpha))$. Its antiholomorphic counterpart is $\tilde{q} = \exp(-(s - i\alpha))$. We can write the closed bosonic string propagator as

$$(6.54) \quad \tilde{b}_0b_0 \int_{|q|\leq 1} \frac{d^2q}{|q|^2} q^{L_0} \tilde{q}^{\tilde{L}_0}.$$

6.2.2. Closed superstrings Closed superstring propagators can be obtained by combining the constructions that we have explained so far.

For the heterotic string, we have to consider the NS and Ramond sectors. In the NS sector, the only moduli of a long tube are the parameters s and α that we already have considered. The derivation of the propagator is precisely the same as in section 6.2.1, except that as in section 6.1.2, we have to include the GSO projection, which comes from summing over fermionic twists in the u direction. So the NS sector propagator of the heterotic string is

$$(6.55) \quad \frac{2\pi\tilde{b}_0b_0\delta_{L_0-\tilde{L}_0} \Pi_{\text{GSO}}}{L_0}.$$

For the Ramond sector of the heterotic string, there is also a gravitino mode. Just as in section 6.1.4, integrating over the corresponding odd modulus gives a factor of $\delta(\beta_0)G_0$. So the Ramond sector propagator is

$$(6.56) \quad \frac{2\pi\tilde{b}_0b_0\delta(\beta_0)\delta_{L_0-\tilde{L}_0} G_0 \Pi_{\text{GSO}}}{L_0} = \frac{2\pi\tilde{b}_0b_0\delta(\beta_0)\delta_{L_0-\tilde{L}_0} \Pi_{\text{GSO}}}{G_0}.$$

Just as for open superstrings, the projection operators in the numerator place some restrictions on what classes of states can propagate, and the factors $1/L_0$ and $1/G_0$ reproduce the poles that one would expect in field theory.

The propagator is one subject for which the generalization from the heterotic string to Type II superstrings merits some comment. The left- and right-movers can independently be placed in the NS or R sector, so overall there are four sectors, namely NS-NS, NS-R, R-NS, and R-R. In each case, the

sum over fermionic twists in the u direction can be carried out separately for holomorphic and antiholomorphic degrees of freedom, giving separate GSO projections Π_{GSO} and $\tilde{\Pi}_{\text{GSO}}$ for the two types of mode. Apart from this, in the NS-NS sector, the derivation is the same as for closed bosonic strings and the propagator is

$$(6.57) \quad \frac{2\pi\tilde{b}_0b_0\delta_{L_0-\tilde{L}_0}\Pi_{\text{GSO}}\tilde{\Pi}_{\text{GSO}}}{L_0}.$$

In the R-NS and NS-R sectors, one also has a holomorphic or antiholomorphic gravitino mode. Integration over the corresponding odd modulus gives the familiar factors $\delta(\beta_0)G_0$ or $\delta(\tilde{\beta}_0)\tilde{G}_0$, so the propagator is

$$(6.58) \quad \frac{2\pi\tilde{b}_0b_0\delta_{L_0-\tilde{L}_0}\delta(\beta_0)G_0\Pi_{\text{GSO}}\tilde{\Pi}_{\text{GSO}}}{L_0}$$

or

$$(6.59) \quad \frac{2\pi\tilde{b}_0b_0\delta_{L_0-\tilde{L}_0}\delta(\tilde{\beta}_0)\tilde{G}_0\Pi_{\text{GSO}}\tilde{\Pi}_{\text{GSO}}}{L_0}.$$

Finally, in the R-R sector, there is both a holomorphic gravitino mode and an antiholomorphic one. Integrating over the corresponding odd moduli gives a factor of $\delta(\beta_0)G_0$ and also a factor of $\delta(\tilde{\beta}_0)\tilde{G}_0$, and the propagator is

$$(6.60) \quad \frac{2\pi\tilde{b}_0b_0\delta(\beta_0)G_0\delta(\tilde{\beta}_0)\tilde{G}_0\delta_{L_0-\tilde{L}_0}\Pi_{\text{GSO}}\tilde{\Pi}_{\text{GSO}}}{L_0}.$$

The only case that really requires discussion is the R-R propagator; the meaning of the factor $G_0\tilde{G}_0$ in the numerator may be unclear. There is no problem of principle. The factor of $1/L_0$ gives the expected poles for on-shell bosons. The factors G_0 and \tilde{G}_0 in the numerator are nonsingular on-shell, so in matching to field theory expectations, those factors could be absorbed in the couplings of the R-R fields rather than regarded as part of the propagator. However, it is possible to learn more by considering the massless states in the R-R sector. For simplicity, we do this for uncompactified superstrings in \mathbb{R}^{10} . The gauge-invariant field strength of a massless R-R field can be regarded as a bispinor $\phi_{\alpha\beta}$, where α and β are spinor indices in \mathbb{R}^{10} (one of which comes by quantizing holomorphic degrees of freedom of the string, and one by quantizing antiholomorphic degrees of freedom). The field $\phi_{\alpha\beta}$ is subject to a chirality projection on each index (coming from the GSO projections), but

this will not affect our remarks here. From a field theory point of view, treating the R-R fields as free fields, the two-point function of $\phi_{\alpha\beta}$ at momentum p is

$$(6.61) \quad \langle \phi_{\alpha\beta}(p)\phi_{\alpha'\beta'}(-p) \rangle = \frac{(\Gamma \cdot p)_{\alpha\alpha'}(\Gamma \cdot p)_{\beta\beta'}}{p^2} + \text{constant}.$$

Here $\Gamma \cdot p$ is the Dirac operator in momentum space. (Eqn. (6.61) has been used in computing the gravitational anomalies of R-R fields; see eqn. (48) of [57].) This agrees, up to an inessential constant factor, with the low energy limit of the massless propagator (6.60), since for massless R-R states, $G_0 = \frac{1}{2}\alpha'^{1/2}(\Gamma \cdot p)_{\alpha\alpha'}$ and $\tilde{G}_0 = \frac{1}{2}\alpha'^{1/2}(\Gamma \cdot p)_{\beta\beta'}$.

We conclude that for massless states, the R-R propagator for Type II superstrings matches the field theory two-point function of the R-R field strength, not the two-point function of the R-R gauge field. Hence in perturbation theory, for oriented closed superstrings, R-R fields couple only via their field strength. In particular they decouple at zero momentum, ensuring that there is no R-R analog of the subtleties involving massless NS tadpoles that we will explore beginning in section 7. For open and/or unoriented superstrings, there is a more complicated story that we will study in section 9.

6.3. The Feynman $i\epsilon$

Our discussion of the propagators is so far missing a crucial detail: the Feynman $i\epsilon$.

6.3.1. The propagator in lorentz signature In Euclidean signature, p^2 is positive-definite, and the propagator $1/(p^2 + m^2)$ is singular only at $p = 0$. In Lorentz signature – where we ultimately must work in order to compute scattering amplitudes – p^2 is no longer positive-definite and the propagator has a pole on-shell at $p^2 + m^2 = 0$. The appropriate treatment of this pole was explained long ago by Feynman; it is essential in Lorentz signature to include the Feynman $i\epsilon$, replacing $1/(p^2 + m^2)$ by $1/(p^2 + m^2 - i\epsilon)$, where ϵ is an infinitesimal positive quantity, and one takes the limit $\epsilon \rightarrow 0$ at the end of any computation. The Feynman $i\epsilon$, in other words, is a recipe to avoid the pole at $p^2 + m^2 = 0$.

The Feynman $i\epsilon$ similarly must be incorporated in any string theory computation. How to do this has been explained in [55, 56]. The latter reference contains a detailed explanation in the spirit of the present paper; we will be much more brief here.

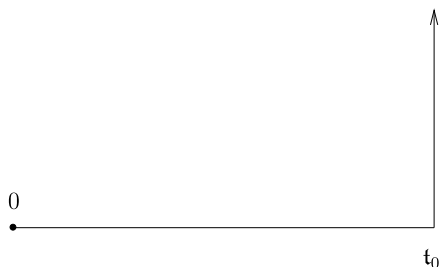


Figure 6: An integration contour that is useful in generalizing the Feynman $i\epsilon$ to string theory.

We start with bosonic open strings. Upon setting $q = e^{-\mathbf{t}}$, the integral in (6.16) becomes

$$(6.62) \quad \frac{1}{p^2 + m^2} = \int_0^\infty dt \exp(-\mathbf{t}(p^2 + m^2)).$$

Here \mathbf{t} is a Euclidean Schwinger parameter. We can think of it as a proper time parameter in Euclidean signature. To get a Lorentz signature propagator, we should integrate over a proper time parameter in Lorentz signature. One way to do this is to set $\mathbf{t} = i\tau$ and integrate over real positive τ . The integral becomes oscillatory and needs a convergence factor $\exp(-\epsilon\tau)$, where ϵ is taken to zero at the end. We get the Feynman propagator

$$(6.63) \quad \frac{1}{p^2 + m^2 - i\epsilon} = i \int_0^\infty d\tau \exp(-i\tau(p^2 + m^2) - \epsilon\tau).$$

What can be the analog of this in string theory? The open string modulus q is naturally real, as is $\mathbf{t} = -\log q$. What is worse, q is only defined when it is small; equivalently, in the context of string theory, \mathbf{t} is only defined when it is large. What can it mean to make \mathbf{t} imaginary?

A procedure that is valid in field theory and is closer to what we can do in string theory is to integrate over real \mathbf{t} from 0 up to some very large value t_0 and then to continue in the positive imaginary direction (fig. 6). This procedure, which gives the right answer in field theory, can be generalized to string theory if we interpret it correctly. We will give just an outline of how this goes, referring to the above-cited papers for more detail.

Let Σ_0 be an open-string worldsheet, that is, a Riemann surface with boundary, and let Γ be the moduli space of conformal structures on Σ . Σ_0

has a closed oriented double cover Σ . Let \mathcal{M} be the moduli space of complex structures on Σ .

Then Γ is a middle-dimensional real cycle in \mathcal{M} . More specifically, Γ is a component of the fixed point set of an antiholomorphic involution of \mathcal{M} . (Such matters are discussed in section 7 of [16].) This implies in particular that we can regard \mathcal{M} as a complexification of Γ . Likewise any covering space \mathcal{T} of \mathcal{M} is a complexification of Γ . Saying that \mathcal{M} or \mathcal{T} is a complexification of Γ means that, locally, a real analytic function on Γ can be analytically continued to a holomorphic function on \mathcal{M} or \mathcal{T} .

In particular, near an open-string degeneration, one can define the real, positive function q on Γ ; likewise we can define $\mathfrak{t} = -\log q$. To be more precise, these functions can be defined when q is sufficiently small or when \mathfrak{t} is sufficiently large. Moreover, \mathfrak{t} can be defined as a real-analytic function on Γ , so it can be analytically continued to a holomorphic function on \mathcal{M} . This function is well-defined for sufficiently large $\text{Re } \mathfrak{t}$. As is explained in [56], one must replace \mathcal{M} by a cover thereof to make $\text{Im } \mathfrak{t}$ single-valued. Once this is done, the integration contour in fig. 6 makes sense in string theory.

Let us explain concretely what this means for the special case that Σ_0 is an annulus without punctures. The moduli space of conformal structures on Σ_0 is a half-line parametrized by a positive parameter \mathfrak{t} (one can think of Σ_0 as an annulus of width π and circumference \mathfrak{t}). Thus in this special case, \mathfrak{t} is naturally-defined for all positive values, not just when it is large. The oriented double cover of Σ_0 is a genus 1 Riemann surface Σ . The Teichmuller space \mathcal{T} of Σ (which is a cover of its moduli space \mathcal{M}) is a copy of the upper half-plane, parametrized by a complex variable τ . If Σ is the oriented double cover of an annulus Σ_0 , then τ is imaginary and the relation between τ and \mathfrak{t} is $\tau = i\mathfrak{t}$. Writing this relation in the form $\mathfrak{t} = -i\tau$, we see that $-i\tau$ is a holomorphic function on \mathcal{T} that coincides with \mathfrak{t} on Γ and thus represents the analytic continuation of \mathfrak{t} from a function on Γ to a holomorphic function on \mathcal{T} .

At this point, we can explain what the Feynman $i\epsilon$ means for the special case of an annulus. Instead of integrating over real \mathfrak{t} , we integrate \mathfrak{t} over the contour that was sketched in fig. 6, or equivalently, we integrate $\tau = i\mathfrak{t}$ over the contour in the upper half-plane that corresponds to this.⁵⁰ We include the

⁵⁰ To show that what we need to integrate is holomorphic in \mathfrak{t} , observe that the partition function on the annulus is $\text{Tr } \exp(-\mathfrak{t}H)$, where H is the Hamiltonian, and this is manifestly holomorphic in \mathfrak{t} for $\text{Re } \mathfrak{t} > 0$. In general, an open-string worldsheet that is close to an open-string degeneration is built by gluing two surfaces Σ_ℓ and Σ_r via a long strip of length \mathfrak{t} , as in fig. 3(a). The path integral on such a worldsheet is the matrix element of $e^{-\mathfrak{t}H}$ between initial and final states that are determined

usual convergence factor $\exp(-\epsilon \operatorname{Im} \mathfrak{t})$. Of course, we do not need to integrate precisely over the stated contour. Any contour homologous to this one and with the same behavior at infinity will do.

The generalization for other topologies proceeds in the same way, treating each open-string degeneration as we have just described. One defines the scattering amplitudes by integrating over an integration cycle Γ that coincides with the naive one Γ except near a degeneration, where \mathfrak{t} is integrated over a contour like that of fig. 6. One can find more detail in [56], but further details are not really essential for the remainder of the present paper. However, one point is really worth spelling out.

6.3.2. An interesting analogy Naively open-string scattering amplitudes are computed by integrating over the moduli space Γ of conformal structures on the open-string worldsheet. However, as we have seen, the Feynman $i\epsilon$ means that in reality, we must integrate (with a convergence factor) over a more general cycle Γ in a complexification of Γ . The appropriate complexification of Γ is a cover \mathcal{T} of the complex moduli space \mathcal{M} .

In the supersymmetric case, given an open-string worldsheet Σ_0 , it seems that, for technical reasons involving supermoduli, there is no natural definition of a moduli space Γ of superconformal structures on Σ_0 . (See section 5 of [15] and section 7 of [16].) Because of this fact, the naive idea that open-superstring scattering amplitudes are computed by integrating over a moduli space Γ of open-superstring worldsheets is not really correct. Instead, one must integrate over a cycle Γ in a complex supermanifold \mathfrak{M} that parametrizes superconformal structures on a closed super Riemann surface Σ that corresponds to the oriented double cover of Σ_0 .

Thus, whether we consider the Feynman $i\epsilon$ or the details of supermoduli, the conclusion is similar. Open-string scattering amplitudes must be defined by integrating not over the naive moduli space Γ , but over a more general cycle Γ in a suitable complexification of Γ , either because the naive Γ does not exist (open superstrings) or because it exists but does not incorporate the Feynman $i\epsilon$ (open bosonic strings). For open superstrings, we have both problems and we must define Γ in a suitable cover of \mathfrak{M} .

In each case, there is no natural choice of Γ . Any homologous cycle with the same behavior at infinity is equally good.

by the path integrals on Σ_ℓ and Σ_r . Keeping $\operatorname{Re} \mathfrak{t}$ large (where this description is meaningful) and giving \mathfrak{t} an imaginary part, such a matrix element is holomorphic in \mathfrak{t} .

The two issues of the Feynman $i\epsilon$ and the Deligne-Mumford compactification are actually complementary in the following sense. At a generic degeneration not associated to mass renormalization or massless tadpoles (see section 7.1), the Feynman $i\epsilon$ is important and we integrate over a contour such as that of fig. 6. In particular, we do not integrate up to $\text{Re } \mathfrak{t} = \infty$, where the compactification occurs, so we do not see the details of the Deligne-Mumford compactification.⁵¹ (We can still use the Deligne-Mumford compactification to analyze the residue of the pole in the propagator, as discussed later.) On the other hand, at special degenerations (associated to mass renormalization or massless tadpoles) at which $p^2 + m^2$ is identically 0, the Feynman propagator $1/(p^2 + m^2 - i\epsilon) = i/\epsilon$ has a pole at $\epsilon = 0$ and the Feynman $i\epsilon$ is not useful. That case, as we will discuss starting in section 7.1, calls for a different approach that does require a knowledge of the compactification.

6.3.3. Closed strings All this persists if we consider the Feynman $i\epsilon$ for closed strings. Let Σ be the worldsheet of a closed bosonic string, and \mathcal{M} the moduli space of conformal structures on Σ . To exploit the fact that the functions and measures usually encountered in string perturbation theory are real-analytic, we can view \mathcal{M} as the diagonal in a product $\mathcal{M}_L \times \mathcal{M}_R$, where \mathcal{M}_R and \mathcal{M}_L parametrize respectively holomorphic and antiholomorphic structures on Σ . (In the bosonic case, \mathcal{M}_L and \mathcal{M}_R are isomorphic spaces with opposite complex structures; each is naturally isomorphic to \mathcal{M} .) A real-analytic function on \mathcal{M} can be analytically continued to a holomorphic function on $\mathcal{M}_L \times \mathcal{M}_R$, or on a covering space thereof.

A closed-string degeneration is described by vanishing of a complex parameter q ; the Euclidean proper time variable is for closed strings $\mathfrak{t} = -\log |q|$. (This coincides with the integration variable s that appears in eqn. (6.53) above.) As \mathfrak{t} is a real-analytic function on \mathcal{M} , it can be continued locally to a holomorphic function on $\mathcal{M}_L \times \mathcal{M}_R$. One wants to replace $\mathcal{M}_L \times \mathcal{M}_R$ by a cover that is large enough so that $\text{Im } \mathfrak{t}$ is single-valued at every degeneration, but small enough that one can find in it an integration cycle that is topologically the same as the diagonal $\mathcal{M} \subset \mathcal{M}_L \times \mathcal{M}_R$. For this, we observe that the universal cover of \mathcal{M} is a Teichmuller space \mathcal{T} , with $\mathcal{M} = \mathcal{T}/F$ where F is a discrete group (the mapping class group). The universal cover of $\mathcal{M}_L \times \mathcal{M}_R$ is therefore $\mathcal{T}_L \times \mathcal{T}_R$, the product of two Teichmuller spaces.

⁵¹This is the reason for something that was explained in a slightly different way in sections 4.4 and 4.5: tree-level scattering amplitudes can be computed correctly with a procedure that treats correctly the interior of moduli space, but does not treat the compactification correctly. At tree level, one does not encounter tadpoles or mass renormalization.

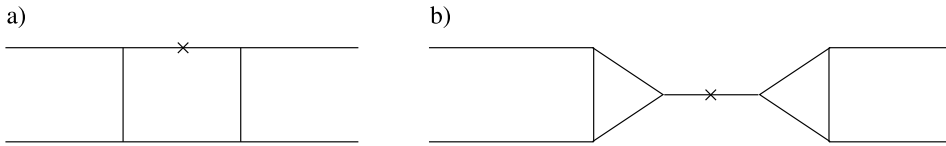


Figure 7: Cutting an internal line (marked \times) in a Feynman diagram will reduce the number of loops by 1, as in (a), or disconnect the diagram, as in (b).

A suitable cover of $\mathcal{M}_L \times \mathcal{T}_R$ is $(\mathcal{T}_L \times \mathcal{T}_R)/F$, with the diagonal action of F on the two factors. This contains \mathcal{M} , embedded as the diagonal, but is “large” enough so that $\text{Im } \mathfrak{t}$ is single-valued at each degeneration. Then the contour of fig. 6 makes sense in this situation: it is part of the definition of an integration cycle $\Gamma \subset (\mathcal{M}_L \times \mathcal{M}_R)/F$ that coincides with the diagonal $\mathcal{M} \subset (\mathcal{M}_L \times \mathcal{M}_R)/F$ except near a degeneration. To incorporate for bosonic strings the Feynman $i\epsilon$, we replace \mathcal{M} with Γ and include in the integral a convergence factor $\exp(-\epsilon \text{Im } \mathfrak{t})$.

Everything is the same for closed superstrings, except that in this case there is no natural moduli space of closed superstring worldsheets to begin with. Even without the Feynman $i\epsilon$, all that one can naturally define in any case is an integration cycle $\Gamma \subset \mathcal{M}_L \times \mathcal{M}_R$ (where \mathcal{M}_R and \mathcal{M}_L parametrize holomorphic and antiholomorphic structures on the worldsheet). This framework, which is forced on us by properties of supermoduli, is in any event what we need to incorporate the Feynman $i\epsilon$. For this, we have to replace $\mathcal{M}_L \times \mathcal{M}_R$ by its cover $(\mathcal{T}_L \times \mathcal{T}_R)/F$ and define Γ in this cover.

6.4. The on-shell and infrared behavior of string theory

6.4.1. Overview At this stage, it is almost clear that string theory amplitudes will match the on-shell and infrared singularities that one would expect in a field theory with the same particles and interactions. In field theory, the singularities come from the poles of propagators. In string theory, the singularities can only come from integration over the length of a long tube or strip, and we have seen that the singularities generated in this way precisely match the poles in a corresponding field theory propagator.

To pursue this further, we need a certain factorization property of the worldsheet path integral of string theory. In field theory, one can always imagine “cutting” an arbitrary line in a Feynman diagram to make a diagram with one less propagator and two more external lines. Cutting a line either causes

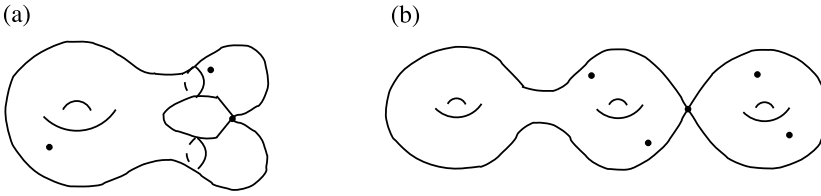


Figure 8: (a) A nonseparating degeneration reduces the genus of Σ by 1 while adding two punctures – one on each side. So this divisor in $\widehat{\mathcal{M}}_{g,n}$ is a copy of $\widehat{\mathcal{M}}_{g-1,n+2}$. (b) A separating degeneration splits Σ in two components, whose genera g_1 and g_2 sum to g . The punctures are divided between the two components and an extra puncture is added on each branch. So this divisor in $\widehat{\mathcal{M}}_{g,n}$ is a copy of $\widehat{\mathcal{M}}_{g_1,n_1+1} \times \widehat{\mathcal{M}}_{g_2,n_2+1}$, where $g_1 + g_2 = g$, $n_1 + n_2 = n$.

a Feynman diagram to become disconnected or else reduces the number of loops by 1 (fig. 7).

The cutting procedure is most useful when the line that is cut is on-shell or almost on-shell. Of course, one can simultaneously cut several lines, taking the corresponding lines to be on-shell. The main purpose of cutting is to understand the singularity that a Feynman diagram develops when one or more internal lines go on-shell. A fundamental fact about Feynman diagrams is the following “factorization” property: evaluating a diagram with the momentum flowing through a given line held fixed is equivalent to cutting that line, treating its ends as external lines, and evaluating what remains of the diagram.

We need an analogous fact in string theory. It can be understood from the Deligne-Mumford compactification of moduli space (for background to much of what follows, see section 6 of [16]). For illustration, until section 6.4.5 we consider bosonic closed strings only. We write $\widehat{\mathcal{M}}_{g,n}$ for the Deligne-Mumford compactification of the moduli space $\mathcal{M}_{g,n}$ of Riemann surfaces Σ of genus g and n punctures. $\widehat{\mathcal{M}}_{g,n}$ has a number of “divisors at infinity” corresponding to the possible ways that Σ can degenerate. A nonseparating degeneration reduces the genus of Σ by 1 while adding two punctures (fig. 8(a)). The corresponding divisor $\mathcal{D}_{\text{nonsep}} \subset \widehat{\mathcal{M}}_{g,n}$ is simply the moduli space of Riemann surfaces with the appropriate genus and number of punctures:

$$(6.64) \quad \mathcal{D}_{\text{nonsep}} \cong \widehat{\mathcal{M}}_{g-1,n+2}.$$

A separating degeneration divides Σ into two components while dividing the punctures in some way between the two sides (fig. 8(b)); in addition, the

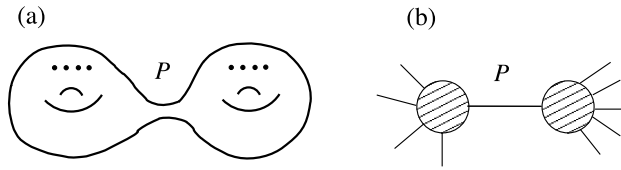


Figure 9: (a) A separating degeneration of a Riemann surface with momentum P flowing between the left and right. (b) The field theory analog, in which a total momentum P enters on the left and exits on the right.

singularity counts as an extra puncture on both sides. So the corresponding divisor $\mathcal{D}_{\text{sep}} \subset \widehat{\mathcal{M}}_{g,n}$ factorizes:

$$(6.65) \quad \mathcal{D}_{\text{sep}} \cong \widehat{\mathcal{M}}_{g_1, n_1+1} \times \widehat{\mathcal{M}}_{g_2, n_2+1},$$

where

$$(6.66) \quad g_1 + g_2 = g, \quad n_1 + n_2 = n.$$

Moreover, as we will now explain, the measure on moduli space that comes from the worldsheet path integral factorizes at these degenerations in a way that is very similar to the behavior of a Feynman diagram when a line is cut.

6.4.2. The separating case In either field theory or string theory, separating degenerations are simpler to analyze. The total momentum P flowing through the separating line from the left to the right of the graph or worldsheet of fig. 9 is equal to the sum of all momenta flowing in from the left or out to the right. The amplitude has a pole $1/(P^2 + m^2)$ whenever a particle σ that propagates between the two parts of the diagram is on-shell at momentum P . In field theory, the residue of the pole is simply obtained by evaluating the rest of the diagram, with the separating line cut and σ and its conjugate (its antiparticle) attached at its two ends. We need a similar result in string theory.

The key assertion is that the worldsheet path integral of string theory behaves in the way that is suggested by the description (6.65) of the compactification divisor. Let us consider a Riemann surface Σ that is degenerating to a pair of surfaces Σ_ℓ, Σ_r . We pick local coordinates x and y on Σ_ℓ and Σ_r . Picking points $x = a$ in Σ_ℓ and $y = b$ in Σ_r , we glue the two surfaces together by

$$(6.67) \quad (x - a)(y - b) = q.$$

Thus, at $q = 0$, the point $x = a$ in Σ_ℓ is glued to the point $y = b$ in Σ_r . For $q \neq 0$, the two branches join together smoothly.

Near the degeneration, the moduli of Σ consist of the following: the gluing parameter q ; the positions a and b of the extra punctures in Σ_ℓ and Σ_r at which the gluing is made; and the other moduli of Σ_ℓ and Σ_r . The worldsheet path integral can be evaluated as follows. The integral over q gives the closed bosonic string propagator

$$(6.68) \quad \tilde{b}_0 b_0 \int_{|q| \leq 1} \frac{d^2 q}{|q|^2} q^{L_0} \bar{q}^{\tilde{L}_0}.$$

An important detail here is that the precise upper limit on the q integral is not important. The pole comes entirely from an arbitrarily small neighborhood of $q = 0$. Indeed, the pole arose in section 6.2.1 as the contribution of a string state that flows through the narrow neck between Σ_ℓ and Σ_r . Let σ be one of the string states that contributes to the pole; let \mathcal{V}^σ be the corresponding vertex operator, and let \mathcal{V}_σ be the vertex operator for the antiparticle of σ . By the state-operator correspondence of conformal field theory, propagation of σ through the neck can be represented in an effective path integral on Σ_r by the insertion of \mathcal{V}^σ at the point b ; similarly it is represented in an effective path integral on Σ_ℓ by insertion of \mathcal{V}_σ . (For more on this, see [17] and also section 6.5 below.) The residue of the pole due to the string state σ is computed by integration over $\mathcal{D}_{\text{sep}} = \widehat{\mathcal{M}}_{g_1, n_1+1} \times \widehat{\mathcal{M}}_{g_2, n_2+1}$ (that is, over a, b , and all the other moduli of Σ_ℓ and Σ_r) with the insertions of \mathcal{V}_σ and \mathcal{V}^σ just described, along with any other vertex operators that may be present. This gives the same sort of description of the residue of a pole that we had in field theory.

We have implicitly assumed that the momentum P flowing between Σ_ℓ and Σ_r is generically off-shell. This is true as long as there are two or more external states attached to both Σ_ℓ and Σ_r (for example, in fig. 9(a), there are four external states on each side). In this case, the $1/(P^2 + m^2)$ singularity of the propagator gives a pole as a function of the external momenta; these poles are important in the physical interpretation of scattering amplitudes. If the number of external particles on Σ_ℓ and/or Σ_r is either 1 or 0, then P is automatically on-shell, independent of the external momenta, and the propagator has a $1/0$ singularity. We will return to these cases in section 7.

6.4.3. The nonseparating case In field theory, when a nonseparating line goes on-shell, one has to integrate over the momentum flowing through the relevant line. In D -dimensional Minkowski spacetime with $d > 2$ (possibly

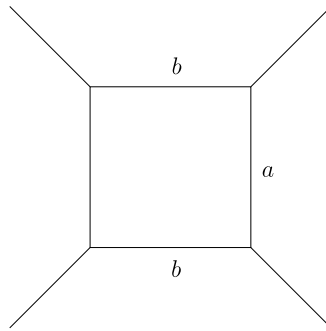


Figure 10: In this contribution to an S -matrix element, when the line labeled a goes to zero momentum, the lines labeled b are forced to be on-shell. For spacetime dimension $d \leq 4$, this leads to infrared divergences in gravity and in gauge theory with massless gauge fields. These divergences follow from very general properties of phase space and momentum conservation, along with the low energy limits of the gauge and gravitational couplings. So they are common to field theory and string theory.

multiplied by a compact manifold of some sort), there is no infrared singularity associated to the momentum integral for a single generic propagator, even for a massless field, since the integral

$$(6.69) \quad \int \frac{d^d p}{(2\pi)^d} \frac{1}{p^2}$$

is convergent at $p = 0$. The divergence at $d \leq 2$ is common to field theory and string theory, since it only depends on the pole of the propagator and on d -dimensional phase space. The usual on-shell infrared divergences for theories with massless fields in $d \leq 4$ dimensions arise because in some cases (fig. 10) taking the momentum of one line to zero forces adjacent lines to be on-shell, leading to an infrared behavior that is less convergent than that of (6.69). Again, this depends only on momentum conservation, the poles of propagators, and d -dimensional phase space, as well as the low energy limits of couplings, so it does not distinguish field theory and string theory in any way.⁵²

⁵²In four-dimensional field theories with massless particles, infrared divergences in the perturbative S -matrix are often dealt with via dimensional regularization. One works in $4 - \varepsilon$ dimensions and one takes the limit $\varepsilon \rightarrow 0$ only after imposing a lower bound on the energy of an observable soft particle. One presumably could do

In field theory, the most interesting singularities that come from cutting of nonseparating lines are the singularities associated to unitarity. Nothing happens in a generic Feynman diagram when a single generic nonseparating line goes on-shell, but if enough nonseparating lines go on-shell that their removal would “cut” the diagram in two disjoint pieces, then one gets singularities associated to unitarity. These singularities come ultimately from the poles of propagators and the way the Feynman rules factorize when a line is cut. The contributions of these singularities depend crucially on the Feynman $i\epsilon$.

In string theory, one has the same poles and – via (6.64) and reasoning that we have just sketched above – the same sort of factorization of the worldsheet path integral. Moreover, the Feynman $i\epsilon$ can be incorporated along lines explained in section 6.3. With this in place, it is reasonable to expect that perturbative string theory is unitary. A proof of unitarity that proceeds roughly along these lines, by adapting to string theory what is known about field theory with the Feynman propagator, has been given recently [22].

An older approach to unitarity in string theory was based on light cone gauge, in which unitarity is manifest. Light cone string diagrams give a triangulation of the moduli space of ordinary Riemann surfaces [58], and this has been used in bosonic string theory to establish the equivalence of light cone string perturbation theory to the covariant description [59]. Analogous arguments have also been developed in superstring theory [60, 61]. Finally, we mention that a rare example in which analytic properties of a string theory loop amplitude have been analyzed explicitly with manifest unitarity can be found in [62].

6.4.4. The leading singularity Much easier than describing unitarity, or giving a full description of all singularities of scattering amplitudes, is to describe the analog in string theory of what in field theory is called the “leading singularity” of an amplitude. (For a modern explanation and application of this notion, see [63].)

Any string worldsheet Σ with $g > 0$ or with $g = 0$ and more than three punctures has a strictly positive number of bosonic moduli and can undergo a separating or nonseparating degeneration. In the nonseparating case, this reduces the genus of Σ ; in the separating case, Σ is decomposed into two components each of which has a smaller genus and/or fewer punctures than Σ . We can continue this process, degenerating Σ again or degenerating its

something similar in string theory, by compactifying from \mathbb{R}^4 to $\mathbb{R}^{4-\varepsilon} \times T^\varepsilon$, where T^ε is a torus of dimension ε .

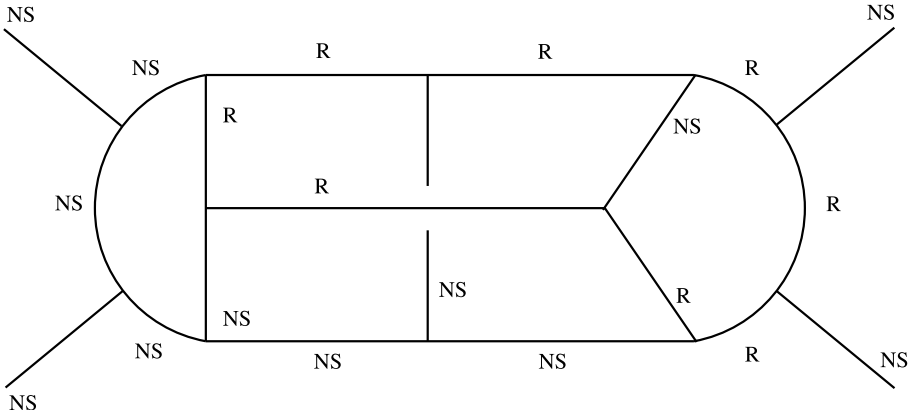


Figure 11: A trivalent graph in which the lines represent NS or R states and the vertices represent genus 0 three-point functions of type NS^3 or $\text{NS} \cdot \text{R}^2$. Leading singularities of string theory amplitudes are associated to such trivalent graphs. The example considered here is a five-loop amplitude with four external NS lines. Internal lines and vertices are of both types.

components, until finally Σ is built out of a collection of three-punctured spheres, glued together pairwise along their punctures. This gives the string theory analog of a Feynman diagram in which all possible internal lines have been placed on-shell. The simultaneous poles in all possible channels give the leading singularity.

From a completely degenerated Riemann surface, we can build a trivalent graph in which a three-punctured sphere is represented by a trivalent vertex, while the gluing of two punctures is represented by a line between two vertices. Leading singularities come from these trivalent graphs. Though we have not yet described degenerations of super Riemann surfaces, the generalization of these trivalent graphs to superstrings is so straightforward that we will describe it without further ado. In – for example – the heterotic string, there are two types of string states – NS and R – and there are two types of three-punctured spheres, which we will call NS^3 (three punctures of NS type) and $\text{NS} \cdot \text{R}^2$ (one NS puncture and two R punctures). Leading singularities of heterotic string amplitudes are associated to trivalent graphs labeled as in fig. 11, with two types of internal line and two types of vertex. The vertices represent the NS^3 genus 0 three-point function described in eqn. (4.27), and the $\text{NS} \cdot \text{R}^2$ genus 0 three-point function described in eqn. (5.60). These trivalent graphs have an obvious analogy with Feynman diagrams of field theory.

This is part of the match between the on-shell singularities of field theory and string theory.

6.4.5. Factorization of superstring amplitudes In our explanation of how to compare the singularities of string theory and field theory amplitudes, there were several key ingredients, which we may roughly summarize as follows:

1. A singularity develops when a certain modulus q vanishes and Σ degenerates. Integration over q produces poles that are analogous to the poles of a Feynman propagator.
2. The locus $q = 0$ is a divisor \mathcal{D} in the moduli space; it can be factored as in (6.64) or (6.65).
3. The integral over \mathcal{D} that gives the residue of a pole due to a given string state σ is itself a string theory scattering amplitude, now with extra insertions of the vertex operator \mathcal{V}_σ and its conjugate on the two sides.

All of these ingredients have analogs in superstring theory, with just a few added wrinkles. The bosonic gluing formula $xy = q$ has two different superanalogs, one for the NS sector and one for the R sector (these were originally constructed by Deligne [40, 41]; for more detail on the following, see section 6 of [16]). In the NS sector, we glue together two copies of $\mathbb{C}^{1|1}$, one with superconformal coordinates $x|\theta$ and one with superconformal coordinates $y|\psi$, via

$$\begin{aligned}
 xy &= -\varepsilon^2 \\
 y\theta &= \varepsilon\psi \\
 x\psi &= -\varepsilon\theta \\
 \theta\psi &= 0.
 \end{aligned}
 \tag{6.70}$$

It is convenient to define

$$q_{\text{NS}} = -\varepsilon^2.
 \tag{6.71}$$

q_{NS} is the closest analog of the bosonic gluing parameter q . For given q_{NS} , ε is determined only up to sign. The sum over the possible signs of ε gives the GSO projection in the NS sector, as explained in section 6.2.3 of [16].

To describe a Ramond degeneration, we again glue two copies of $\mathbb{C}^{1|1}$ with local coordinates $x|\theta$ and $y|\psi$. But now we endow each copy with a

superconformal structure defined by the odd vector fields

$$(6.72) \quad D_\theta^* = \frac{\partial}{\partial \theta} + \theta x \frac{\partial}{\partial x}, \quad D_\psi^* = \frac{\partial}{\partial \psi} + \psi y \frac{\partial}{\partial y}.$$

This is chosen to describe Ramond punctures at $x = 0$ and $y = 0$, respectively. The gluing formulas are now simply

$$(6.73) \quad \begin{aligned} xy &= q_R \\ \theta &= \pm \sqrt{-1} \psi. \end{aligned}$$

The GSO projection in the Ramond sector comes from the sum over the sign that is explicitly written in eqn. (6.73). The parameter q_R plays the role of the bosonic gluing parameter q and the NS gluing parameter q_{NS} .

Now let us move on to the second item in the above list. In the NS sector, the factorizations (6.64) and (6.65) have immediate analogs. For example, let $\widehat{\mathfrak{M}}_{g, n_{NS}, n_R}$ be the Deligne-Mumford compactification of $\mathfrak{M}_{g, n_{NS}, n_R}$, the moduli space of super Riemann surfaces of genus g with n_{NS} NS punctures and n_R Ramond punctures. The nonseparating divisor $\mathcal{D}_{\text{nonsep}}$ has the same sort of description as for bosonic Riemann surfaces. At a nonseparating degeneration, the genus is reduced by 1 and n_{NS} increases by 2; an extra NS puncture appears on each branch:

$$(6.74) \quad \mathcal{D}_{\text{nonsep}} \cong \widehat{\mathfrak{M}}_{g, n_{NS}+2, n_R}.$$

This is the obvious analog of eqn. (6.64). Similarly the obvious analog of eqn (6.65) holds for separating degenerations. Any separating divisor has the form

$$(6.75) \quad \mathcal{D}_{\text{sep}} \cong \widehat{\mathfrak{M}}_{g_1, n_{NS,1}+1, n_{R,1}} \times \widehat{\mathfrak{M}}_{g_2, n_{NS,2}+1, n_{R,2}},$$

with $g_1 + g_2 = g$, $n_{NS,1} + n_{NS,2} = n_{NS}$, and $n_{R,1} + n_{R,2} = n_R$.

Given this, the analog of the third item is clear. We can describe the residue of the pole due to an on-shell NS state in the same way that we did for bosonic strings in sections 6.4.2 and 6.4.3. Integration over q_{NS} gives poles due to on-shell string states; if a string state σ contributes such a pole, then according to the state-operator correspondence of superconformal field theory, vertex operators \mathcal{V}_σ and \mathcal{V}^σ appear on the two branches. Integration over the remaining moduli gives a description of the residue of the pole due to σ in terms of a scattering amplitude with the extra insertions of \mathcal{V}_σ and \mathcal{V}^σ .

By contrast, the most obvious analog of eqns. (6.74) and (6.75) does *not* hold for Ramond degenerations. The reason is that Ramond punctures are really divisors. For brevity, we will describe the situation for separating degenerations; the same idea holds for nonseparating ones. Suppose that we are given two super Riemann surfaces Σ_ℓ and Σ_r each with a distinguished Ramond divisor. Suppose further that near each of these divisors we are given local coordinates $x|\theta$ or $y|\psi$, with the superconformal structure being as in (6.72). The distinguished divisors are given respectively by $x = 0$ and by $y = 0$, and are parametrized by θ and by ψ . We want to glue together these two divisors to make a super Riemann surface Σ . We can do this by gluing $\theta = \pm\sqrt{-1}\psi$ as in (6.73). But we can also introduce an odd parameter η and make the gluing $\theta = \pm\sqrt{-1}(\psi + \eta)$. Here η is an odd modulus that we should for our present purposes associate to the gluing and not to Σ_ℓ or Σ_r . We call η the fermionic gluing parameter.

One can generalize (6.74) or (6.75) for a Ramond degeneration if one takes proper account of the fermionic gluing parameter. For example, instead of a separating Ramond divisor \mathcal{D}_{sep} being a product $\widehat{\mathfrak{M}}_{g_1, n_{\text{NS},1}, n_{\text{R},1}+1} \times \widehat{\mathfrak{M}}_{g_2, n_{\text{NS},2}, n_{\text{R},2}+1}$, it is a fiber bundle over such a product. We denote the fibration as

$$(6.76) \quad \mathbf{\Pi} : \mathcal{D}_{\text{sep}} \rightarrow \widehat{\mathfrak{M}}_{g_1, n_{\text{NS},1}, n_{\text{R},1}+1} \times \widehat{\mathfrak{M}}_{g_2, n_{\text{NS},2}, n_{\text{R},2}+1}.$$

The fibers have dimension $0|1$ and are parametrized by the fermionic gluing parameter. There is a precisely analogous fibration for nonseparating degenerations.

Now let us consider the pole due to an on-shell Ramond string state. What is the residue of such a pole? In the separating case, imitating what we have done for bosonic strings and for the NS sector of superstrings, we would like to express this residue in terms of an integral over $\widehat{\mathfrak{M}}_{g_1, n_{\text{NS},1}, n_{\text{R},1}+1} \times \widehat{\mathfrak{M}}_{g_2, n_{\text{NS},2}, n_{\text{R},2}+1}$. If we simply integrate over q_{R} , we will extract in the usual way a pole $1/L_0$, and we learn that its residue can be computed as an integral over \mathcal{D}_{sep} . To further reduce to an integral over $\widehat{\mathfrak{M}}_{g_1, n_{\text{NS},1}, n_{\text{R},1}+1} \times \widehat{\mathfrak{M}}_{g_2, n_{\text{NS},2}, n_{\text{R},2}+1}$, we have to integrate over the fibers of the fibration $\mathbf{\Pi}$. In other words, we have to integrate over the fermionic gluing parameter η .

This parameter was introduced in another way in section 6.1.4. As we explained there, integration over η gives a factor of G_0 . So the contribution of an on-shell Ramond state to a scattering amplitude is $G_0/L_0 = 1/G_0$ times an integral over $\widehat{\mathfrak{M}}_{g_1, n_{\text{NS},1}, n_{\text{R},1}+1} \times \widehat{\mathfrak{M}}_{g_2, n_{\text{NS},2}, n_{\text{R},2}+1}$. Just as in the case of an NS degeneration, the operator-state correspondence of conformal field theory tells us how to compute the residue of the pole associated to an on-shell Ramond

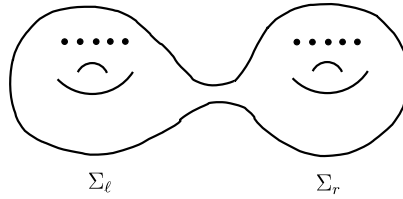


Figure 12: A super Riemann surface Σ decomposes to a union of two components Σ_ℓ and Σ_r , meeting at a point (or at a Ramond divisor).

sector string state σ : we integrate over $\widehat{\mathfrak{M}}_{g_1, n_{NS,1}, n_{R,1}+1} \times \widehat{\mathfrak{M}}_{g_2, n_{NS,2}, n_{R,2}+1}$ with insertions of the appropriate vertex operators \mathcal{V}_σ and \mathcal{V}^σ on the two sides.

To compare the description of the gluing parameter that we have given here with that of section 6.1.4, one may observe the following. In the local model (6.73), including the fermionic gluing parameter is equivalent to acting on $y|\psi$ with a superconformal transformation $\exp(\eta G_0)$, where G_0 is the superconformal vector field

$$(6.77) \quad G_0 = \frac{\partial}{\partial \psi} - \psi y \frac{\partial}{\partial y}$$

that acts nontrivially on the Ramond divisor at $y = 0$. (In particular, at $y = 0$, $\exp(\eta G_0)$ transforms ψ to $\psi + \eta$.) Because of this geometrical fact, a shift in the fermionic gluing parameter acts on the string state flowing between the two branches of Σ by $\exp(\eta G_0)$, which is the result that we obtained in another but related way in section 6.1.4.

Remark. A further remark will be important when we study anomalies in section 9. There is an important special case in which the fermionic gluing parameter does *not* represent a modulus. If Σ_r has a superconformal symmetry that acts on the divisor at $y = 0$ by $\psi \rightarrow \psi + \eta$ (or if Σ_ℓ has a superconformal symmetry that acts as such a shift at $x = 0$) then the fermionic gluing parameter is not a modulus of the super Riemann surface Σ made by gluing of Σ_ℓ and Σ_r . In practice, this happens only for the Ramond-Ramond sector of open and/or unoriented superstring theory, and only for very particular cases of Σ_r or Σ_ℓ .

6.4.6. Picture numbers and picture-changing operators As we have just explained, the physical interpretation of superstring theory depends on the fact that the residue of a pole at $L_0 = 0$ can be interpreted as a sum of

contributions of on-shell states. Thus (fig. 12), if a super Riemann surface Σ splits locally into a product of two branches Σ_ℓ and Σ_r meeting at a point (or at a Ramond divisor), then the contribution of an on-shell string state σ to the pole can be evaluated by inserting vertex operators \mathcal{V}_σ and \mathcal{V}^σ on Σ_ℓ and Σ_r , respectively.

This statement reflects the factorization of the divisor at infinity, which for a separating NS degeneration takes the form (6.75):

$$(6.78) \quad \mathcal{D}_{\text{sep}} \cong \widehat{\mathfrak{M}}_{g_1, n_{\text{NS},1}+1, n_{\text{R},1}} \times \widehat{\mathfrak{M}}_{g_2, n_{\text{NS},2}+1, n_{\text{R},2}},$$

where Σ_ℓ has genus g_1 with $n_{\text{NS},1} + 1$ NS punctures and $n_{\text{R},1}$ Ramond punctures, while the corresponding values for Σ_2 are g_2 , $n_{\text{NS},2} + 1$, and $n_{\text{R},2}$. At a separating Ramond degeneration, the divisor at infinity has an analogous but slightly more complicated structure described in eqn. (6.76).

There is one very simple but important observation to add to what we have already said. Interpreting the residue of the pole by using the factorization of the divisor at infinity requires that the vertex operators \mathcal{V}_σ and \mathcal{V}^σ should have the canonical picture numbers, namely -1 in the NS sector and $-1/2$ in the R sector. Indeed, as we know from sections 4.1 and 5.1, in a superconformal formalism, only vertex operators of canonical picture can be inserted at a puncture of NS or R type, assuming that those punctures are defined in the conventional way. One can modify the definition of a puncture by endowing it with more structure in a way that increases the odd dimension of the moduli space (see section 4.3 of [16]) and this makes it possible to compute in a superconformal formalism using vertex operators of picture number more negative than -1 or $-1/2$. If we wish, though there is no apparent benefit in doing so, we can do this for the punctures at which external vertex operators in fig. 12 are inserted. But we do not have any such freedom for the “internal” punctures that appear when Σ degenerates; they are the standard punctures described in sections 4.1 and 5.1, and in a superconformally-invariant formalism, the vertex operators inserted at those punctures will have canonical picture number.

Another type of comment about picture number may be helpful. In this paper, we never make any fundamental statements in terms of positions of picture-changing operators (PCO’s); we view the PCO’s as a possible method of integration over odd moduli, but not a basic definition. Still, it may be helpful to spell out in terms of PCO’s what it means to treat correctly the compactification of the moduli space. To the extent that the decomposition (6.75) of the divisor \mathcal{D}_{sep} is important, one would like the procedure for integrating over odd moduli using the PCO’s to respect this factorization. This

means that when Σ decomposes to two intersecting components Σ_ℓ and Σ_r , the location of the PCO's should have a limit as $q_{\text{NS}} \rightarrow 0$, and Σ_ℓ and Σ_r should each have the right number of PCO's to describe its odd moduli. Moreover, the PCO's should be placed to avoid spurious singularities on Σ_ℓ and on Σ_r . (To compute correctly with PCO's, they must be used piecewise, avoiding spurious singularities, as described in section 3.5. One follows this same procedure along both Σ_ℓ and Σ_r .) With notation as above, the odd dimension of the moduli space of Σ_ℓ is $(2g_1 - 2) + (\mathfrak{n}_{\text{NS},1} + 1) + \mathfrak{n}_{\text{R},1}/2$, and that is the number of PCO's that should be placed on Σ_ℓ . The corresponding number of PCO's on Σ_r is then $(2g_2 - 2) + (\mathfrak{n}_{\text{NS},2} + 1) + \mathfrak{n}_{\text{R},2}/2$. And if Σ_ℓ and Σ_r undergo further separating degenerations, the conditions just stated should be satisfied again on each of the resulting components.

All this has a close analog for Ramond degenerations, with one correction: for a Ramond degeneration, there is one odd modulus associated to the gluing rather than to Σ_ℓ or Σ_r . So as Σ degenerates via formation of a long neck, one of the PCO's should be placed in the neck.

It is a slightly tricky question to what extent one will actually get wrong answers in the picture-changing formalism if one distributes the PCO's incorrectly when Σ degenerates. In genus 0, as long as the momentum flowing between Σ_ℓ and Σ_r is generically off-shell, one can distribute the PCO's between Σ_ℓ and Σ_r in an arbitrary fashion, treating the compactification of the moduli space incorrectly, and still get the correct tree-level S -matrix. This was explained in section 4.4. It is not entirely clear if this remains true in higher genus; the spurious singularities may cause trouble. (The recipe explained in section 3.5 to compute piecewise and avoid spurious singularities certainly cannot be extended over the compactification if one treats the compactification incorrectly.) At any rate, in higher genus, one will run into degenerations (associated to massless tadpoles and mass renormalization) for which the momentum flowing between Σ_ℓ and Σ_r is generically on-shell. In our treatment of such questions, starting in section 7, it will be clear that one should expect trouble if one does not treat the compactification of moduli space correctly.

6.5. The states that contribute to the pole

Now we want to explain that only physical states contribute to the singularities described in section 6.4. (For a slightly different view with some further details, see [17].)

6.5.1. Bosonic open strings To illustrate the idea, we consider bosonic open strings. The propagator is b_0/L_0 , and the residue of the pole at $L_0 = 0$ is

$$(6.79) \quad R = b_0 \Pi_0,$$

where Π_0 is the projector onto states with $L_0 = 0$. To define Π_0 , one expands any state $\mathcal{V} \in \mathcal{H}$ in eigenstates of L_0 and defines $\Pi_0 \mathcal{V}$ as the component of \mathcal{V} with $L_0 = 0$. This can be done in a sector of fixed spacetime momentum; in such a sector, the spectrum of L_0 is discrete. The residue R is BRST-invariant:

$$(6.80) \quad \{Q_B, R\} = \{Q_B, b_0\} \Pi_0 = L_0 \Pi_0 = 0.$$

Of course, R has ghost number $N_{\text{gh}} = -1$.

Let \mathcal{H} be the space of all string states. We can view R as a linear map $R : \mathcal{H} \rightarrow \mathcal{H}$, or equivalently as an element of $\mathcal{H} \otimes \mathcal{H}^*$, where \mathcal{H}^* is the dual space to \mathcal{H} . Since R has ghost number -1 , it maps states of ghost number $n + 1$ to states of ghost number n , so it is, more specifically, an element of

$$(6.81) \quad \bigoplus_{n \in \mathbb{Z}} \mathcal{H}_n \otimes \mathcal{H}_{n+1}^*,$$

where \mathcal{H}_n is the subspace of \mathcal{H} with $N_{\text{gh}} = n$, and \mathcal{H}_n^* is the dual of \mathcal{H}_n .

So far, we have viewed the propagator and its residue as maps from one string state to another. But for what follows, it will be more convenient to view R more symmetrically as an element of $\mathcal{H} \otimes \mathcal{H}$, representing a pair of string states that are to be inserted at the two ends of the long strip. For this, we simply use the duality $\mathcal{H}_n^* \cong \mathcal{H}_{3-n}$ of eqn. (5.65), via which we can identify R with an element

$$(6.82) \quad R' \in \bigoplus_{n \in \mathbb{Z}} \mathcal{H}_n \otimes \mathcal{H}_{2-n},$$

or more simply as an element

$$(6.83) \quad R' \in (\mathcal{H} \otimes \mathcal{H})_2,$$

where $(\mathcal{H} \otimes \mathcal{H})_2$ is the subspace of $\mathcal{H} \otimes \mathcal{H}$ consisting of pairs of states with total ghost number 2.

Since R and the duality map are both BRST-invariant, it follows that $R' \in (\mathcal{H} \otimes \mathcal{H})_2$ is BRST-invariant. Hence, it helps to know the cohomology of the BRST operator Q_B acting on $\mathcal{H} \otimes \mathcal{H}$. This is simply the tensor product of

two copies of the cohomology of Q_B acting on \mathcal{H} . The non-zero cohomology groups of Q_B acting on \mathcal{H} are as follows⁵³:

1. In $N_{\text{gh}} = 0$, there is the identity operator 1, with zero spacetime momentum.
2. In $N_{\text{gh}} = 1$, the cohomology can be identified with the space of states $\mathcal{V} = cU$, where U is a dimension 1 conformal primary of the matter system.
3. In $N_{\text{gh}} = 2$, the cohomology can be identified with the space of states $\mathcal{V} = c\partial cU$, with U as before.
4. In $N_{\text{gh}} = 3$, there is the operator $c\partial c\partial^2 c$, with zero spacetime momentum.

It follows immediately that the $N_{\text{gh}} = 2$ cohomology of Q_B acting on $\mathcal{H} \otimes \mathcal{H}$ is, at any non-zero momentum, the tensor product of two copies of the $N_{\text{gh}} = 1$ cohomology of Q_B acting on \mathcal{H} . It is convenient now to let U_i be a basis of matter primaries of $L_0 = 1$, and U^i a dual basis (so that the genus zero two-point functions in the matter theory are $\langle U_i(0)U^j(1) \rangle = \delta_i^j$). Then we have

$$(6.84) \quad R' = \sum_i cU_i \otimes cU^i + Q_B \mathcal{X},$$

for some $\mathcal{X} \in (\mathcal{H} \otimes \mathcal{H})_1$, of the general form

$$(6.85) \quad \mathcal{X} = \sum_j \mathcal{S}_j \otimes \mathcal{T}_j, \quad \mathcal{S}_j, \mathcal{T}_j \in \mathcal{H}.$$

The contribution $\sum_i cU_i \otimes cU^i$ in (6.84) is important, so we will explain it in detail. Let us go back to the original definition of R as the map $b_0\Pi_0$ from \mathcal{H}_{n+1} to \mathcal{H}_n . As such, R maps $c\partial cU_i$ to cU_i ; indeed, $c\partial cU_i$ is invariant under Π_0 , and b_0 removes the ∂c factor. On the other hand, the dual of $c\partial cU_i$ is cU^i . Combining these statements, the cohomologically nontrivial part of R is $\sum_i cU_i \otimes cU^i$, as claimed in (6.84).

Eqn. (6.84) implies that the residue of a scattering amplitude at $L_0 = 0$, with all other vertex operators being BRST-invariant, can be computed from

⁵³Here and also in the corresponding discussion below of open superstrings, we omit possible Chan-Paton factors. If present, these must be included in an obvious way, taking the tensor product of what is described momentarily with an algebra of matrices acting on the Chan-Paton factors.

a sum over physical states where we insert cU_i on one side of the degeneration, cU^i at the other side, and sum over i . The residue receives no contribution from the BRST-trivial term $Q_B\mathcal{X}$. That $Q_B\mathcal{X}$ decouples is possibly most obvious for a separating degeneration, at which the worldsheet path integral factorizes as a product of path integrals on the two components of the worldsheet. In this case, the path integral with an insertion of $Q_B(\mathcal{S}_j \otimes \mathcal{T}_j) = Q_B\mathcal{S}_j \otimes \mathcal{T}_j + (-1)^{|\mathcal{S}_j|} \mathcal{S}_j \otimes Q_B\mathcal{T}_j$ vanishes, since a BRST-trivial state $Q_B\mathcal{S}_j$ or $Q_B\mathcal{T}_j$ is inserted on one branch or the other. To show the decoupling of $Q_B\mathcal{X}$ at a nonseparating degeneration, we have to use in a slightly more general way than we have done before the fundamental identity (2.36) for the decoupling of BRST-trivial states:

$$(6.86) \quad dF_\Omega + F_{Q_B\Omega} = 0.$$

Given any BRST-invariant vertex operators $\mathcal{V}_1, \dots, \mathcal{V}_n$, we set $\Omega = \mathcal{V}_1\mathcal{V}_2 \cdots \mathcal{V}_n\mathcal{X}$ (where the \mathcal{V}_i and the two factors of \mathcal{X} are inserted at $n + 2$ distinct punctures) and learn that

$$(6.87) \quad dF_{\mathcal{V}_1 \dots \mathcal{V}_n \mathcal{X}} + F_{\mathcal{V}_1 \dots \mathcal{V}_n Q_B \mathcal{X}} = 0.$$

The contribution of $Q_B\mathcal{X}$ to a genus g non-separating residue is therefore

$$(6.88) \quad \int_{\mathcal{M}_{g-1, n+2}} F_{\mathcal{V}_1 \dots \mathcal{V}_n Q_B \mathcal{X}} = - \int_{\mathcal{M}_{g-1, n+2}} dF_{\mathcal{V}_1 \dots \mathcal{V}_n \mathcal{X}}.$$

Thus this contribution vanishes, modulo possible anomalous contributions at infinity in moduli space. The contributions at infinity in arguments such as this will be analyzed in section 7 and do not affect our present discussion.

For future reference, we make a small aside. In studying gauge-invariance and BRST anomalies, it will be useful to replace the residue $R = b_0\Pi_0$ with $\tilde{R} = \Pi_0$. \tilde{R} is BRST-invariant but of course not annihilated by b_0 . Duality maps \tilde{R} to a Q_B -invariant element $\tilde{R}' \in (\mathcal{H} \otimes \mathcal{H})_3$. At non-zero spacetime momentum, by arguments similar to those above,

$$(6.89) \quad \tilde{R}' = \sum_i \left(c\partial c U_i \otimes c U^i + c U_i \otimes c\partial c U^i \right) + Q_B\mathcal{Y},$$

with $\mathcal{Y} \in (\mathcal{H} \otimes \mathcal{H})_2$. Note that \tilde{R}' is invariant under exchanging the two factors of $\mathcal{H} \otimes \mathcal{H}$. At zero momentum, there is an additional term

$$(6.90) \quad \left(c\partial c\partial^2 c \otimes 1 + 1 \otimes c\partial c\partial^2 c \right).$$

Returning to our main theme, we now have to discuss the role of conformal invariance in the above analysis. A careful reader may have noted a small sleight of hand in our explanation of the decoupling of $Q_B\mathcal{X}$. We have based the present paper on a conformally-invariant formalism in which the only vertex operators considered are conformal or superconformal vertex operators (primary fields that do not depend on derivatives of the ghost fields). But as we briefly described in section 2.4.4, it is also possible to develop a more general formalism in which conformal invariance is not assumed and vertex operators are more general Q_B -invariant local operators (annihilated by $b_0 - \tilde{b}_0$ in the case of closed strings). In such a formalism, the argument for decoupling of $Q_B\mathcal{X}$ proceeds exactly as stated above. However, in our conformally-invariant formalism, we cannot define a correlation function with insertion of $Q_B\mathcal{X}$ unless $Q_B\mathcal{X}$ is a conformal vertex operator. More precisely, $Q_B\mathcal{X} \in \mathcal{H} \otimes \mathcal{H}$ must be annihilated by L_n and b_n , $n \geq 0$, acting on either of the two factors of \mathcal{H} in $\mathcal{H} \otimes \mathcal{H}$. On the other hand, in a conformally-invariant formalism, the residue of a pole at $L_0 = 0$ must be conformally invariant, so the BRST-trivial contribution to the residue must come from an operator $Q_B\mathcal{X} \in \mathcal{H} \otimes \mathcal{H}$ that obeys these conditions. So at least what we have to show to vanish is well-defined. Given that $Q_B\mathcal{X}$ is annihilated by L_n and b_n , $n \geq 0$, in each factor, the argument of appendix B, applied to $\mathcal{H} \otimes \mathcal{H}$ rather than \mathcal{H} , shows that we can assume that \mathcal{X} obeys the same conditions, and then the argument for decoupling of $Q_B\mathcal{X}$ proceeds as above.

6.5.2. Open superstrings In open superstring theory, we let $\mathcal{H}_{n,m}$ be the space of string states of ghost number n and picture number m .

In the NS sector, there is not much new to say. As explained in section 6.4.6, we set the picture number to the canonical value $m = -1$. The residue of a scattering amplitude at $L_0 = 0$ is still $R = b_0\Pi_0$, which we can understand as an element of $\oplus_n \mathcal{H}_{n,-1} \otimes \mathcal{H}_{n+1,-1}^*$. Via the duality $\mathcal{H}_{n,-1}^* \cong \mathcal{H}_{1-n,-1}$ (eqn. (5.66)), we can identify R with a Q_B -invariant element $R' \in (\mathcal{H} \otimes \mathcal{H})_{0,-1 \otimes -1}$, where the notation means that $R' \in \mathcal{H} \otimes \mathcal{H}$ has ghost number 0 and has picture number -1 in each factor.

At this stage, we can now give a slightly different explanation of the fact that we should use vertex operators of picture number -1 in each factor, relying on the form of the duality map rather than on a knowledge of the factorization of the divisor at infinity in the moduli space. If we take R to act on states of picture number m , then after applying the duality map, R' will take values in $(\mathcal{H} \times \mathcal{H})_{0,m \otimes (-2-m)}$, with picture numbers m and $-2 - m$ in the two factors. There appears to be no moduli space suitable for computing superstring scattering amplitudes using vertex operators of positive or zero

picture number; on the other hand, the only way to make both m and $-2 - m$ negative is to set $m = -2 - m = -1$. So to understand the pole at $L_0 = 0$, we will have to use NS vertex operators with picture number -1 . A similar argument can be applied later in the Ramond sector to show that the residue of a pole should be computed with vertex operators of the canonical picture number $-1/2$.

The cohomology of Q_B acting on $(\mathcal{H} \otimes \mathcal{H})_{-1 \otimes -1}$ (where we allow states of any ghost number but set the picture number to -1 in each factor) is as follows:

1. In $N_{\text{gh}} = -1$, there is the operator $c\delta'(\gamma)$, with zero spacetime momentum. (In conventional language, this operator is related to the identity operator by picture-changing.)
2. In $N_{\text{gh}} = 0$, the cohomology can be identified with the space of states $\mathcal{V} = c\delta(\gamma)U$, where U is a dimension $1/2$ superconformal primary of the matter system.
3. In $N_{\text{gh}} = 1$, the cohomology can be identified with the space of states $\mathcal{V} = c\partial c\delta(\gamma)U$, with U as before.
4. In $N_{\text{gh}} = 2$, there is the operator $c\partial c\delta(\gamma)\partial\gamma$, with zero spacetime momentum.

Given this, the results of section 6.5.1 have straightforward analogs. The analog of eqn. (6.84), at non-zero spacetime momentum, is

$$(6.91) \quad R' = \sum_i c\delta(\gamma)U_i \otimes c\delta(\gamma)U^i + Q_B \mathcal{X}.$$

Eqn. (6.89) similarly has an obvious analog constructed by including factors of $\delta(\gamma)$, and eqn. (6.90) has an analog with $c\delta'(\gamma)$ and $c\partial c\delta(\gamma)\partial\gamma$ replacing 1 and $c\partial c\partial^2 c$.

The Ramond sector involves some new elements. The residue of the pole at $L_0 = 0$ is

$$(6.92) \quad R = b_0 \delta(\beta_0) G_0 \Pi_0.$$

This operator commutes with the ghost number, and increases the picture number by 1. We are interested in the case that the image of R – which is the operator that will be inserted on one of the two branches of a degenerating string worldsheet – has the canonical picture number $-1/2$. So we can view R as a linear transformation from $\mathcal{H}_{n,-3/2}$ to $\mathcal{H}_{n,-1/2}$, or in other words as an

element of $\oplus_n \mathcal{H}_{n,-1/2} \otimes \mathcal{H}_{n,-3/2}^*$. The duality (5.68) lets us identify R with an element

$$(6.93) \quad R' \in (\mathcal{H} \otimes \mathcal{H})_{1;-1/2 \otimes -1/2},$$

where the notation means that R' has overall ghost number 1 and has picture number $-1/2$ in each factor.

At picture number $-1/2$, the cohomology of Q_B acting on \mathcal{H} is nonvanishing only for ghost number $1/2$ or $3/2$. At ghost number $1/2$, the cohomology is generated by states $c\Theta_{-1/2}\Phi$, where $\Theta_{-1/2}$ is the $\beta\gamma$ ground state of picture number $-1/2$, and Φ is a matter primary of dimension $5/8$. Similarly, at ghost number $3/2$, the cohomology is generated by states $c\partial c\Theta_{-1/2}\Phi$.

Given the facts stated in the last paragraph, the general form of R' must be

$$(6.94) \quad R' = \sum_i c\Theta_{-1/2}\Phi_i \otimes c\Theta_{-1/2}\Phi'_i + Q_B\mathcal{X},$$

where Φ_i and Φ'_i are matter primaries of dimension $5/8$. The exact term $Q_B\mathcal{X}$ will decouple as usual, so the pole can be evaluated by insertions on the two branches of the physical states $c\Theta_{-1/2}\Phi_i$ and $c\Theta_{-1/2}\Phi'_i$. However, because of the factor of G_0 in the original formula (6.92) for R , the relation between Φ_i and Φ'_i is not as simple as in the cases that we have considered so far.

We will explain in detail what happens for the important case of massless Ramond states of uncompactified ten-dimensional superstring theory. The residue R maps the picture number $-3/2$ state $c\partial c\Theta_{-3/2}\Sigma^\alpha(p)$ to $c\Theta_{-1/2}p^I\Gamma_I^{\alpha\beta}\Sigma_\beta(p)$. On the other hand, the dual of $c\partial c\Theta_{-3/2}\Sigma^\alpha(p)$ is $c\Theta_{-1/2}\Sigma_\alpha(-p)$. So the contribution of massless Ramond states of momentum p to the residue is

$$(6.95) \quad \sum_{\alpha\beta} (p \cdot \Gamma)^{\alpha\beta} c\Theta_{-1/2}\Sigma_\alpha(p) \otimes c\Theta_{-1/2}\Sigma_\beta(-p).$$

To match this to what one would expect from field theory, we simply observe that $(p \cdot \Gamma)^{\alpha\beta}$ is the numerator of the Dirac propagator (describing the propagation of a massless fermion field of momentum p and specified $SO(1,9)$ chirality), while $c\Theta_{-1/2}\Sigma_\alpha(p)$ is the string theory operator that describes the coupling of such a fermion to other fields.

6.5.3. Analog for closed strings For the most part, the analog of all this for closed strings is fairly clear, so we will be brief. But a few points are worthy of note.

The residue of the pole in the closed bosonic string propagator is $R = \tilde{b}_0 b_0 \tilde{\Pi}_0 \Pi_0$, where $\tilde{\Pi}_0$ and Π_0 are respectively the projectors to $\tilde{L}_0 = 0$ and to $L_0 = 0$. By analogy to what we did for open strings, this can be converted to an element $R' \in (\mathcal{H} \otimes \mathcal{H})_4$, with \mathcal{H} being the space of physical states and the subscript indicating the total ghost number. Reasoning as for open strings gives at non-zero momentum

$$(6.96) \quad R' = \sum_i \tilde{c}c U_i \otimes \tilde{c}c U^i + \{Q_B, \mathcal{X}\}.$$

The sum runs over a complete set of conformal vertex operators. The conditions for dropping the $\{Q_B, \mathcal{X}\}$ term are the same as they were for open strings.

In studying anomalies, it will be important to know what happens if the residue R is replaced by $\tilde{R} = (b_0 - \tilde{b}_0) \tilde{\Pi}_0 \Pi_0$. This can be mapped to an element $\tilde{R}' \in (\mathcal{H} \times \mathcal{H})_5$. At non-zero momentum, the analog of (6.89) reads

$$(6.97) \quad \tilde{R}' = \sum_i (\tilde{c}c(\tilde{\partial}\tilde{c} + \partial c)U_i \otimes \tilde{c}c U^i + \tilde{c}c U_i \otimes \tilde{c}c(\tilde{\partial}\tilde{c} + \partial c)U^i) + \{Q_B, \mathcal{X}\}.$$

At zero momentum, there are exceptional contributions analogous to (6.90). The reason for this is that at zero momentum, the closed bosonic string has cohomology at ghost number 1, generated (for bosonic strings in \mathbb{R}^{26}) by the operators $c\partial X^I$ and $\tilde{c}\tilde{\partial} X^I$, $I = 0, \dots, 25$, that are related to momentum and winding number symmetries (see section 8.2.1), as well as cohomology at ghost number 4, obtained by multiplying those operators by, respectively, $\tilde{c}\tilde{\partial}\tilde{c}\tilde{\partial}^2\tilde{c}$ and $c\partial c\partial^2 c$. We can make an element of $(\mathcal{H} \otimes \mathcal{H})_5$ by multiplying exceptional classes of ghost number 1 and 4. At zero momentum, these classes appear in an exceptional contribution to \tilde{R}' :

$$(6.98) \quad \sum_I \left(\tilde{c}\tilde{\partial}\tilde{c}\tilde{\partial}^2\tilde{c} \cdot c\partial X^I \otimes c\partial X_I + c\partial X_I \otimes \tilde{c}\tilde{\partial}\tilde{c}\tilde{\partial}^2\tilde{c} \cdot c\partial X^I \right) + z \leftrightarrow \tilde{z}.$$

The ingredients needed to generalize this to closed superstrings should be fairly clear from what we said about open superstrings in section 6.5.2. Perhaps the only real subtlety concerns the generalization of the exceptional zero-momentum contribution (6.98). In the NS sector, the analog of $c\partial c\partial^2 c$ is $c\partial c\partial^2 c\delta(\gamma)\delta(\partial\gamma)$, and the analog of $c\partial X^I$ is $c\delta(\gamma)DX^I$. In the Ramond sector, there is no analog of the identity operator or of $c\partial c\partial^2 c$, but $c\partial X^I$ has an analog, namely the supersymmetry generator $\mathcal{S}_\alpha = c\Theta_{-1/2}\Sigma_\alpha$ studied in section 8.2.2. This is a $p = 0$ limit of the chiral Ramond vertex operator $c\Theta_{-1/2}\Sigma_\alpha e^{ip \cdot X}$. Because of the factor of p in the residue formula (6.95), which

arose from integration over the fermionic gluing parameter, it appears that there is no exceptional zero-momentum contribution to \tilde{R}' in Ramond sectors. Under certain conditions, there is a loophole in this reasoning, but we defer this to section 9.

6.5.4. Another look at the pole of the propagator Since the pole of the string theory propagator is so important, we will explain another way to look at it. For brevity, we consider only open strings or a chiral sector of closed strings. We start with bosonic string theory.

We consider gluing of two Riemann surfaces Σ_ℓ and Σ_r with local parameters x and y . We slightly generalize the usual gluing relation $xy = q$ so that the gluing is centered at $x = a$ and $y = b$:

$$(6.99) \quad (x - a)(y - b) = q.$$

We want to determine how the worldsheet path integral depends on a , b , and q , keeping other moduli fixed. Focusing on these three moduli makes sense when (but only when) q is small. The analysis we will give is suitable for understanding the singular behavior for $q \rightarrow 0$.

In a first pass, we will ignore insertions of operators other than the identity operator and represent the dependence of the path integral on the moduli a , b , and q by a three-form $\Omega = f(a, b, q) da db dq$. By translation symmetry, we assume that f is independent of a and b . The dependence of f on q can be determined from the scaling symmetry

$$(6.100) \quad (x, a) \rightarrow (\lambda x, \lambda a), \quad (y, b) \rightarrow (\tilde{\lambda} y, \tilde{\lambda} b), \quad q \rightarrow \lambda \tilde{\lambda} q.$$

of the gluing equation (6.99). This implies that f must be a constant multiple of $1/q^2$, so the contribution to the worldsheet path integral in which the operator inserted on each branch is the identity operator is

$$(6.101) \quad \Omega \sim da db \frac{dq}{q^2}.$$

The formula in the last paragraph involves integrated vertex operators in the sense of section 2.5. An analogous formulas with unintegrated vertex operators is simply

$$(6.102) \quad \Omega \sim c(a) \otimes c(b) \frac{dq}{q^2}.$$

The meaning of this formula is that to evaluate this contribution to the scattering amplitude, we are supposed to insert $c(a)$ on Σ_ℓ , and $c(b)$ on Σ_r , and integrate over q , as well as over the moduli of Σ_ℓ and Σ_r . Those latter moduli include a and b . The passage from unintegrated to integrated vertex operators replaces $c(a)$ and $c(b)$ by one-forms da and db . Notice that $c(a)$ and da transform the same way under scaling; indeed, $c(a)$ has mass dimension -1 (since it has $L_0 = -1$) and hence length dimension 1, just like da . The factor dq/q^2 in (6.102) is $dq q^{L_0-1}$ where $L_0 = -1$ for the operator c .

This discussion is rather formal, since the operator c is not Q_B -invariant and there is no way to isolate or define its contribution to a scattering amplitude. In fact, that is clear from eqn. (6.102): to evaluate the contribution of the operator c to a scattering amplitude, we would have to integrate the form dq/q^2 near $q = 0$, where that form is unintegrable. The operator c can be viewed as the tachyon vertex operator $c \exp(ik \cdot X)$ at $k = 0$, so the dq/q^2 singularity reflects the existence of the tachyon; there is no such singularity in tachyon-free theories. Let us consider a more general contribution in which (unintegrated) vertex operators \mathcal{V}_ℓ and \mathcal{V}_r are inserted on Σ_ℓ and Σ_r , respectively, and suppose that they have conformal dimension h_ℓ and h_r . This means that they scale under (6.100) as $\mathcal{V}_\ell \rightarrow \lambda^{-h_\ell} \mathcal{V}_\ell$, $\mathcal{V}_r \rightarrow \tilde{\lambda}^{-h_r} \mathcal{V}_r$. A scale-invariant generalization of (6.102) exists only if $h_\ell = h_r$, and takes the form

$$(6.103) \quad \mathcal{V}_\ell(a) \otimes \mathcal{V}_r(b) dq q^{L_0-1}$$

where we consider L_0 to act on \mathcal{V}_ℓ or \mathcal{V}_r , so that it can be set to h_ℓ or h_r . The integral over q near $q = 0$ now has a pole at $L_0 = 0$ and the residue of the pole can be computed by inserting the operator $\mathcal{V}_\ell(a) \otimes \mathcal{V}_r(b)$ on the two sides.

In generalizing this to superstring theory, for brevity we will consider only the NS sector. We start with super Riemann surfaces Σ_ℓ and Σ_r with local coordinates $x|\theta$ and $y|\psi$, respectively. We glue them by slightly generalizing (6.70) so that the gluing is centered at $a|\alpha \in \Sigma_\ell$ and $b|\beta \in \Sigma_r$:

$$(6.104) \quad \begin{aligned} (x - a - \alpha\theta)(y - b - \beta\psi) &= -\varepsilon^2 \\ (y - b - \beta\psi)(\theta - \alpha) &= \varepsilon(\psi - \beta) \\ (x - a - \alpha\theta)(\psi - \beta) &= -\varepsilon(\theta - \alpha) \\ (\theta - \alpha)(\psi - \beta) &= 0. \end{aligned}$$

For the moment, all we really need to know of this formula is the scaling symmetry

$$(x, a, \alpha) \rightarrow (\lambda x, \lambda a, \lambda^{1/2} \alpha)$$

$$(6.105) \quad \begin{aligned} (y, b, \beta) &\rightarrow (\tilde{\lambda}y, \tilde{\lambda}b, \tilde{\lambda}^{1/2}\beta) \\ \varepsilon &\rightarrow (\lambda\tilde{\lambda})^{1/2}\varepsilon. \end{aligned}$$

This is enough to determine that the analog of (6.101) is, in an obvious notation,

$$(6.106) \quad \Omega \sim [da|d\alpha] \otimes [db|d\beta] \frac{d\varepsilon}{\varepsilon^2}.$$

The analog of (6.102) is similarly

$$(6.107) \quad \Omega \sim c\delta(\gamma) \otimes c\delta(\gamma) \frac{d\varepsilon}{\varepsilon^2}.$$

Finally with general operator insertions, the analog of (6.103) is

$$(6.108) \quad \mathcal{V}_\ell(a|\alpha) \otimes \mathcal{V}_r(b|\beta) d\varepsilon \varepsilon^{2L_0-1}.$$

We can also write this in terms of $q_{\text{NS}} = -\varepsilon^2$ using

$$(6.109) \quad d\varepsilon \varepsilon^{2L_0-1} \sim dq_{\text{NS}} q_{\text{NS}}^{L_0-1}.$$

Again the integral over ε or q_{NS} has a pole at $L_0 = 0$, whose residue can be computed by inserting $\mathcal{V}_\ell(a|\alpha) \otimes \mathcal{V}_r(b|\beta)$ on the two sides.

Several well-known but relatively subtle facets of superstring theory follow from these simple formulas. Since the GSO projection comes from the sum over the sign of ε , and the form $d\varepsilon/\varepsilon^2$ in (6.106) is odd under $\varepsilon \rightarrow -\varepsilon$, we see that the identity operator – or equivalently the RNS tachyon – is removed by the GSO projection. Also, the behavior $d\varepsilon/\varepsilon^2$ in (6.106), compared to dq/q^2 in (6.101), shows that the tachyon mass squared in superstring theory is one-half what it is in bosonic string theory.

Returning to the bosonic string (the superanalog of what we are about to explain is fairly evident), the only transformations of the local coordinates x and y that we have considered so far are the affine linear transformations $x \rightarrow \lambda(x - a)$, $y \rightarrow \tilde{\lambda}(y - b)$. Suppose that we replace x and y with general local coordinates $\hat{x}(x)$ and $\hat{y}(y)$, and that the points $x = a$, $y = b$ correspond in the new coordinate system to $\hat{x} = \hat{a}$ and $\hat{y} = \hat{b}$. Then to first order in q , the gluing is described by

$$(6.110) \quad (\hat{x} - \hat{a})(\hat{y} - \hat{b}) = \hat{q},$$

where

$$(6.111) \quad \widehat{q} = q \frac{\partial \widehat{a}}{\partial a} \frac{\partial \widehat{b}}{\partial b}.$$

The fact that q is rescaled under a change of local parameters means that it is really not best understood as a complex number but as a section of a complex line bundle over the divisor \mathcal{D} at infinity; the fact that this rescaling is by the product of a function $\partial \widehat{a} / \partial a$ of the local parameters on Σ_ℓ times a function $\partial \widehat{b} / \partial b$ of the local parameters on Σ_r means that the line bundle in question is the tensor product of a line bundle over $\widehat{\mathcal{M}}_\ell$ with one over $\widehat{\mathcal{M}}_r$. For more on this, see, for instance, section 6 of [16].

Beyond first order in q , the relation between \widehat{q} and q becomes nonlinear. Ω can be expanded in powers of q with contributions of the form written in (6.103) that come from both primary fields and descendants. The descendant contributions can be related to the primary contributions using invariance under reparametrizations of Σ_ℓ and Σ_r , but in doing so, one has to use the nonlinear transformation of the gluing parameter under reparametrizations.

6.6. The Deligne-Mumford compactification and the integration cycle

Having introduced the Deligne-Mumford compactification in section 6.4, we can now complete the description given in section 3.3.2 of the integration cycle of superstring perturbation theory. In section 6.6.1, we consider the heterotic string. Since the idea of choosing an integration cycle as a step in formulating superstring perturbation theory may be unfamiliar, in section 6.6.2 we consider an example. Finally, in section 6.6.3, we consider the other superstring theories. The description we give here of the integration cycle resolves what in the 1980s was described as an ambiguity in superstring perturbation theory [48].

6.6.1. The Heterotic string We first recall the problem as presented in section 3.3.2. The worldsheet Σ of a heterotic string has holomorphic moduli $m_1 \dots m_p | \eta_1 \dots \eta_s$ and antiholomorphic moduli $\widetilde{m}_1 \dots \widetilde{m}_p$. Roughly speaking, one wants to integrate over the cycle Γ defined by taking the \widetilde{m}_i to be the complex conjugates of the m_i . But this is not well-defined (unless supermoduli space is projected) because of the possibility of shifting the m_i by expressions of quadratic and higher order in the η_j . In the interior of moduli space, we simply choose any integration cycle Γ such that $\overline{\widetilde{m}_i} = m_i$, modulo fermion bilinears. There is no natural choice, but any two choices are homologous. If,

therefore, Γ were compact, the choice of Γ would not matter at all; as it is, we have to specify how Γ should behave at infinity in moduli space.

We are now in a position to answer this question. The inspiration for the answer comes from the factorization described in section 6.4. For simplicity in terminology, let us consider a separating degeneration in the NS sector. The idea behind the factorization was roughly that the integral over the bosonic gluing parameter q_{NS} gives a pole $1/L_0$. At $q_{\text{NS}} = 0$, the other moduli of Σ are the moduli of Σ_ℓ and Σ_r (including the moduli of the points on Σ_ℓ and Σ_r at which these surfaces are glued together to make Σ), and the integral over those other moduli will then give the residue of the pole.

In more detail, the heterotic string worldsheet Σ has an antiholomorphic gluing parameter \tilde{q} as well as the holomorphic gluing parameter q_{NS} . The $1/L_0$ pole is supposed to come from an integration over \tilde{q} and q_{NS} keeping fixed the moduli of Σ_ℓ and Σ_r , and the pole specifically arises at $q_{\text{NS}} = \tilde{q} = 0$. This suggests that the definition of Γ should let us set $\tilde{q} = q_{\text{NS}} = 0$ without affecting the other moduli. To this end, the definition of Γ should include a condition

$$(6.112) \quad \tilde{q} = q_{\text{NS}}$$

or

$$(6.113) \quad \tilde{q} = q_{\text{NS}}(1 + \mathcal{O}(\eta_i\eta_j)),$$

but not

$$(6.114) \quad \tilde{q} = q_{\text{NS}} + \mathcal{O}(\eta_i\eta_j).$$

If as in (6.114), we were to add to the relation between \tilde{q} and q_{NS} a fermion bilinear with a coefficient that does not vanish at $q_{\text{NS}} = 0$, then naively speaking, we could not set $\tilde{q} = q_{\text{NS}} = 0$ without disturbing the other moduli.

It turns out that the right procedure in defining Γ is to impose (6.112) or (6.113), as opposed to (6.114), but the reason that this is necessary is a little subtle. The most critical case to understand is the case that the momentum flowing between Σ_ℓ and Σ_r is on-shell regardless of the external momenta. As usual, the problem first arises when there are two odd moduli. It is possible for the measure that must be integrated to compute a heterotic string amplitude to behave near $q_{\text{NS}} = \tilde{q} = 0$ as

$$(6.115) \quad \Xi = [d\tilde{q}; dq_{\text{NS}}|d\eta_1, d\eta_2] \tilde{q}^{-1}.$$

When this is the case, $\int \Xi$ is invariant under $q_{\text{NS}} \rightarrow q_{\text{NS}}(1 + \eta_1\eta_2)$ but not under $q_{\text{NS}} \rightarrow q_{\text{NS}} + \eta_1\eta_2$. Under any infinitesimal change of coordinates, $\int \Xi$ changes by a total derivative. For the particular case $q_{\text{NS}} \rightarrow q_{\text{NS}} + \eta_1\eta_2$, the shift in $\int \Xi$ is

$$(6.116) \quad \int \Xi \rightarrow \int \Xi + \int [d\tilde{q}; dq_{\text{NS}}|d\eta_1, d\eta_2] \eta_1\eta_2 \frac{\partial}{\partial q_{\text{NS}}} \frac{1}{\tilde{q}}.$$

But this is non-zero, because once we interpret q_{NS} as \tilde{q} , the expression $\partial_{q_{\text{NS}}}(1/\tilde{q})$ has a delta function contribution at $q_{\text{NS}} = \tilde{q} = 0$. For more detail on how this phenomenon arises in heterotic string computations, see section 6.6.2 and appendix D.

Once one knows that a condition is required on the behavior of the integration cycle at infinity, the right condition must be (6.112) or more precisely (6.113), since this is the only condition that can be stated using the data at hand.

Indeed, because of facts explored in section 6.3 of [16] and also at the end of section 6.5.4 above, only the more general condition (6.113) really makes sense, and not the more restrictive (6.112). In brief, \tilde{q} and q_{NS} are really not complex numbers but sections of complex line bundles that we will call $\tilde{\mathcal{L}}$ and \mathcal{L} , respectively. ($\tilde{\mathcal{L}}$ and \mathcal{L} are line bundles over the left and right moduli spaces \mathcal{M}_L and \mathcal{M}_R , respectively.) By \tilde{q} or q_{NS} we mean a holomorphic section of $\tilde{\mathcal{L}}$ or of \mathcal{L} that has a simple zero on the compactification divisor. This defines them only up to

$$(6.117) \quad \tilde{q} \rightarrow e^{\tilde{\varphi}} \tilde{q}, \quad q_{\text{NS}} \rightarrow e^{\varphi} q_{\text{NS}},$$

for some functions $\tilde{\varphi}$ and φ . Because of this, a condition as precise as (6.112) would not be well-defined. However, once one reduces modulo the odd moduli $\eta_1 \dots \eta_s$ and identifies $\mathcal{M}_{L,\text{red}}$ as the complex conjugate of $\mathcal{M}_{R,\text{red}}$, $\tilde{\mathcal{L}}$ is the complex conjugate of \mathcal{L} . Therefore, a well-defined condition is that $\tilde{q} = q_{\text{NS}} e^{\phi}$ where ϕ vanishes modulo the odd moduli. This is what we have written in (6.113).

At this point, we can complete the description of the heterotic string integration cycle Γ . An inductive procedure is involved. When Σ degenerates, we impose the condition (6.113). Beyond this, as long as Σ_ℓ and Σ_r are smooth, Γ may be any cycle of the right dimension whose reduced space is the diagonal in $\mathcal{M}_{L,\text{red}} \times \mathcal{M}_{R,\text{red}}$. When Σ_ℓ or Σ_r degenerates, one needs a further condition with the same form as (6.113). This continues until we finally reach a maximal degeneration, such as the one depicted schematically in fig. 11.

It is noteworthy that the condition (6.113) that we actually need makes sense, and a more precise condition (6.112) that we do not need does not make sense. This is the story of superstring perturbation theory: precisely what one needs is true, and in general no more.

The condition (6.113) is formulated in terms of $q_{\text{NS}} = -\varepsilon^2$ rather than in terms of ε , so for given \tilde{q} , the sign of ε is not fixed; the sum over this sign leads to the GSO projection. (Changing the sign of ε while keeping the other variables fixed only makes sense for $\varepsilon \rightarrow 0$, but the poles at $L_0 = 0$ come from the behavior for $\varepsilon \rightarrow 0$, so the sum over the two signs of ε does make sense in analyzing which states contribute a pole at $L_0 = 0$.)

For a Ramond degeneration of the heterotic string, we simply replace q_{NS} by q_{R} in the foregoing. For the other superstring theories, just a few minor modifications are needed; see section 6.6.3. For a treatment of some of these issues in terms of conditions on the placement of picture-changing operators, see [64].

6.6.2. An example An example explored in the literature [39, 42, 43] gives a good illustration of these ideas. In this case, the degeneration involves a collision of two NS vertex operators. On a heterotic string worldsheet with local coordinates $\tilde{z}; z|\theta$, we consider two NS punctures, at say $\tilde{z}; z|\theta = \tilde{u}_1; u_1|\zeta_1$ and $\tilde{u}_2; u_2|\zeta_2$. The degeneration occurs at $\tilde{u}_1 \rightarrow \tilde{u}_2$, $u_1 \rightarrow u_2$. The gluing parameters are

$$(6.118) \quad \begin{aligned} \tilde{q} &= \tilde{u}_1 - \tilde{u}_2 \\ q_{\text{NS}} &= u_1 - u_2 - \zeta_1 \zeta_2. \end{aligned}$$

The only nontrivial point is the term $-\zeta_1 \zeta_2$ in q_{NS} . This term is easily motivated from global supersymmetry (it is determined by invariance under the global supersymmetry generator $G_{-1/2} = \partial_\theta - \theta \partial_z$) and is derived⁵⁴ in section 6.3.2 of [16].

At this stage, we need to know one fact that will be more systematically developed in section 7.6. Infrared singularities in string theory are regularized by placing a lower bound on the magnitude of the gluing parameters. In the present example, the cutoff can be $|\tilde{q}| \geq \epsilon$, where we take $\epsilon \rightarrow 0$ at the end of the computation. Since the integration cycle is defined by a condition such as $\tilde{q} = q_{\text{NS}}$, the infrared cutoff also bounds q_{NS} away from zero.

Now let $u_{12} = u_1 - u_2$ and $\tilde{u}_{12} = \tilde{q} = \tilde{u}_1 - \tilde{u}_2$. The reasoning of section 6.6.1, specialized to this situation, together with the infrared cutoff procedure

⁵⁴The factor of 1/2 in eqn. (6.28) of that paper is inessential, since $q_{\text{NS}} = -\varepsilon^2$ is only defined up to $q_{\text{NS}} \rightarrow q_{\text{NS}} e^\phi$.

stated in the last paragraph, amounts to saying that near $u_{12} = \tilde{u}_{12} = 0$, instead of integrating over \tilde{u}_{12} , u_{12} , ζ_1 , and ζ_2 with $\overline{\tilde{u}_{12}} = u_{12}$, we want to integrate over \tilde{q} , q_{NS} , ζ_1 , and ζ_2 with $\overline{\tilde{q}} = q_{\text{NS}}$. This is precisely the conclusion of the above-cited papers, where it is shown that the alternative procedure leads to the wrong answer.

In appendix D, we give a more detailed and precise account of this example. See also [28] for much more detail.

6.6.3. Other superstring theories Here we extend the description of the integration cycle to Type II and Type I superstring theories. There are only a few details to explain.

Let Σ be a Type II superstring worldsheet. It has holomorphic and antiholomorphic moduli $m_1 \dots m_p | \eta_1 \dots \eta_s$ and $\tilde{m}_1 \dots \tilde{m}_p | \tilde{\eta}_1 \dots \tilde{\eta}_{\tilde{s}}$. (The numbers of holomorphic and antiholomorphic even moduli are always equal, but this is not so in general for the odd moduli; s and \tilde{s} can be unequal if there are operator insertions of types NS-R and/or R-NS.)

Let us consider for definiteness in the notation a degeneration of NS-NS type. There are then holomorphic and antiholomorphic gluing parameters q_{NS} and \tilde{q}_{NS} . (For other degenerations, simply replace \tilde{q}_{NS} by \tilde{q}_{R} and/or q_{NS} by q_{R} in the following.) As in the case of the heterotic string, roughly speaking, we want to constrain the behavior at infinity of the integration cycle Γ by setting $\overline{\tilde{q}_{\text{NS}}} = q_{\text{NS}}$. However, there is a problem here: \tilde{q}_{NS} is a function on a complex supermanifold (the moduli space of super Riemann surfaces with punctures) and there is no natural notion of complex-conjugating a function on a complex supermanifold. The problem becomes obvious if we recall that \tilde{q}_{NS} is only well-defined up to multiplication by a function $\exp(\tilde{\phi})$ where $\tilde{\phi}$ may depend on the $\tilde{\eta}_j$. So it does not make much sense to complex conjugate \tilde{q}_{NS} without being able to complex conjugate the $\tilde{\eta}_j$, but the complex conjugates of the $\tilde{\eta}_j$ are certainly not part of the formalism.

Writing the relation between \tilde{q}_{NS} and q_{NS} in the form $\tilde{q}_{\text{NS}} = \overline{q}_{\text{NS}}$ would merely move the problem from antiholomorphic to holomorphic degrees of freedom. For the heterotic string, we were able to avoid this issue, because \tilde{q} was a function on an ordinary complex manifold, where it makes sense to take the complex conjugate of a function.

The appropriate procedure (see [15], especially section 5, for background) is to describe the relation between \tilde{q}_{NS} and q_{NS} in parametric form. Γ is a smooth cs supermanifold, meaning that it is parametrized by even and odd coordinates $t^1 \dots t^{2p} | \zeta^1 \dots \zeta^{s+\tilde{s}}$, where the t^i can be considered real modulo the odd variables ζ^j . The locus at infinity in Γ is of codimension $2|0$ and we

pick coordinates so that this locus is defined by $t^1 = t^2 = 0$. Then q_{NS} and \tilde{q}_{NS} are defined in terms of functions on Γ by any condition of the form

$$(6.119) \quad \begin{aligned} q_{\text{NS}}(1 + \mathcal{O}(\eta^2)) &= t^1 + it^2 \\ \tilde{q}_{\text{NS}}(1 + \mathcal{O}(\tilde{\eta}^2)) &= t^1 - it^2. \end{aligned}$$

In this way, one formulates a relation that roughly corresponds to $\bar{q}_{\text{NS}} = q_{\text{NS}}$ or $\tilde{q}_{\text{NS}} = \bar{q}_{\text{NS}}$ without complex-conjugating either \tilde{q}_{NS} or q_{NS} .

Since the asymptotic behavior of Γ has been defined via conditions on q_{NS} and \tilde{q}_{NS} , the signs of the gluing parameters ε and $\tilde{\varepsilon}$ are unspecified. The sum over these signs leads to the chiral GSO projection – separate GSO projections for holomorphic and antiholomorphic modes. As usual, reversing the signs of ε and $\tilde{\varepsilon}$ while keeping fixed the other variables only makes sense for $\varepsilon, \tilde{\varepsilon} \rightarrow 0$, but this suffices for studying the pole at $L_0 = 0$.

For open superstrings, a basic fact to start with (see section 7.4 of [16] as well as section 9.2.1 below for more detail on much that follows) is that on an ordinary Riemann surface with boundary, the gluing parameter q associated to an open-string degeneration is real and positive (modulo the odd variables). In particular, an open-string degeneration is associated to a boundary of moduli space at $q = 0$. The same is true of certain closed-string degenerations of open and/or unoriented strings. Because of this, the integration cycle Γ of open and/or unoriented superstring theory is going to be a supermanifold with boundary. This notion is a little delicate; see for example section 3.5 of [15]. The essential point is that to make sense of supermanifolds with boundary and especially to make sense of integration on them, one needs an equivalence class of nonnegative functions ρ with a first order zero along the boundary, modulo rescaling $\rho \rightarrow \rho e^f$ for a function f . Roughly speaking, ρ will be the gluing parameter q_{NS} or q_{R} . A fuller explanation involves the following.

For bosonic strings, there is a natural moduli space of open and/or unoriented Riemann surfaces. But this does not appear to have a superstring analog.⁵⁵ Just as we explained for closed oriented superstrings in section 3.3.2, it seems that open and/or unoriented superstring world sheets have a natural integration cycle (up to homology) but not a natural moduli space. To construct the integration cycle, we begin by associating to an open and/or unoriented superstring worldsheet Σ a certain double cover that is a closed oriented super Riemann surface $\hat{\Sigma}$. Let \mathfrak{M} be the moduli space that parametrizes

⁵⁵It does have an analog in the nonsupersymmetric and tachyonic Type 0 string theory, in which holomorphic and antiholomorphic odd coordinates are complex conjugates.

deformations of $\widehat{\Sigma}$. It is a complex supermanifold, say of complex dimension $p|s$, with even and odd moduli $m_1 \dots m_p | \eta_1 \dots \eta_s$. The integration cycle for open and/or unoriented super Riemann surfaces is a (real) codimension $p|0$ cycle $\Gamma \subset \widehat{\mathfrak{M}}$ that is defined, roughly speaking, by taking the m_i to be real, modulo the odd variables. To be more exact, inside the reduced space $\widehat{\mathfrak{M}}$, there is a natural real cycle Γ of codimension p that parametrizes the deformations of the reduced space of Σ , and Γ is defined by thickening this in the fermionic directions.

Just as for closed oriented superstrings, Γ is not uniquely determined, but the possible choices are homologous. So as usual, all we need to do is to put a condition on how Γ should behave at infinity. For this, we require first a few generalities. Γ will be again a smooth cs supermanifold with coordinates $t^1 \dots t^p | \zeta^1 \dots \zeta^s$, where the t^i can be considered real modulo the ζ^j . But now, as observed above, Γ will be a supermanifold with boundary. This means that we can pick the coordinates, locally, so that t^1 plays a special role; $t^1 \geq 0$ on Γ and the boundary of Γ is defined by $t^1 = 0$. Of course, the function t^1 with these properties is not uniquely determined; it is determined modulo $t^1 \rightarrow e^f t^1$, where f is a function on Γ . (Importantly, this does not allow a substitution such as $t^1 \rightarrow t^1 + \zeta^1 \zeta^2$, though it does allow $t^1 \rightarrow t^1(1 + \zeta^1 \zeta^2)$.) The integration cycle Γ for open superstrings near, say, an NS degeneration is constrained to obey

$$(6.120) \quad t^1 = q_{\text{NS}}(1 + \mathcal{O}(\eta^2)).$$

Here q_{NS} is the complex gluing parameter of the double cover $\widehat{\Sigma}$, defined in the usual way. (For a Ramond degeneration, just replace q_{NS} by q_{R} .) What we have done, roughly speaking, is to put the assertion that q_{NS} is real and nonnegative near the boundary of Γ in parametric form, without mentioning the complex conjugate of q_{NS} or making any other claim that does not make sense on a complex supermanifold.

Another way to express the idea of the last paragraph is to say that the boundary of the integration cycle Γ is contained in the divisor $\mathcal{D} \subset \mathfrak{M}$ defined by $q_{\text{NS}} = 0$.

For Type II superstrings as well as for open and/or unoriented superstrings, what we have explained is the condition that should be imposed on the integration cycle when a single degeneration occurs. For multiple degenerations, one places such conditions on the gluing parameters at each degeneration.

7. BRST anomalies, massless tadpoles, and all that

7.1. BRST anomalies

At many points in the present paper, beginning in section 2.4.3, we have seen that the proof of gauge-invariance and related properties relies on integration by parts on moduli space. One simply uses eqn. (2.48) together with Stokes's theorem (or the superanalogs of these formulas) to express a scattering amplitude with insertion of a BRST-trivial state such as $Q_B \mathcal{W}_1$ in terms of a boundary integral:

$$(7.1) \quad \int_{\Gamma} F_{\{Q_B \mathcal{W}_1\}, \nu_2, \dots, \nu_n} = - \int_{\partial \Gamma} F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}.$$

Here Γ is the integration cycle of the worldsheet path integral, as described in section 6.6. (For bosonic string theory, it is simply the moduli space $\mathcal{M}_{g,n}$.) The scattering amplitude with insertion of $Q_B \mathcal{W}_1$ is zero if the boundary integral on the right hand side of (7.1) vanishes.

Actually, this phrasing of the problem is a little too schematic. In closed oriented string theories, Γ has no boundary, but the forms that must be integrated over Γ are singular along a divisor \mathcal{D} (which has multiple components, corresponding to different ways that Σ may degenerate). We place a cutoff on the integral on the left of (7.1) by removing a small neighborhood of \mathcal{D} ; details of this will be described. By $\partial \Gamma$, we mean the boundary of the resulting cutoff version of Γ . Then we consider the limit in which the cutoff is removed; there is no anomaly if the right hand side of (7.1) vanishes in this limit.

By now we have assembled the ingredients to understand when a problem may actually arise. Infinity in moduli space is the region in which one of the gluing parameters vanishes. The gluing parameters are $q = e^{-s}$ for open strings or $q = e^{-(s+i\alpha)}$ for closed strings, and vanish for $s \rightarrow \infty$. We cut off Γ by placing an upper bound on s , so to get a component of $\partial \Gamma$, we simply set s to a large value⁵⁶ rather than integrating over it.

The basic idea of what will happen is visible for open bosonic strings. Propagation of an open string through a proper time s is described by a factor

⁵⁶More precisely, we set s to be the sum of a large constant and an arbitrary, but fixed, function of the remaining worldsheet moduli. This more careful statement is needed because q is a section of a complex line bundle rather than a complex number, as explained at the end of section 6.5.4. We cannot simply set q to 0 or s to ∞ , since $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$ is typically singular at $q = 0$. See section 7.7 for more on these points.

$\exp(-sL_0)$. To construct the open bosonic string propagator in section 6.1.1, we multiplied by b_0 and integrated over s , giving

$$(7.2) \quad b_0 \int_0^\infty ds \exp(-sL_0).$$

In looking for a possible boundary contribution on the right hand side of a formula like eqn. (7.1), we do not want to integrate over s , but rather to set it to a large value. So we drop the b_0 factor and the integration in (7.2). We are left simply with a factor $\exp(-sL_0)$, so a boundary term can only arise if $\exp(-sL_0)$ is nonvanishing for large s .

An immediate consequence is that only certain types of degeneration can generate anomalies. Let us consider first a typical separating degeneration, such as that of fig. 9 in section 6.4.2, with more than one external particle on each side. The momentum P flowing through the separating line is a sum of several external momenta and is generically not on-shell. Moreover, for suitable external momenta, P^2 can have any real or complex value. In this situation, there is no possible anomaly. This follows from the relation $L_0 = (\alpha'/4)P^2 + N$, where N is the mass squared operator of the string. In a region of external momenta in which the real part of P^2 is sufficiently positive, $\exp(-sL_0)$ vanishes for large s . So for such external momenta, there is no surface term at large s . The general result for the scattering amplitude can be obtained by analytic continuation from the region just indicated, and so possesses no anomaly associated to a degeneration of this type.

A somewhat similar argument shows that there are never anomalies associated to nonseparating degenerations. Nonseparating degenerations only arise in loop amplitudes (such as the one-loop diagram of fig. 4(b) in section 6.1.1), and loop amplitudes do not make sense in theories with tachyons. So to discuss anomalies associated to nonseparating degenerations, we should consider only tachyon-free theories, that is we should assume that $N \geq 0$ (possibly after projecting to the GSO-invariant part of the spectrum in the case of superstring theory). In the case of a nonseparating degeneration, the momentum P flowing through the separating line is an integration variable (the loop momentum). The loop integration can be performed over the cycle on which P is real in Euclidean signature, so that L_0 is positive semidefinite and vanishes only when N and P both vanish. Away from $N = P = 0$, $\exp(-sL_0)$ vanishes exponentially. Since $P = 0$ has measure zero, there is no anomaly associated to the fact that $L_0 = 0$ just for $P = 0$. (Even for $N = 0$, the integral $\int d^D P \exp(-sL_0) = \int d^D P \exp(-s(\alpha'/4)P^2)$ vanishes for large s , though only as a power of s .)

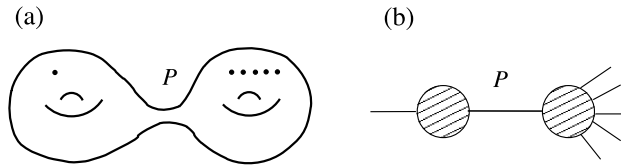


Figure 13: With only one external particle to the left (or right) of a separating degeneration, the momentum P flowing through the separating line is constrained to be on-shell. This is the configuration associated with mass renormalization. The configuration is sketched in string theory and in field theory in (a) and (b), respectively.

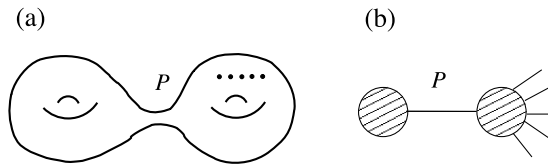


Figure 14: With no external particles at all to the left (or right) of a separating degeneration, the momentum P flowing through the separating line automatically vanishes. Hence a massless particle flowing through this line is on-shell. This is the most troublesome configuration for superstring perturbation theory.

Going back to separating degenerations, there are two critical cases that actually do cause trouble. These are the cases, sketched in figs. 13 and 14, in which the momentum P flowing through the separating line is automatically on-shell. (The figures are drawn for closed strings, though this involves jumping slightly ahead of our story.) In fig. 13, there is just one external particle to the left of the separating line, so P is equal to the momentum of that external particle and in particular is constrained to be on-shell. In fig. 14, there are no external particles at all on the left, and momentum conservation forces $P = 0$, which is on-shell in the case of a massless particle.

These then are the troublesome cases for superstring perturbation theory. We will review the analogous issues in field theory in section 7.2, after which we explain some basics of what happens in string theory. The issues that arise here will occupy us for most of the rest of this paper.

What we have said so far generalizes straightforwardly to the other string theories. For example, consider closed bosonic string theory. To extract a

boundary term involving a closed-string degeneration, we modify eqn. (6.53) for the closed-string propagator in the following way. We keep the integration over α and the associated factor of $\Psi_\alpha = b_0 - \tilde{b}_0$. But we set s to a large constant and omit the factor of Ψ_s that is associated to the s integral. We are left with

$$(7.3) \quad 2\pi(b_0 - \tilde{b}_0)\delta_{L_0 - \tilde{L}_0} \exp(-s(L_0 + \tilde{L}_0)).$$

By essentially the same arguments that we have given for open strings, an anomaly can only arise from states of $L_0 = \tilde{L}_0 = 0$, and only for separating degenerations of the types sketched in figs. 13 and 14. In passing from bosonic string theory to superstring theory, the only modifications of the above formulas are the obvious factors of $\delta(\beta_0)G_0$ in the Ramond sector. These factors do not affect the discussion of which types of degeneration can contribute anomalies.

7.2. Review of field theory

Before trying to understand what is happening in string theory, it helps to review how the troublesome cases are understood in field theory.

The two cases are completely different. In fig. 13, let the external particle flowing in from the left have momentum p and mass m . Momentum conservation implies that the momentum P of the separating internal line is simply equal to p . In an S -matrix element, the external particle is always on-shell, so $p^2 + m^2 = 0$. Consequently, if we directly evaluate the contribution to the S -matrix from a Feynman diagram such as that of fig. 13(b), we are sitting on top of the pole $1/(P^2 + m^2)$ of the propagator. We cannot simply calculate the S -matrix by summing Feynman diagrams with on-shell external particles.

The solution to this problem is well-known. Instead of computing an S -matrix element directly, one introduces a local field \mathcal{O} that can create the particle in question from the vacuum, and one computes matrix elements of \mathcal{O} . This enables one to vary the momentum p (and therefore also P) away from its mass-shell, and then by searching for a pole in the matrix element of \mathcal{O} , one recovers the appropriate S -matrix element. The effect of diagrams such as that of fig. 13(b) is to shift the position of the pole away from its tree-level value; this is called mass renormalization. Mass renormalization generally⁵⁷ does not affect the existence of a sensible perturbation expansion for a given theory. A physically sensible perturbation expansion remains physically sensible, whatever the mass renormalization effects may be.

⁵⁷ There is a potential exception in the case of a theory that at tree level has a massless scalar; perturbative effects of mass renormalization might make the scalar

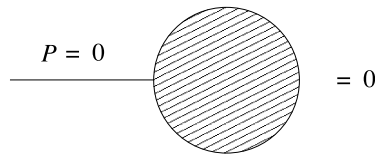


Figure 15: The shaded region represents symbolically the “tadpole” of a massless scalar field – its amplitude to disappear into the vacuum. In general in quantum field theory, to make possible a sensible perturbation expansion around a chosen classical vacuum, the tadpoles of massless scalar fields must vanish.

Quite different is the configuration of fig. 14. There are no particle insertions at all to the left of the separating line, so the momentum flowing through that line is $P = 0$. Thus the separating line is on-shell if and only if it is associated to a field of mass 0. Let us call this field w . The left half of fig. 14(b) represents a contribution to the “tadpole” of w – the amplitude for a single w quantum to disappear into the vacuum. By Lorentz invariance, this tadpole vanishes unless w is Lorentz-invariant. In practice, w is either a massless scalar field or else the trace of the metric tensor.

If such a massless tadpole is non-zero, then the associated contributions to the scattering amplitude are proportional to $1/0$ – that is, they are proportional to the scalar propagator $1/(P^2 + m^2)$ at $P = m = 0$. Accordingly, for a theory to have a perturbation expansion around an initially assumed classical vacuum, the tadpoles for all massless scalars, computed in that vacuum, must vanish (fig. 15). Otherwise, one has to search for an extremum of an appropriate quantum effective potential. Nonvanishing of the tadpole means that this extremum will not coincide with the initially assumed value.

What can happen in detail depends on the underlying classical potential for the field w . Rather than trying to be general, let us discuss the situation that prevails in ten-dimensional supergravity and in other theories that arise

tachyonic, in which case the theory has no perturbative expansion near the originally considered classical vacuum. In practice, if there is a good reason – such as supersymmetry or a spontaneously broken continuous bosonic symmetry – to have a massless scalar at tree level, this usually prevents mass renormalization for the scalar in question. However, if (as in three-dimensional theories with only two supercharges, for example), supersymmetry does allow perturbative mass renormalization for massless scalars, then perturbation theory must be developed around a minimum of an appropriate effective potential, not around an arbitrary classical vacuum.

as low energy limits of superstring compactifications to $d > 2$ Lorentzian dimensions. Whenever there is a perturbative string theory, there is always at least one massless scalar field at the classical level – the dilaton field ϕ . Perturbation theory is an expansion in $g_{\text{st}} = e^\phi$, which is called the string coupling constant. In particular, the effective action at string tree level is (in the closed-string sector) proportional to $1/g_{\text{st}}^2 = e^{-2\phi}$, and so the classical potential for ϕ , if not identically zero, is a multiple of $e^{-2\phi}$. Since this function has no stationary point as a function of ϕ , a classical vacuum that could represent the starting point of superstring perturbation theory can only exist if the potential is identically zero as a function of ϕ at the classical level. By familiar arguments [68], the relationship $g_{\text{st}} = e^\phi$ implies that a stable vacuum accessible to perturbation theory exists if and only if the effective potential for the dilaton field is identically zero in perturbation theory, not just classically (in particular, the dilaton mass renormalization must vanish as well as the dilaton tadpole). The basic idea here is that if a non-zero effective potential is generated in loops, then for weak coupling or $\phi \ll 0$, it is dominated by the non-zero contribution of lowest order and has no critical point.⁵⁸ If other moduli are present – in addition to the expectation value of ϕ – this argument holds for any values they may have. So for perturbation theory to work, the effective potential must vanish identically as a function of the moduli of the compactification – or at least all moduli that parametrize vacua that are accessible to perturbation theory.

7.2.1. The importance of spacetime supersymmetry Such a result will not hold without a special reason. The reason usually considered is spacetime supersymmetry. Typically, in the field theory that arises as the low energy limit of a supersymmetric string compactification to $d > 3$ dimensions (or a compactification with more than 2 unbroken supercharges to $d = 3$), it is straightforward to show that supersymmetry remains unbroken to all orders of perturbation theory. Conversely, in the relatively rare cases that supersymmetry is unbroken at tree level but is spontaneously broken by loop effects, it is typically possible to predict this using the low energy effective field theory.

⁵⁸The genus g contribution to the effective potential is a multiple of $g_{\text{st}}^{2g-2} = \exp((2g-2)\phi)$. When one adds this to the kinetic energy of the metric and the gravitational field, which at the classical level is a multiple of $e^{-2\phi}(R + 4\partial_I\phi\partial^I\phi)$, one finds that to get a classical solution in a homogeneous spacetime – even with a cosmological constant – one needs a cancellation between contributions to the effective potential associated to different values of g . Such cancellations may occur but only if g_{st} is sufficiently large, so they are not accessible to perturbation theory.

For example, the arguments that are relevant for analyzing the behavior in perturbation theory of a compactification to 4 dimensions with $\mathcal{N} = 1$ supersymmetry are well-known. If absent at tree level, a tadpole for the dilaton and other classical moduli can be triggered only by loop corrections to the superpotential for chiral superfields, or a D -term for a $U(1)$ gauge field. A simple argument using holomorphy⁵⁹ shows that a superpotential is not generated in loops. As for a D -term, this can be generated in perturbation theory for a $U(1)$ gauge field that lacks a D -term at tree level, but only at the one-loop level.⁶⁰ These are the necessary facts to understand the fate in perturbation theory of a classical supersymmetric vacuum of closed oriented strings. For open and/or unoriented strings, one also has to analyze one-loop anomalies, which we defer to section 9.

Supersymmetry breaking by loop effects is only harder to come by in models with more unbroken supersymmetry at tree level, including all supersymmetric compactifications with $d > 4$. Rather than reviewing all the possible cases here, we will just jump to $d = 10$. A rather general reasoning shows that if a supersymmetric action is going to have a non-vanishing potential energy function $V(\phi)$ (or possibly $V(\phi, a)$ for Type IIB supergravity, which has a second scalar field a as well as the dilaton ϕ), then the supersymmetrization of this interaction will require a non-derivative Yukawa-like coupling $W_{ij}(\phi)\psi^i\psi^j$ that is bilinear in fermion fields ψ^i . In general, these fermion fields might have spin 1/2 and/or spin 3/2. In most of the ten-dimensional supergravity theories, Lorentz invariance makes it impossible to write such a Yukawa-like coupling. (For example, in the supergravity limit of the heterotic string, the massless neutral fermions are a spin 3/2 field ψ_I of one chirality and a spin 1/2 field λ of opposite chirality; no Lorentz-invariant bilinear ψ^2 , $\psi\lambda$, or λ^2 exists. There are also massless chiral fermions in the adjoint representation of the gauge group, and again there is no Lorentz-invariant and gauge-invariant fermion bilinear.) The only ten-dimensional supergravity theory in which a non-zero effective potential (including a constant potential – a cosmological constant) is not excluded by this simple argument is the Type IIA supergravity. This theory actually does admit a supersymmetric deformation with a nontrivial potential [70]. However, this deformation violates

⁵⁹The superpotential depends holomorphically on a complex field whose real part is the dilaton and whose imaginary part is an axion-like field that in perturbation theory decouples at zero momentum. This decoupling, plus the known dependence of loop corrections on the dilaton field, implies [69] that a superpotential cannot be generated in loops.

⁶⁰The Fayet-Iliopoulos D -term is a constant, independent of all fields. The known dependence of loop corrections on the dilaton field implies [44] that such a constant can only be generated at one-loop order.

the symmetry $(-1)^{F_L}$, which assigns the value -1 (or $+1$) to states in the left-moving Ramond (or NS) sector, and which is conserved in perturbative Type IIA superstring theory. When supplemented by that last statement, low energy field theory predicts that a nontrivial potential cannot be generated in any of the ten-dimensional superstring theories in perturbation theory.

We have here emphasized tadpoles for massless scalars, rather than for the trace of the metric tensor. From a field theory point of view, in the absence of massless scalars (or at least in the absence of massless scalar tadpoles), a tadpole for the trace of the metric tensor simply represents the generation of a cosmological constant. Assuming that this tadpole is finite and small, a perturbative expansion might exist in de Sitter or Anti de Sitter spacetime. This situation is not realized in perturbative string theory (at least not for $d > 2$), because the structure of string perturbation theory is such that if the dilaton tadpole vanishes, the vacuum energy also vanishes and there is no tadpole for the trace of the metric. It follows that in string theory compactifications, it is not necessary to study a tadpole for the trace of the metric separately from the dilaton tadpole.

7.2.2. A subtlety at zero momentum It probably is no coincidence that in addition to it being unnecessary to analyze a tadpole for the trace of the metric separately from that for the dilaton, it is also difficult to do so. The reason for the last statement is that although at non-zero momentum, there are separate conformal or superconformal vertex operators for the dilaton and the graviton, at zero momentum there is only a single Lorentz-invariant conformal or superconformal vertex operator, which represents the coupling of a linear combination of the dilaton and the trace of the metric.

For example, for bosonic strings, the operator in question is $\tilde{c}\eta_{IJ}\tilde{\partial}X^I\partial X^J$, while for the heterotic string, it is $\tilde{c}\delta(\gamma)\eta_{IJ}\tilde{\partial}X^ID_\theta X^J$. (Here η_{IJ} is the Lorentz metric.) These formulas have simple modifications for the other superstring theories. In each case, there is no other conformal or superconformal primary with the same quantum numbers. This operator couples to a linear combination of the dilaton and the trace of the metric, as analyzed in [71].

It is also true from the point of view of the low energy effective field theory that there is only one tadpole condition at g -loop order. That is because the g -loop contribution to the effective potential has a known dependence on the dilaton ϕ and the spacetime metric G , namely it is proportional to $\sqrt{\det G}\exp((2g-2)\phi)$.

It is possible to construct separate Q_B -invariant vertex operators for the dilaton and the trace of the metric, but the “second” vertex operator is not conformal or superconformal. As we remarked in section 2.4.4, methods to

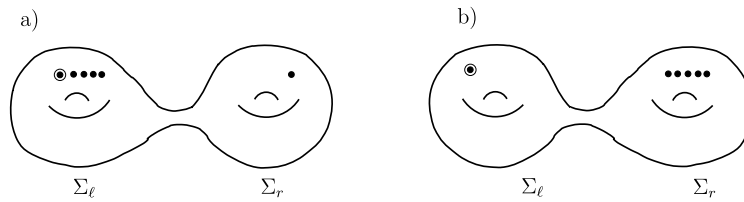


Figure 16: A string worldsheet Σ degenerates to a union of two branches Σ_ℓ and Σ_r in a manner related to mass renormalization and anomalies. A dot represents a generic conformal vertex operator which may be any of $\mathcal{V}_2, \dots, \mathcal{V}_n$, while a dot surrounded by a circle represents the Q_B -trivial vertex operator $\mathcal{V}_1 = \{Q_B, \mathcal{W}_1\}$. There are two interesting configurations – a separating degeneration (a) on which all but 1 of $\mathcal{V}_2, \dots, \mathcal{V}_n$ are on the same branch as \mathcal{V}_1 , and a separating degeneration (b) in which they are all on the opposite branch.

compute with such more general vertex operators are known [17, 18], but involve some extra machinery beyond what we have described in the present paper. For our purposes, we do not need to develop this machinery, since the tadpole of an operator that is not superconformal cannot appear in a superconformal formalism.

7.3. Back to string theory

The reader may have noticed that although we began in section 7.1 with a discussion of BRST anomalies in string theory, the field theory discussion of section 7.2 focused on infrared singularities – poles associated to mass renormalization and massless tadpoles – rather than anomalies. The two effects are different, but closely related. Both involve the effects of on-shell intermediate particles.

It is possible to understand heuristically why, in string theory, mass renormalization and massless tadpoles are associated to BRST anomalies. BRST symmetry in string theory incorporates the mass shell condition for string states, so if the mass shell condition is modified by quantum mass renormalization, this will show up as an anomaly in the classical BRST symmetry. Likewise, BRST symmetry incorporates the equations of motion for background fields. A non-zero massless tadpole means that the equations of motion have not been satisfied, so BRST symmetry will fail.

In figs. 16 and 17, we have sketched the separating degenerations that can lead to BRST anomalies. In each case, we consider a product of n conformal

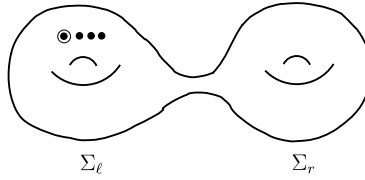


Figure 17: A separating degeneration related to tadpoles and the associated anomalies. All vertex operators, including the BRST-trivial vertex operator labeled by a circle, are on the same branch.

or superconformal vertex operators $\mathcal{V}_1, \dots, \mathcal{V}_n$ on a string worldsheet Σ , of genus g . These operators are all Q_B -invariant, have canonical ghost number and (in the superstring case) picture number, and are annihilated by b_n, β_m , $n, m \geq 0$. We single out one of these operators, say \mathcal{V}_1 , to be BRST-trivial, obeying $\mathcal{V}_1 = \{Q_B, \mathcal{W}_1\}$, where \mathcal{W}_1 obeys the same conditions as the \mathcal{V}_i except that its ghost number is 1 less (and, of course, \mathcal{W}_1 is not Q_B -invariant). Σ degenerates to a union of two components Σ_ℓ and Σ_r , where we can assume that \mathcal{V}_1 is inserted on, say, Σ_ℓ . There are two interesting cases related to mass renormalization – the operators $\mathcal{V}_2, \dots, \mathcal{V}_n$ may be all but one on Σ_ℓ or all on Σ_r , as in figs. 16(a) and (b), respectively. There is essentially only one configuration associated to massless tadpoles, with all vertex operators on Σ_ℓ , as in fig. 17.

To properly understand any of these three cases, we need a sharper result than the one that we obtained in section 7.1. Let us first recall the description in section 6.5 of the residue of the pole at $L_0 = 0$. In that analysis, we showed that the residue of the pole can be computed by inserting on each side conformal vertex operators associated to physical states. For example, for bosonic open strings, according to eqn. (6.84), the residue is computed by inserting $\sum_i cU_i \otimes cU^i$, where the sum runs over a complete set of matter primaries of $L_0 = 1$ (modulo null vectors), and the two factors are to be inserted on Σ_ℓ and Σ_r , respectively. An essential point here is that the residue of the pole can be computed by inserting only the vertex operators associated to physical states, not to more general states of $L_0 = 0$. (In fact, the conformally-invariant formalism assumed in this paper does not enable us to define insertions of such more general vertex operators.)

By contrast, in section 7.1, we associated the anomaly to states of $L_0 = 0$ propagating between Σ_ℓ and Σ_r , but we did not restrict these to physical states. To understand the BRST anomalies associated to the processes

sketched in the above figures, we need a sharper result, more similar to that of section 6.5, relating the anomaly to couplings of physical states.

There is an obvious and almost correct guess for how to compute the anomaly. One step in the derivation in section 7.1 was to omit the usual factor of b_0 that appears in the residue of the $1/L_0$ pole. According to (6.89), this means that we need to replace $\sum_i cU_i \otimes cU^i$ by

$$(7.4) \quad \sum_i (c\partial cU_i \otimes cU^i + cU_i \otimes c\partial cU^i).$$

So on either Σ_ℓ or Σ_r we must insert an operator $c\partial cU_i$ rather than cU_i , with ghost number greater by 1 than the canonical value. On the other hand, when we calculate the anomaly by evaluating the right hand side of eqn. (7.1), one of the operators inserted on Σ_ℓ is the operator \mathcal{W}_1 , whose ghost number is less by 1 than the canonical value. So to get a non-zero path integral, we should insert $c\partial cU_i$ on Σ_ℓ and the conformal vertex operator cU^i on Σ_r .

Both the derivation and the conclusion in the last paragraph are oversimplified. This must be so for the following simple reason: our formalism does not allow an insertion of the operator $c\partial cU_i$, since it is not annihilated by b_0 . A more precise explanation is given in section 7.7. However, it is true that, as the heuristic argument in the last paragraph suggests, the anomaly can be evaluated as a sum of contributions in each of which a conformal or superconformal vertex operator is inserted on Σ_r . Thus, the anomaly can be evaluated as a path integral on Σ_r with an insertion of

$$(7.5) \quad \mathcal{O} = \sum_i a_i \mathcal{Y}^i,$$

where the sum runs over all conformal or superconformal vertex operators \mathcal{Y}^i of the appropriate momentum and the coefficients a_i are computed by path integrals on Σ_ℓ . For open bosonic strings, the \mathcal{Y}^i are cU^i , where U^i are matter primaries of $L_0 = 1$. The analogs for the other string theories are familiar.

We can now get a clearer understanding of the various BRST anomalies. Let us begin with the tadpole-like diagram of fig. 17. The genus g_r of Σ_r is strictly positive (or the indicated degeneration would not arise in the Deligne-Mumford compactification).⁶¹ For such a degeneration, the anomaly can be

⁶¹ With the right definition of the genus, this statement holds in all string theories, not just closed oriented string theory. For any possibly open or unorientable string worldsheet Σ of Euler characteristic χ , we define the genus g by $\chi = 2 - 2g$. For example, by this definition, a disc has $g = 1/2$. The general topological constraint for fig. 17 to be relevant in string perturbation theory is then indeed simply $g_r > 0$.

computed by inserting $\mathcal{O} = \sum_i a_i \mathcal{Y}^i$ on Σ_r , where the \mathcal{Y}^i are conformal or superconformal primaries of spacetime momentum $P = 0$. Thus the anomaly is $\sum_i a_i \langle \mathcal{Y}^i \rangle_{g_r}$, where $\langle \mathcal{Y}^i \rangle_{g_r}$ is the tadpole of \mathcal{Y}^i in genus g_r . In particular, if the massless tadpoles vanish, then the anomalies associated to the tadpole diagrams also vanish. The reader would probably guess this result based on sections 7.1 and 7.2.

The basic fact about fig. 17 is really the tadpole (the perturbative vacuum does not satisfy the loop-corrected equations of motion) rather than the BRST anomaly (which reflects the fact that in string theory the equations of motion are built into the BRST symmetry). For oriented closed strings, there is not really a BRST anomaly independent of the equations of motion. Matters are different for open and/or unoriented strings, as we analyze in section 9.4. In that case, anomalies have a life of their own, independent of the equations of motion.

The next most obvious case is perhaps fig. 16(a) in which there is just one vertex operator, say \mathcal{V}_n , supported on Σ_r . The anomaly is

$$(7.6) \quad \sum_i a_i \langle \mathcal{Y}^i \mathcal{V}_n \rangle_{g_r}.$$

By momentum conservation, we can assume that the spacetime momentum carried by \mathcal{Y}^i is minus that of \mathcal{V}_n . The expression $\langle \mathcal{Y}^i \mathcal{V}_n \rangle_{g_r}$ represents a genus g_r coupling of two physical states; it is a contribution to mass renormalization for the physical states in question. Thus if the state corresponding to \mathcal{V}_n undergoes no mass renormalization in genus g_r , then there is no anomaly of this type. This again is probably the result that the reader would guess.

We are left with the anomaly of fig. 16(b). This is best-understood as a loop correction to classical BRST symmetry. At tree-level, the state $\{Q_B, \mathcal{W}_1\}$ decouples from the S -matrix. The genus g_ℓ of Σ_ℓ is strictly positive,⁶² or the configuration of fig. 16(b) does not arise in the Deligne-Mumford compactification. The anomaly of fig. 16(b) means that in genus $g \geq g_\ell$, the coupling of $\{Q_B, \mathcal{W}_1\}$ is not zero but equals the coupling of $\mathcal{O} = \sum_i a_i \mathcal{Y}^i$ in genus g_r . In other words, what decouples is not $\{Q_B, \mathcal{W}_1\}$ but

$$(7.7) \quad \mathcal{Q}_B(\mathcal{W}_1) = \{Q_B, \mathcal{W}_1\} - g_{st}^{2g_\ell} \mathcal{O}(\mathcal{W}_1),$$

where we write $\mathcal{O}(\mathcal{W}_1)$ (and not just \mathcal{O}) to emphasize that \mathcal{O} is a linear function of \mathcal{W}_1 . Since $\mathcal{O}(\mathcal{W}_1)$ is a physical vertex operator, obeying $\{Q_B, \mathcal{O}(\mathcal{W}_1)\} = 0$, the deformation preserves the fact that $\mathcal{Q}_B^2 = 0$, to this

⁶²See footnote 61 for the most general version of this statement.

order. Thus, BRST symmetry is retained in this order, but deformed. We discuss the physical interpretation of this phenomenon in section 7.4.

Going back to fig. 16(a), one might be slightly puzzled about why mass renormalization of the string state associated to \mathcal{V}_n causes an anomaly in the BRST symmetry $\mathcal{V}_1 \rightarrow \mathcal{V}_1 + \{Q_B, \mathcal{W}_1\}$ associated to particle 1. An intuitive reason for this has already been explained. To show vanishing of $\langle \{Q_B, \mathcal{W}_1\} \mathcal{V}_2 \dots \mathcal{V}_n \rangle$, one needs BRST invariance of $\mathcal{V}_2, \dots, \mathcal{V}_n$. But since classical BRST symmetry incorporates the mass shell conditions, mass renormalization of the string states associated to $\mathcal{V}_2, \dots, \mathcal{V}_n$ implies a correction to the condition of BRST invariance for those vertex operators, leading to a failure of BRST-invariance for \mathcal{V}_1 .

We can write a formula for the correction to BRST symmetry of physical states by representing the anomaly by the operator insertion in (7.4). Given this and defining $y^i = \langle cU^i \mathcal{V}_n \rangle_{\Sigma_r}$, the anomaly amounts to replacing $\{Q_B, \mathcal{V}_n\}$ by $\sum_i y^i c \partial c U_i$. The corrected BRST symmetry is

$$(7.8) \quad \mathcal{Q}_B(\mathcal{V}_n) = \{Q_B, \mathcal{V}_n\} + \sum_i y^i c \partial c U_i,$$

a formula described in [67]. Eqns. (7.7) and (7.8) give the leading corrections to the BRST symmetry for two adjacent values of the ghost number, corresponding to gauge parameters and physical states, respectively. A systematic framework to study such deformations to all orders has been developed [22], though this has not been expressed in the super Riemann surface language.

7.4. Gauge symmetry breaking

Our next task is to explain the physical meaning of the correction to BRST symmetry described in eqn. (7.7). At least for gauge symmetries of massless string states, this correction represents spontaneous breaking of gauge symmetry. (For massive string states, the gauge symmetries are already spontaneously broken at tree level, so the corrections to the BRST transformations do not seem to have an interpretation in terms of spontaneous symmetry breakdown.)

For example, \mathcal{V}_1 might be the vertex operator of a gauge field A that is massless at tree level. In bosonic open string theory, we might have $\mathcal{V}_1 = c \varepsilon \cdot \partial X \exp(ip \cdot X)$, where p and ε are momentum and polarization vectors, obeying $p^2 = \varepsilon \cdot p = 0$. In this example, the gauge parameter is $\mathcal{W}_1 = \exp(ip \cdot X)$; it generates the gauge transformation $\varepsilon \rightarrow \varepsilon - i\lambda p$ (where λ is a constant). We could consider in a similar way any other string theory

that has a massless gauge field at tree level, coming from an open- or closed-string state; the specific formulas for \mathcal{V}_1 and \mathcal{W}_1 will not be important for now. The operator \mathcal{W}_1 is not the vertex operator of a physical state, since it is not Q_B -invariant, but $\mathcal{O}(\mathcal{W}_1)$, if not zero, is the vertex operator of a physical state with the same Poincaré quantum numbers as \mathcal{W}_1 . So it carries spacetime momentum p and is invariant under the little group (the subgroup of Lorentz transformations that keep p fixed). In other words, $\mathcal{O}(\mathcal{W}_1)$ is the vertex operator of a scalar field σ that is massless at tree level. That is the key point. The formula $\mathcal{Q}_B(\mathcal{W}_1) = Q_B(\mathcal{W}_1) - g_{\text{st}}^{2g_\ell} \mathcal{O}(\mathcal{W}_1)$ tells us that a gauge transformation $\varepsilon \rightarrow \varepsilon - i\lambda p$ generated by \mathcal{W}_1 must be accompanied by an insertion of the $\mathcal{O}(\mathcal{W}_1)$ vertex operator with a coefficient $-g_{\text{st}}^{2g_\ell} \lambda$ or in other words by the shift

$$(7.9) \quad \sigma \rightarrow \sigma - g_{\text{st}}^{2g_\ell} \lambda.$$

This leads to what is usually called gauge symmetry breaking. The essential phenomenon (but without the factor of $g_{\text{st}}^{2g_\ell}$) is exhibited by the basic Stueckelberg model of $U(1)$ gauge symmetry breaking,

$$(7.10) \quad \int \left(F_{IJ} F^{IJ} + (\partial_I \sigma + A_I)^2 \right),$$

which is invariant under $A \rightarrow A + \partial\lambda$, $\sigma \rightarrow \sigma - \lambda$. In vacuum, $\sigma = 0$ up to a gauge transformation; in expanding around this vacuum, A is a massive spin 1 particle.

The power of g_{st} given in eqn. (7.9) is correct if A and σ are both closed-string states or both open-string states; it requires some modification otherwise, as we discuss presently. This power of g_{st} , which is not a constant but equals e^ϕ with ϕ the dilaton field, is unnatural in the context of gauge symmetry breaking and this gives one reason to suspect that the phenomenon just described will be hard to realize in string theory. This is true and will be explained below.

For another example, \mathcal{V}_1 might be the vertex operator for the massless spin 3/2 gravitino field ψ_I in a supersymmetric string compactification. In that case, \mathcal{W}_1 would have the Poincaré quantum numbers of a spin 1/2 field obeying the massless Dirac equation, and $\mathcal{O}(\mathcal{W}_1)$ is the vertex operator for a physical fermion field λ that is massless at tree level. (The fermion fields corresponding to \mathcal{V}_1 and to $\mathcal{O}(\mathcal{W}_1)$ have the same spacetime chirality.) The anomaly means that a gravitino gauge transformation, which at the linearized level is $\psi_I \rightarrow \psi_I + \partial_I \zeta$, where ζ is a c -number spinor field with the same

chirality as ψ , must be accompanied by

$$(7.11) \quad \lambda \rightarrow \lambda - g_{\text{st}}^{2g_\ell} \zeta.$$

λ is usually called the Goldstino, and what we have just described is spontaneous breaking of supersymmetry.

For a further example, one might be tempted to consider p -form gauge fields arising in the Ramond-Ramond sector of superstrings. However, these are not a useful example because in perturbative string theory there are no Ramond sector gauge parameters⁶³ at the massless level; the first Ramond sector gauge parameter arises at the first massive level (see eqn. C.9). Instead, the last example relevant to our discussion arises for the B -field of closed oriented strings. If \mathcal{V}_1 is the B -field vertex operator, then the corresponding gauge parameter

$$(7.12) \quad \mathcal{W}_1 = \varepsilon_I \cdot (c^z \partial_z X^I - \tilde{c}^{\bar{z}} \partial_{\bar{z}} X^I) \exp(ip \cdot X), \quad p^2 = \varepsilon \cdot p = 0,$$

has the Poincaré quantum numbers of a $U(1)$ gauge field, and $\mathcal{O}(\mathcal{W}_1)$, if not zero, actually is the vertex operator for a $U(1)$ gauge field A that is massless at tree level. The anomaly means that a B -field gauge transformation $B_{IJ} \rightarrow B_{IJ} + \partial_I \lambda_J - \partial_J \lambda_I$ must be accompanied by $A_I \rightarrow A_I - g_{\text{st}}^{2g_\ell} \lambda_I$. This results in the breakdown of B -field gauge-invariance; A can be gauged away and B becomes massive [72].

Based on field theory intuition, one might surmise – correctly – that the phenomena we have described here are actually rather rare among string theory compactifications. We have already explained in section 7.2.1 that spontaneous supersymmetry breaking by loops – in models in which supersymmetry is unbroken at tree level – can occur only under rather special circumstances. The only known models are those described in [44], involving the generation of a Fayet-Iliopoulos D-term at one-loop order. For ordinary bosonic gauge symmetries, the difficulty in triggering symmetry breaking by loops is even more obvious. If all charged scalar fields have positive masses at tree level, then perturbative quantum corrections cannot trigger gauge symmetry breaking; if there is a charged scalar field with a negative mass squared at tree level, then the gauge symmetry was spontaneously broken already classically. For a gauge symmetry unbroken at tree level to be spontaneously broken by weak

⁶³This is related to the fact that the vertex operator of a Ramond-Ramond field, in the canonical picture which we use to compute the S -matrix, describes the $p+1$ -form field strength rather than the p -form gauge field. See section 6.2.2 for one explanation of that fact.

quantum corrections, there must be a charged scalar field that is massless at tree level and acquires a negative mass squared from perturbative quantum corrections.

This is obviously rather special, but it can be natural if supersymmetry accounted for existence of massless charged scalars at tree level and is spontaneously broken by loop effects, as in the models of [44]. But even then, this mechanism for gauge symmetry breaking as a result of loop effects does not correspond to the BRST anomaly described in eqn. (7.7). That anomaly is a loop correction to the gauge transformation laws of the fields, not to the masses.

Field theory offers only less guidance about how spontaneous breakdown of B -field gauge symmetry might be triggered by small quantum corrections.

7.4.1. Examples Given all this, one might despair of finding an example in which the perturbative BRST anomaly of eqn. (7.7) occurs for bosonic gauge symmetries of massless modes of string theory. In fact, this happens only in a very restricted way. To understand where to look, it helps to start with something that was explained in sections 2.5.3 and 4.3.1 for closed bosonic strings and the NS sector of closed superstrings. (The phenomenon under discussion does not occur for the R-R sector, as explained above.)

The key point is that in those cases, the gauge invariances of massless fields can be proved by integration by parts on the worldsheet Σ , as opposed to the moduli space of Riemann surfaces. We consider a closed-string vertex operator \mathcal{V} whose integrated form is V . V is a $(1, 1)$ -form constructed from the matter fields only, and couples via the integral $\int_{\Sigma} V$. A gauge transformation acts by $V \rightarrow V + dW$, where W is a 1-form constructed from the matter fields. For example, if V is the vertex operator of the B -field of the closed oriented bosonic string, then

$$(7.13) \quad W = \varepsilon_I dX^I \exp(ip \cdot X), \quad p^2 = \varepsilon \cdot p = 0.$$

The gauge symmetry is simply the vanishing of $\int_{\Sigma} dW$. In closed-string theory, there is no problem with the vanishing of that integral. The only singularities in the integral occur when W meets another vertex operator \mathcal{X} , in which case the worldsheet Σ splits off a genus zero component containing W and \mathcal{X} . This is not a degeneration of the form of fig. 16(b) that can lead to quantum corrections to gauge-invariance; the total momentum carried by W and \mathcal{X} is generically off-shell, so a collision between them does not lead to an anomaly.

The situation for open-string vertex operators is no different. If \mathcal{W} is an open-string vertex operator, its integrated form is a 1-form W , constructed

from the matter fields, that couples via $\oint_{\partial\Sigma} W$. A gauge transformation acts by $W \rightarrow W + dP$, where now P is a scalar or 0-form operator constructed from the matter fields. But dP decouples, since $\oint_{\partial\Sigma} dP = 0$. Again, there may be singularities where P meets another open-string vertex operator, but they do not affect the vanishing of the integral.

But something new does happen in the combined theory of open and closed strings, specifically for the case of a closed-string gauge parameter. Let us go back to the closed-string vertex operator \mathcal{V} with integrated form V and gauge-invariance $V \rightarrow V + dW$. Clearly, if Σ has a nonempty boundary, then $\int_{\Sigma} dW$ need not vanish. Rather, by Stokes's theorem, it equals $\int_{\partial\Sigma} W$. Here we can regard W as an open-string vertex operator, in integrated form. For example, in the case that \mathcal{V} was the vertex operator for the BRST-invariant B -field, the gauge parameter W was presented in (7.13), and we recognize it as the classic formula for the vertex operator of a massless open-string gauge field. When W is understood as an open-string vertex operator, rather than a closed-string gauge parameter, we will denote it as W_{open} . The corresponding unintegrated vertex operator is $\mathcal{W}_{\text{open}} = cW_{\text{open}}$.

Clearly, in this situation, the closed-string mode that is represented (in integrated form) by $V = dW$ does not decouple from the S -matrix; rather its coupling is that of the open-string mode with integrated vertex operator W_{open} , as originally explained long ago [72]. (The original construction was for breakdown of B -field gauge invariance; to get a similar model with breaking of a $U(1)$ gauge symmetry, it suffices to compactify the target space on a circle, whereupon part of the B -field becomes a $U(1)$ gauge field which undergoes the same mechanism.) In our terminology, the fact that the closed-string pure gauge mode $\mathcal{V} = \{Q, \mathcal{W}\}$ (or in integrated form $V = dW$) does not decouple, but rather couples via the open-string vertex operator W_{open} , is expressed in the form

$$(7.14) \quad \mathcal{O}(\mathcal{W}) = \mathcal{W}_{\text{open}}, \quad \mathcal{W}_{\text{open}} = cW_{\text{open}}.$$

Indeed, this type of example perfectly illustrates eqn. (7.7). The boundary term that prevents decoupling of dW occurs when W approaches the boundary of Σ . From the point of view of the Deligne-Mumford compactification, this means that Σ splits off an additional component Σ_{ℓ} which is a disc that contains \mathcal{W} , but contains none of the other vertex operators that may originally have been present. (See section 9.2.1 for more on such open-string degenerations.) In other words, this example fits the framework of fig. 16(b), with Σ_{ℓ} being a disc, $\mathcal{V}_1 = \{Q, \mathcal{W}_1\}$ being a closed-string vertex operator, $\mathcal{O}(\mathcal{W}_1)$ an open-string vertex operator, and no restriction on Σ_r or $\mathcal{V}_2, \dots, \mathcal{V}_n$.

Though it perfectly fits the framework of fig. 16(b), the example challenges the terminology that we used in describing that figure. A worldsheet with disc topology generates the classical action of open strings, but is a correction of some sort – it is not clear that one should call it a quantum correction – for the closed strings. With the most natural way to normalize the vertex operators, there is no factor of g_{st} in the correction to BRST symmetry in this example; instead the transformation $\mathcal{W} \rightarrow \mathcal{O}(\mathcal{W})$ maps closed-string gauge parameters to open string vertex operators (whose natural scaling with g_{st} is different). For example, for the breakdown of B -field gauge symmetry, the action is naturally written

$$(7.15) \quad \int d^D x \left(\frac{1}{g_{\text{st}}^2} H_{IJK}^2 + \frac{1}{g_{\text{st}}} (\partial_I A_J - \partial_J A_I + B_{IJ})^2 \right), \quad H = dB.$$

The gauge symmetry is $\delta B_{IJ} = \partial_I \lambda_J - \partial_J \lambda_I$, $A_I \rightarrow A_I - \lambda_I$, with no factor of g_{st} .

Symmetry breaking by mixing of open and closed strings can also occur for spacetime supersymmetry; this is discussed in section 9.2.3. But, even though there are severe constraints, there are additional possibilities for spontaneous breaking of spacetime supersymmetry by loop effects that do not have analogs for gauge symmetries associated to massless bosonic gauge fields. That is because (section 5.4) there is no forgetful map for Ramond punctures. Decoupling of a pure gauge mode of the gravitino field can only be proved by a full integral over the moduli space of super Riemann surfaces, not simply by an integral over Σ . As a result, some of the arguments that apply for bosonic gauge symmetries cannot be used to constrain supersymmetry breaking by loops. In particular, in the models described and studied in [44, 39, 42], the phenomenon of fig. 16(b) does occur, with Σ_ℓ having genus 1. From the point of view of low energy effective field theory, what happens is the following. At tree level, the theory has a massless neutral spin 1/2 field λ with a supersymmetry transformation law $\delta\lambda = D + \dots$, where D is the auxiliary field in the vector multiplet associated to a $U(1)$ gauge symmetry. At tree level, D vanishes in a vacuum with unbroken $U(1)$ gauge symmetry. At one-loop order, the value of D in such a vacuum is shifted by a constant ζ (called the Fayet-Iliopoulos term) and so the transformation law becomes $\delta\lambda = \zeta + \dots$, where the \dots terms are bilinear and higher order in charged fields. The constant term means that, at least if we keep the $U(1)$ unbroken, supersymmetry is spontaneously broken. (For a framework, not expressed in terms of super Riemann surfaces, in which one can study the restoration of supersymmetry with the $U(1)$ symmetry spontaneously broken, see [23].)

From the vantage point of the present paper, the main importance of the phenomenon depicted in fig. 16(b) is that it is a failure mode for superstring perturbation theory. We have to know that loop corrections do not trigger the breaking of spacetime supersymmetry in order to have any hope of using supersymmetry to prove the vanishing of massless tadpoles. Happily, there is no difficulty, as supersymmetry breaking by loops is highly constrained and it is possible to effectively determine when it occurs.

7.4.2. Some miscellaneous remarks We conclude this discussion with some miscellaneous remarks.

First of all, spontaneous gauge symmetry breaking always leads to mass renormalization, since the gauge field gains a non-zero mass. When this happens, as in any situation with mass renormalization, the S -matrix diverges if one attempts to calculate on the classical mass shell.

Although spontaneous gauge symmetry breaking always leads to mass renormalization for a massless gauge field, the converse is not true. (In contrast to gauge symmetry breaking, the phenomenon that we are about to describe is limited to abelian gauge fields.) For example, consider a closed-string compactification to four dimensions with a $U(1)$ gauge field A and also a two-form gauge field B . At tree level, the action is schematically

$$(7.16) \quad \int_{\mathbb{R}^4} d^4x e^{-2\phi} \left((dA)^2 + (dB)^2 \right).$$

At one-loop order, one may generate a gauge-invariant interaction

$$(7.17) \quad \int_{\mathbb{R}^4} B \wedge dA.$$

This interaction can only be generated at one-loop order, since the presence of a non-zero power of $g_{\text{st}} = e^\phi$ would spoil the invariance under the B -field gauge transformations $B \rightarrow B + d\Lambda$; the proof of this invariance requires integration by parts. If generated, the interaction (7.17) leads to mass renormalization for A and B , without modifying the classical gauge-invariance. The fact that this interaction can only be generated at one-loop order means that it is possible, in a given string theory compactification, to effectively determine whether it is generated or not. The main known case in which such a one-loop interaction is generated is the class of models studied in [44, 39, 42]. The mechanism for generating this interaction in these models is as follows: in certain heterotic string compactifications to four dimensions, the ten-dimensional Green-Schwarz interaction $B \wedge \text{Tr } F^4 + \dots$ (which arises

in heterotic string theory at one-loop order) generates the interaction (7.17). When this interaction is generated at one-loop order, the one-loop S -matrix elements of A or B are divergent, because of the mass renormalization or more precisely the on-shell pole that results from the A - B mixing.⁶⁴

From a field theory point of view, in such a four-dimensional model, one can dualize B to a scalar field σ . In the dual description, A gets a mass from gauge symmetry breaking. Outside of perturbation theory, there is no distinction between massless spin-zero particles that are associated to spin-zero fields and those that are associated to two-form fields; nonperturbative dualities can exchange different descriptions of the same massless modes. In string perturbation theory, however, there are two kinds of massless spin-zero mode in four dimensions – those whose vertex operator is derived from a zero-form and those whose vertex operator is derived from a two-form. Therefore, in perturbation theory, the two mechanisms for an abelian gauge field to gain a mass, by mixing with a spin-zero mode of one of the two types, are distinct. Both mechanisms are severely restricted, in ways that we have explained.

7.4.3. Restricting to the massless S -matrix Although perturbative mass renormalization is heavily constrained for massless gauge fields, it is almost ubiquitous for massive string states. The only obvious exception is that in some models, massive BPS states are not subject to mass renormalization.

This leads us to a problem that will place a significant restriction on what we can accomplish in the rest of this paper. As soon as there is mass renormalization, the usual conformal or superconformal framework of superstring perturbation theory does not work. It is necessary to go slightly off-shell in order to proceed. This can be done by endowing all Riemann surfaces or super Riemann surfaces with local parameters at punctures (in the super Riemann surface case, the analog of a local parameter is a local system of superconformal coordinates). This approach goes back to [17, 18] and has been much developed recently for superstrings (in the language of picture-changing operators, not super Riemann surfaces); for example see [19–24].

However, in the present paper, to avoid an extra layer of complication, we prefer to avoid these issues. Therefore, we will limit ourselves in the rest of this paper to studying the S -matrix of massless particles. Moreover we will

⁶⁴This example is rather tricky. Even when the interaction (7.17) is generated at one-loop level, the on-shell one-loop two-point function $\langle AB \rangle$ vanishes. (This follows just from conservation of angular momentum.) Accordingly there is no one-loop BRST anomaly. However, as stated in the text, the mixing and the resulting pole do cause a divergence in one-loop S -matrix elements of A or B .

consider only supersymmetric theories and more specifically only theories in which spacetime supersymmetry and general considerations of low energy field theory suffice to show that in perturbation theory there is no spontaneous supersymmetry breaking and no mass renormalization for any of the particles that are massless at tree level. (This includes most supersymmetric theories, as we have explained.)

The restriction to the massless S -matrix is unfortunate. However, in one sense the restriction is less severe than it might appear. Massive particles that decay to stable ones are most precisely understood as resonances in the S -matrix of massless particles, so it is really only the existence of *stable* massive particles that obstructs our ability to compute the complete S -matrix in perturbation theory. For example, in four of the five ten-dimensional superstring theories, the massive string states are all unstable against decay to massless ones, so the massless S -matrix is the complete S -matrix in perturbation theory. The exception is the Spin(32)/ \mathbb{Z}_2 heterotic string, whose perturbative spectrum includes stable massive particles in the spinor representation of the gauge group.

7.5. Massless tadpoles in string theory

Ultimately, the most critical question for superstring perturbation theory is to make sure that the vacuum state that one attempts to construct in perturbation theory is not destabilized by massless tadpoles. To show this, we will use spacetime supersymmetry. As just explained, we consider only the massless S -matrix, and only string compactifications in which there is no perturbative mass renormalization for massless particles.

Additionally, to begin with, we consider only closed oriented strings. This leads to several simplifications. One immediate simplification is that for closed, oriented strings, we only have to consider tadpoles in the NS sector (or the NS-NS sector in the case of Type II superstrings). R-R tadpoles are prevented both by the $(-1)^{F_L}$ symmetry⁶⁵ of closed oriented Type II string theory, and by the decoupling of R-R fields at zero momentum. In open and/or unoriented superstring theory, there are some new ingredients (there is no $(-1)^{F_L}$ symmetry, and the decoupling of R-R fields at zero momentum can fail under certain circumstances), and R-R tadpoles become relevant. We discuss these matters in section 9.

⁶⁵The superconformal formalism only works when the expectation values of R-R fields vanishes. This ensures the existence in perturbation theory of a symmetry $(-1)^{F_L}$ that acts as +1 or -1 on states from the left-moving NS or R sector.

The proof that the amplitudes are spacetime supersymmetric and free of massless tadpoles will proceed by induction in the genus g of the string worldsheet. Assuming that there are no massless tadpoles and that spacetime supersymmetry holds up to genus $g - 1$, we will show that this is also true in genus g .

The arguments are deferred to section 8. However, one important detail about the interpretation of the result will be explained here.

Whenever one has a perturbative superstring S -matrix, it depends on at least one modulus, namely the string coupling constant $g_{\text{st}} = e^\phi$. There may possibly be other moduli. Let \mathcal{Z} be the moduli space that is parametrized by g_{st} and any other moduli. The output of our discussion will be to show that the perturbative S -matrix is well-defined as a function on \mathcal{Z} . But in general there is no reason to expect \mathcal{Z} to have a natural parametrization in terms of fields such as ϕ .

We will find that in general there is no such natural parametrization. This will be the conclusion of an analysis in section 7.6 of infrared divergences and their regularization. We will find that when massless tadpoles vanish, the integrals that have to be evaluated to compute the S -matrix are convergent, but only conditionally so. Different infrared regulators lead to results for the S -matrix that differ by reparametrizations of \mathcal{Z} . (Some of the issues have been treated in [73].)

What we have just explained has an analog for wavefunction renormalization (section 7.6.3). When mass renormalization vanishes, this causes the cancellation of certain logarithmically divergent contributions to the scattering amplitudes. One is left with integrals that are conditionally convergent. The result of regulating these conditionally convergent integrals is unique up to a coupling-dependent renormalization of the vertex operators that are used to compute the scattering amplitudes. So the S -matrix is naturally determined, but the vertex operators that are used to evaluate it are not.

Though this will not be explored in the present paper, one anticipates that in compactifications with much unbroken supersymmetry, \mathcal{Z} will often have a natural parametrization or at least a small class of natural parametrizations. This would correspond in our analysis in section 7.6 to using supersymmetry to find a distinguished infrared regulator.

7.6. What happens when the tadpole vanishes

7.6.1. Conditionally convergent integrals In describing how to proceed when tadpoles vanish, we will use the language of bosonic string theory

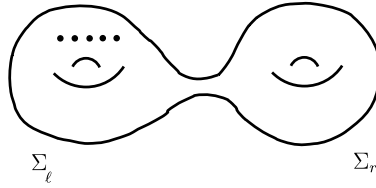


Figure 18: A Riemann surface Σ splits into two components Σ_ℓ and Σ_r . All external vertex operators are on Σ_ℓ .

since the subtleties of worldsheet and spacetime supersymmetry play no essential role. We ignore the tachyon of bosonic string theory as we are really interested in applying the reasoning that follows to tachyon-free superstring theories. (For a systematic approach to these issues, not expressed in the super Riemann surface language, see [22].)

We consider the separating degeneration of fig. 18 in which a Riemann or super Riemann surface Σ splits into two components Σ_ℓ and Σ_r ; the delicate case is that all external vertex operators are on one branch, say Σ_ℓ . In bosonic string theory, the splitting is described by an equation

$$(7.18) \quad xy = q,$$

where x and y are local parameters on Σ_ℓ and Σ_r respectively, and q is a complex parameter that controls the degeneration. In superstring theory, q is replaced by an analogous parameter q_{NS} or q_{R} , defined in section 6.4.5, and the analysis then proceeds in the same way.

In general, Σ is a Riemann surface of genus g with n punctures, and Σ_ℓ and Σ_r are of genus g_ℓ and g_r with $g_\ell + g_r = g$. The complex structures of Σ is parametrized by $\mathcal{M}_{g,n}$, the moduli space of Riemann surfaces of genus g with n punctures. We write $\widehat{\mathcal{M}}$ for the Deligne-Mumford compactification of $\mathcal{M}_{g,n}$ that arises if one allows degenerations such as the one at $q = 0$.

If we set $q = 0$ in (7.18), then Σ decomposes as the union of Σ_ℓ and Σ_r , joined together at the single point $x = y = 0$. This defines a divisor $\mathcal{D} \subset \widehat{\mathcal{M}}$, and this divisor has a very simple structure, as already explained in eqn. (6.66):

$$(7.19) \quad \mathcal{D} = \widehat{\mathcal{M}}_\ell \times \widehat{\mathcal{M}}_r.$$

Here $\widehat{\mathcal{M}}_\ell$ and $\widehat{\mathcal{M}}_r$ are the Deligne-Mumford compactifications of \mathcal{M}_ℓ and \mathcal{M}_r , which parametrize complex structures on Σ_ℓ and Σ_r , respectively. This just

says that at $q = 0$, the moduli of Σ_ℓ and Σ_r (including the choices of the points at which they are glued together to make Σ) can be varied independently, and a choice of those moduli determines Σ .

If all external vertex operators are on Σ_ℓ , then $\widehat{\mathcal{M}}_\ell$ is a copy of $\widehat{\mathcal{M}}_{g_\ell, n+1}$, which parametrizes a genus g_ℓ surface with $n + 1$ punctures. The $n + 1^{\text{th}}$ puncture is the point at which Σ_ℓ is glued to Σ_r . Similarly in this situation, $\widehat{\mathcal{M}}_r$ is a copy of $\widehat{\mathcal{M}}_{g_r, 1}$.

Near \mathcal{D} , $\widehat{\mathcal{M}}$ is parametrized by q as well as \mathcal{D} and can be approximated as a small neighborhood of \mathcal{D} embedded as the zero section of a complex line bundle $\mathcal{N} \rightarrow \mathcal{D}$. \mathcal{N} is the normal bundle to \mathcal{D} in $\widehat{\mathcal{M}}$. The parameter q in (7.18) is not well-defined as a complex number, because its definition depends on the choices of local parameters x and y . Rather, q is a linear function on the line bundle \mathcal{N} (or equivalently a section of \mathcal{N}^{-1}). The transformation of q under a change in the local parameter was described explicitly in eqn. (6.111). The only aspect of that formula that we will use is that q is well-defined up to a transformation

$$(7.20) \quad q \rightarrow e^{f_\ell + f_r} q.$$

Here f_ℓ is a function on $\widehat{\mathcal{M}}_\ell$ and reflects the dependence of q on the choice of x ; similarly, f_r is a function on $\widehat{\mathcal{M}}_r$ and reflects the dependence of q on the choice of y . The factorized form of (7.20) reflects the fact that \mathcal{N} is the tensor product of a line bundle over $\widehat{\mathcal{M}}_\ell$ and one over $\widehat{\mathcal{M}}_r$:

$$(7.21) \quad \mathcal{N} = \mathcal{L}_\ell \otimes \mathcal{L}_r,$$

as described in more detail in section 6.1.3 of [16].

The g -loop contribution to a scattering amplitude is obtained by integrating over $\widehat{\mathcal{M}}$ a differential form $F(g|\delta g)$, which of course depends on the momenta p_i and the other quantum numbers ζ_i of the external string states:

$$(7.22) \quad \mathcal{A}_g(p_1, \zeta_1; \dots; p_n, \zeta_n) = \int_{\widehat{\mathcal{M}}} F_{p_1, \zeta_1; \dots; p_n, \zeta_n}(g|\delta g)$$

We usually write just $F(g|\delta g)$ or simply F rather than $F_{p_1, \zeta_1; \dots; p_n, \zeta_n}(g|\delta g)$. The contribution to F of a massless scalar of zero momentum that propagates between Σ_ℓ and Σ_r was analyzed in section 6.2. It factors as the product of $d^2q/\bar{q}q$ multiplied by a form \mathcal{G}_ℓ on $\widehat{\mathcal{M}}_\ell$ and a form \mathcal{G}_r on $\widehat{\mathcal{M}}_r$. (We use here the language of bosonic string theory and do not distinguish \tilde{q} from \bar{q} ; d^2q is short for $-id\bar{q}dq$.) The forms \mathcal{G}_ℓ and \mathcal{G}_r describe the coupling of a single massless scalar of zero momentum to the surfaces Σ_ℓ and Σ_r , respectively, in

addition to the n external strings that are already coupled to Σ_ℓ . The singular part of the integral at $q = 0$ is thus

$$(7.23) \quad \mathcal{A}_{g,\text{sing}} = \sum_{\alpha=1}^s \int \frac{d^2q}{\bar{q}q} \int_{\widehat{\mathcal{M}}_\ell} \mathcal{G}_{\ell,\alpha} \int_{\widehat{\mathcal{M}}_r} \mathcal{G}_{r,\alpha}.$$

The sum runs over all massless scalars ϕ_α , $\alpha = 1, \dots, s$, that might contribute tadpoles. The factor $1/\bar{q}q$ is $\tilde{q}^{\tilde{L}_0-1} q^{L_0-1}$ with $\tilde{L}_0 = L_0 = 0$.

The integral (7.23) diverges at $q = 0$. On the other hand, our hypothesis that the g -loop tadpoles vanish means that

$$(7.24) \quad \int_{\widehat{\mathcal{M}}_r} \mathcal{G}_{r,\alpha} = 0, \quad \alpha = 1, \dots, s.$$

So the divergence at $q = 0$ is multiplied by 0, rather like the $0/0$ that one gets in field theory if the tadpole vanishes.

Suppose that we cut off the integral by placing a lower bound $|q| \geq \epsilon$ for some small $\epsilon > 0$. The integral⁶⁶

$$(7.25) \quad \int_{\epsilon < |q| < \Lambda} \frac{d^2q}{\bar{q}q}$$

diverges for $\epsilon \rightarrow 0$ as $-4\pi \log \epsilon$, but the coefficient of $\log \epsilon$ vanishes after integration over $\widehat{\mathcal{M}}_\ell \times \widehat{\mathcal{M}}_r$, because of (7.24). So the cutoff integral

$$(7.26) \quad \mathcal{A}_{g,\epsilon} = \int_{\widehat{\mathcal{M}}_\epsilon} F(g|\delta g),$$

where $\widehat{\mathcal{M}}_\epsilon$ is the region in $\widehat{\mathcal{M}}$ defined by $|q| \geq \epsilon$, has a limit as $\epsilon \rightarrow 0$. Vanishing of the tadpoles has made the scattering amplitude convergent.

This is not the whole story, however, since q is really not a complex number, as assumed in the above derivation, but a section of the line bundle \mathcal{N}^{-1} . This bundle is topologically nontrivial, so there is no way to trivialize it (even if we do not ask to do so holomorphically). However, what we need to make sense of the cutoff that we used is not a trivalization of \mathcal{N} but merely a hermitian metric on \mathcal{N} , or equivalently on \mathcal{N}^{-1} . The condition $|q| > \epsilon$ means that the norm of q , computed using some chosen hermitian metric on \mathcal{N}^{-1} , is greater than ϵ . A change in the hermitian metric on \mathcal{N}^{-1} would be equivalent to replacing the cutoff condition $|q| > \epsilon$ by $|q| > e^h \epsilon$, for some real-valued

⁶⁶ Λ is an irrelevant cutoff at large $|q|$. The approximation (7.23) to the integral over $\widehat{\mathcal{M}}$ is only valid for q small.

function h on $\widehat{\mathcal{M}}_\ell \times \widehat{\mathcal{M}}_r$. This has the effect of shifting $\log \epsilon$ to $\log \epsilon + h$. This does not affect the fact that \mathcal{A}_g has a limit for $\epsilon \rightarrow 0$, but it shifts the value of the limit by

$$(7.27) \quad \mathcal{A}_g \rightarrow \mathcal{A}_g - 4\pi \sum_\alpha \int_{\widehat{\mathcal{M}}_\ell \times \widehat{\mathcal{M}}_r} h \mathcal{G}_{\ell,\alpha} \mathcal{G}_{r,\alpha}.$$

This is certainly not zero in general, so we seem to be in trouble.

What saves the day is the following. Since \mathcal{N}^{-1} is the tensor product of a line bundle on $\mathcal{L}_\ell^{-1} \rightarrow \widehat{\mathcal{M}}_\ell$ and a line bundle $\mathcal{L}_r^{-1} \rightarrow \widehat{\mathcal{M}}_r$, it is natural to choose the metric on \mathcal{N}^{-1} to be the tensor product of a metric on \mathcal{L}_ℓ^{-1} and a metric on \mathcal{L}_r^{-1} . This means that we can naturally restrict to the case that $h = h_\ell + h_r$, where h_ℓ and h_r are functions on $\widehat{\mathcal{M}}_\ell$ and $\widehat{\mathcal{M}}_r$, respectively. Using the vanishing tadpole condition (7.24), we see that h_ℓ does not contribute to the integral in (7.27). Defining

$$(7.28) \quad \Delta_{g_r} \phi_\alpha = -\frac{1}{4\pi} \int_{\widehat{\mathcal{M}}_r} h_r \mathcal{G}_{r,\alpha},$$

the metric dependence of \mathcal{A}_g is

$$(7.29) \quad \mathcal{A}_g \rightarrow \mathcal{A}_g + \sum_\alpha \Delta_{g_r} \phi_\alpha \int_{\widehat{\mathcal{M}}_\ell} \mathcal{G}_{\ell,\alpha}.$$

Here $\int_{\widehat{\mathcal{M}}_\ell} \mathcal{G}_{\ell,\alpha}$ is the genus g_ℓ contribution to a scattering amplitude with the \mathfrak{n} vertex operators that we started with plus one more vertex operator that represents a ϕ_α field at zero momentum. This insertion gives the derivative of the scattering amplitude with respect to ϕ_α . So we can interpret (7.29) as

$$(7.30) \quad \mathcal{A}_g \rightarrow \mathcal{A}_g + \sum_\alpha \Delta_{g_r} \phi_\alpha \frac{\partial}{\partial \phi_\alpha} \mathcal{A}_{g_\ell},$$

where \mathcal{A}_{g_ℓ} is the genus g_ℓ contribution to the scattering amplitude under study.

However, what we have analyzed here is a particular degeneration corresponding to the decomposition $g = g_\ell + g_r$. For the full story, we have to sum over all such degenerations, analyzing each one in the same way. This gives

$$(7.31) \quad \mathcal{A}_g \rightarrow \mathcal{A}_g + \sum_{g_\ell + g_r = g} \sum_\alpha \Delta_{g_r} \phi_\alpha \frac{\partial}{\partial \phi_\alpha} \mathcal{A}_{g_\ell}.$$

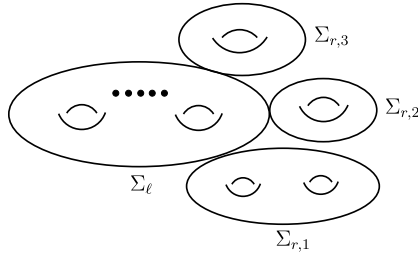


Figure 19: A Riemann surface that degenerates to a union of Σ_ℓ with several other branches $\Sigma_{r,i}$, $i = 1, 2, 3$.

The full perturbative scattering amplitude is $\mathcal{A} = \sum_{g=0}^{\infty} \mathcal{A}_g$. We include factors of the string coupling constant g_{st} in the definition of \mathcal{A}_g (so explicitly \mathcal{A}_g is proportional to g_{st}^{2g-2}), so this infinite sum is really an expansion in powers of g_{st} . In perturbation theory, one views this sum over genus as a formal power series in powers of g_{st} . The reason to not write explicitly the powers of g_{st} is that we do not want to treat the dilaton, whose expectation value determines g_{st} , differently from the rest of the ϕ_α . Similarly, we define $\Delta\phi_\alpha = \sum_{g_r=0}^{\infty} \Delta_{g_r} \phi_\alpha$. Now, summing (7.31) over g , we find that the dependence of \mathcal{A} on the choice of infrared cutoff used in calculating it is

$$(7.32) \quad \mathcal{A} \rightarrow \mathcal{A} + \sum_{\alpha} \Delta\phi_{\alpha} \frac{\partial}{\partial\phi_{\alpha}} \mathcal{A}.$$

This is still not the full answer because in arriving at this formula, we have only included contributions in which Σ_ℓ , to which the external vertex operators are attached, couples through a massless scalar of zero momentum to a single additional component Σ_r . In reality, Σ_ℓ may couple to any number $k \geq 0$ of such components (fig. 19). Associated to each such component is its own gluing parameter q with its logarithmically divergent measure $d^2q/\bar{q}q$. We regulate each such integral as before by a suitable choice of hermitian metric. The dependence on the choice of infrared regulator always gives the same factor $\sum_{\alpha} \Delta\phi_{\alpha} \cdot \mathcal{V}_{\alpha}$ found above, where by \mathcal{V}_{α} we mean the insertion in the scattering amplitude of a zero momentum ϕ_{α} vertex operator. As usual, \mathcal{V}_{α} can be replaced by a derivative of the scattering amplitude with respect to ϕ_{α} . Summing over k and remembering to include a factor of $1/k!$ since the components disappearing into the vacuum are equivalent, we exponentiate the result in (7.32). At this level, the full dependence of the scattering amplitude

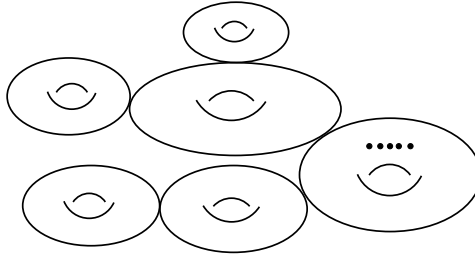


Figure 20: A “tree” of Riemann surfaces, with all vertex operators inserted on one component.

on the choice of infrared regulator takes the form

$$(7.33) \quad \mathcal{A} \rightarrow \exp \left(\sum_{\alpha} \Delta\phi_{\alpha} \frac{\partial}{\partial\phi_{\alpha}} \right) \mathcal{A}.$$

The operator

$$(7.34) \quad \mathcal{K} = \sum_{\alpha} \Delta\phi_{\alpha} \frac{\partial}{\partial\phi_{\alpha}}$$

is a vector field on the parameter space of the string compactifications under consideration, or in other words on the moduli space \mathcal{Z} of string theory vacua. So $\exp(\mathcal{K})$ is a diffeomorphism of that parameter space (in the sense of formal power series in g_{st}). What we learn from this analysis is that when massless tadpoles vanish, string perturbation theory constructs a natural family of perturbative S -matrices parametrized by \mathcal{Z} , but – at this level of generality – without a natural choice of parametrization of \mathcal{Z} . Different infrared regulators will give results that differ by a reparametrization of \mathcal{Z} .

What we have just described is really only part of an inductive procedure. The moduli spaces \mathcal{M}_{ℓ} and \mathcal{M}_r are themselves not compact, and in evaluating the integrals $\int_{\mathcal{M}_{\ell}} \mathcal{A}_{\ell}$ and $\int_{\mathcal{M}_r} \mathcal{A}_r$, one may have met tadpoles in lower order. Inductively, once all integrals up to genus $g - 1$ have been defined, one applies the above reasoning to the genus g integrals. The full analysis amounts to regularizing the contributions of “trees” of Riemann surfaces, with all external vertex operators on the same component, as indicated in fig. 20. Consider any component Σ^* in such a tree other than the component that contains the external vertex operators; let g^* be its genus. In general, Σ^* is joined at double points to some number $s > 0$ of other components, and the analysis of the

simultaneous $d^2q/\bar{q}q$ singularities associated to all the double points leads to a sum of terms, each of which is proportional to a correlation function $\langle \mathcal{V}_{\alpha_1} \dots \mathcal{V}_{\alpha_s} \rangle_{\mathcal{G}^*}$, where the \mathcal{V}_{α_i} are vertex operators of moduli fields at zero momentum. If all such correlation functions vanish (after integration over the moduli of Σ^*) then the analysis of general trees is similar to what we have already explained and the result is the same, except that \mathcal{K} becomes a nonlinear function of the $\Delta\phi_\alpha$. Vanishing of the zero-momentum correlation functions amounts to the condition that the effective potential is identically zero as a function of the ϕ_α . It is desirable but more difficult to develop the theory without this assumption, so as to include scalar fields that are massless without being moduli.

One can consider in a similar way infrared divergences associated to massless open strings. An open-string degeneration is still described by a gluing formula $xy = q$ (or its superanalog), but now q is real. The $d^2q/\bar{q}q$ singularity that we have analyzed is replaced by a slightly simpler dq/q singularity, which has the same logarithmic divergence and can be treated the same way. Different regularizations of the massless open-string degenerations differ by shifts in the expectation values of massless open-string fields of spin zero.

7.6.2. Another infrared regulator Here, we will explain another type of infrared regularization that can be used in analyzing these questions. The results will be equivalent to what we had before. One reason to describe this alternative regulator is to emphasize that different types of infrared regulator are possible and do lead to equivalent results.

We return to the basic question of studying the integral (7.22) that defines a scattering amplitude near the separating degeneration at $q = 0$. The form F can be approximated near $q = 0$ as

$$(7.35) \quad F_0 = \sum_{\alpha} \frac{d^2q}{\bar{q}q} \wedge \mathcal{G}_{\ell,\alpha} \wedge \mathcal{G}_{r,\alpha}.$$

An important point, which applies also to some expressions written below, is that the form F_0 is invariant under $q \rightarrow e^f q$, where f is a function on $\widehat{\mathcal{M}}_\ell \times \widehat{\mathcal{M}}_r$, because $\mathcal{G}_{\ell,\alpha} \wedge \mathcal{G}_{r,\alpha}$ is a form of top degree on $\widehat{\mathcal{M}}_\ell \times \widehat{\mathcal{M}}_r$. So F_0 is well-defined, independent of precisely how we define the q parameter, as a differential form on what we will call $\widehat{\mathcal{M}}_0$, the total space of the line bundle $\mathcal{N}^{-1} \rightarrow \widehat{\mathcal{M}}_\ell \times \widehat{\mathcal{M}}_r$. Suppose we can write $F_0 = d\Lambda$, where Λ is a form of real codimension 1. Then we modify the cutoff definition (7.26) by adding a boundary term:

$$(7.36) \quad \mathcal{A}_{g,\epsilon} = \int_{\widehat{\mathcal{M}}_\epsilon} F - \int_{\partial\widehat{\mathcal{M}}_\epsilon} \Lambda.$$

Here $\widehat{\mathcal{M}}_\epsilon$ is defined by $|q| \geq \epsilon$, and $\partial\widehat{\mathcal{M}}_\epsilon$ its boundary, defined by $|q| = \epsilon$. If $F = d\Lambda$, Stokes's theorem implies that the right hand side of (7.36) is invariant under $\epsilon \rightarrow e^h\epsilon$. If instead we actually have $d\Lambda = F_0$, where F_0 coincides with F only near $q = 0$, then the correct conclusion from Stokes's theorem is that the limit of $\mathcal{A}_{g,\epsilon}$ for $\epsilon \rightarrow 0$ is invariant under $\epsilon \rightarrow e^h\epsilon$.

We have eliminated the need to choose a hermitian metric on \mathcal{N} , but now we need to find a way to write $F_0 = d\Lambda$. For this, let us restrict to a point in $\widehat{\mathcal{M}}_\ell$ and view q as a section of the line bundle $\mathcal{L}_r^{-1} \rightarrow \widehat{\mathcal{M}}_r$. Consider the differential form

$$(7.37) \quad \chi_{r,\alpha} = \frac{d^2q}{\bar{q}q} \mathcal{G}_{r,\alpha}.$$

We view this as a top form on what we will call \mathcal{X}_r – the total space of the line bundle $\mathcal{L}_r^{-1} \rightarrow \widehat{\mathcal{M}}_r$, with the zero section removed. The reason that we remove the zero section is that $\chi_{r,\alpha}$ is singular at $q = 0$. We note that $\chi_{r,\alpha}$ is invariant under $q \rightarrow e^f q$ for any function f , so in particular it is well-defined. We want to find a form $\lambda_{r,\alpha}$ with

$$(7.38) \quad \chi_{r,\alpha} = d\lambda_{r,\alpha}.$$

Abstractly, $\lambda_{r,\alpha}$ exists since (because of the noncompactness of \mathcal{X}_r) the top-dimensional cohomology of \mathcal{X}_r vanishes. We will explain shortly how to make a fairly nice choice of $\lambda_{r,\alpha}$, but for the moment suppose this has been done. Then in the cutoff formula (7.36), we take

$$(7.39) \quad \Lambda = \sum_{\alpha} \mathcal{G}_{\ell,\alpha} \wedge \lambda_{r,\alpha}.$$

Clearly $d\Lambda = F_0$, as desired.

Let us now investigate the extent to which (7.36) depends on the choice of $\lambda_{r,\alpha}$. Since we are supposed to obey $d\lambda_{r,\alpha} = \chi_{r,\alpha}$, we can only add to $\lambda_{r,\alpha}$ a closed form. On the other hand, if we add an exact form to $\lambda_{r,\alpha}$, then Λ will change by an exact form and (7.36) is invariant. So we are only interested in shifting $\lambda_{r,\alpha}$ by a form $\Delta\lambda_{r,\alpha}$ that is closed but not exact. Let us write \mathcal{X}_ϵ for the subspace of \mathcal{X} defined by $|q| = \epsilon$. \mathcal{X} is contractible onto \mathcal{X}_ϵ (and topologically, \mathcal{X}_ϵ does not depend on the metric used in the condition $|q| = \epsilon$). The only invariant information in $\Delta\lambda_{r,\alpha}$, modulo exact forms, is the “period”

$$(7.40) \quad \Delta\phi_\alpha = - \int_{\mathcal{X}_\epsilon} \Delta\lambda_{r,\alpha}.$$

If we do shift $\lambda_{r,\alpha}$ in this fashion, then Λ is shifted by

$$(7.41) \quad \Lambda \rightarrow \Lambda + \sum_{\alpha} \mathcal{G}_{\ell,\alpha} \wedge \Delta \lambda_{r,\alpha},$$

and when we insert this in (7.36), and take the limit $\epsilon \rightarrow 0$, we find that $\mathcal{A}_{\mathfrak{g}}$ shifts by

$$(7.42) \quad \mathcal{A}_{\mathfrak{g}} \rightarrow \mathcal{A}_{\mathfrak{g}} + \sum_{\alpha} \Delta \phi_{\alpha} \int_{\widehat{\mathcal{M}}_{\ell}} \mathcal{G}_{\ell,\alpha}.$$

This is the familiar result of eqn. (7.29), and the rest of the analysis proceeds from there.

To conclude, we will describe a nice class of choices for $\lambda_{r,\alpha}$. A naive way to proceed is as follows. The vanishing tadpole condition (7.24) implies that the form $\mathcal{G}_{r,\alpha}$ on \mathcal{M}_r is exact, $\mathcal{G}_{r,\alpha} = d\beta_{r,\alpha}$, for some $\beta_{r,\alpha}$. So one might try

$$(7.43) \quad \lambda_{r,\alpha}^{(0)} = \frac{d^2 q}{\bar{q}q} \beta_{r,\alpha}.$$

But this expression is not invariant under $q \rightarrow e^f q$, so it only makes sense once one is given a trivialization of \mathcal{L}_r^{-1} . Under $q \rightarrow e^f q$, we have $\lambda_{r,\alpha}^{(0)} \rightarrow \lambda_{r,\alpha}^{(0)} - i \left(d\bar{f}(dq/q) + d\bar{q}/\bar{q} \cdot df \right) \beta_{r,\alpha}$. Note that $\Delta \lambda_{r,\alpha}^{(0)} = -i \left(d\bar{f}(dq/q) + d\bar{q}/\bar{q} \cdot df \right) \beta_{r,\alpha}$ is closed (since $d\bar{q}/\bar{q}$, dq/q , $df \wedge \beta_{r,\alpha}$, and $d\bar{f} \wedge \beta_{r,\alpha}$ are all closed; the last two are closed because they are top forms on $\widehat{\mathcal{M}}_r$), so this shift in $\lambda_{r,\alpha}^{(0)}$ does not affect the condition $d\lambda_{r,\alpha}^{(0)} = \chi_{r,\alpha}$. The upshot of this is that instead of (7.43), we should try

$$(7.44) \quad \lambda_{r,\alpha} = \frac{d^2 q}{\bar{q}q} \beta_{r,\alpha} + \frac{dq}{q} \gamma_{r,\alpha} + \frac{d\bar{q}}{\bar{q}} \tilde{\gamma}_{r,\alpha},$$

where $\gamma_{r,\alpha}$ and $\tilde{\gamma}_{r,\alpha}$ are top forms on $\widehat{\mathcal{M}}_r$. There is no problem in globally solving $d\lambda_{r,\alpha} = \chi_{r,\alpha}$ with $\lambda_{r,\alpha}$ of this form, though the way of writing $\lambda_{r,\alpha}$ as a sum of the three indicated terms depends on a choice of local trivialization of \mathcal{L}_r^{-1} . To prove that a global choice of $\lambda_{r,\alpha}$ of the claimed form does exist, one covers \mathcal{M}_r by small open sets on which one can pick a trivialization of \mathcal{L}_r^{-1} , so that one can choose $\lambda_{r,\alpha}$ in the form (7.43). Since $\beta_{r,\alpha}$ is globally defined, two such local solutions differ by terms of the form $\Delta \lambda = (dq/q) \Delta \gamma_{r,\alpha} + (d\bar{q}/\bar{q}) \Delta \tilde{\gamma}_{r,\alpha}$. The obstruction to modifying the local choices of $\lambda_{r,\alpha}$ by adding terms of the form $\Delta \lambda$ so that they fit together into a global $\lambda_{r,\alpha}$ is given by

a one-dimensional cohomology class of a smooth manifold, namely $\widehat{\mathcal{M}}_r$, with values in a coherent sheaf. Such cohomology always vanishes above dimension zero, so there is no obstruction to finding $\lambda_{r,\alpha}$.

7.6.3. Wavefunction renormalization Starting in section 7.4.3, we restricted attention to the case that mass renormalization vanishes. To understand mass renormalization in superstring perturbation theory requires a more general formalism than the one developed in the present paper. But there is something to say even if mass renormalization vanishes.

When mass renormalization vanishes, we meet a question that is exactly analogous to the question that we have been studying in the context of tadpoles. Mass renormalization is associated to a degeneration of the type sketched in fig. 16(b), with a single external vertex operator \mathcal{V} on one side, say on Σ_ℓ , and arbitrary insertions on Σ_r . Near such a degeneration, the genus g contribution to the scattering amplitude \mathcal{A}_g has a singular behavior that is just like that of the integral (7.23) that we have studied in the tadpole case. The behavior near the degeneration is

$$(7.45) \quad \mathcal{A}_{g,\text{sing}} = \sum_{\alpha=1}^s \int \frac{d^2q}{\bar{q}q} \int_{\widehat{\mathcal{M}}_\ell} \mathcal{G}_{\ell,\alpha} \int_{\widehat{\mathcal{M}}_r} \mathcal{G}_{r,\alpha}.$$

The label α now runs over all physical string states that at tree level are degenerate in mass with the external particle associated to the vertex operator \mathcal{V} .

The integral has a potential logarithmic divergence near $q = 0$, but absence of mass renormalization means that

$$(7.46) \quad \int_{\widehat{\mathcal{M}}_\ell} \mathcal{G}_{\ell,\alpha} = 0$$

for all α . This ensures the cancellation of the logarithmic divergence in $\mathcal{A}_{g,\text{sing}}$. We are now in a familiar situation. Vanishing of mass renormalization ensures that the integral for \mathcal{A}_g converges, but only conditionally so. If one introduces an infrared cutoff by restricting to $|q| \geq \epsilon$, then \mathcal{A}_g has a limit for $\epsilon \rightarrow 0$, but this limit depends on the hermitian metric that was used to define what we mean by $|q|$.

The interpretation is quite similar to what it was in the tadpole case. By imitating the previous arguments, one shows that the change in \mathcal{A}_g resulting from a change in the hermitian metric is equivalent to a scattering amplitude computed on Σ_r with an insertion of one more vertex operator, which reflects the output of the path integral on Σ_ℓ . This vertex operator is a

linear combination of the vertex operators \mathcal{V}_α that represent physical string states that have the same momentum and therefore the same mass as the one corresponding to \mathcal{V} . Thus the dependence on the choice of cutoff is equivalent to a wavefunction renormalization $\mathcal{V} \rightarrow \mathcal{V} + \sum_\alpha c_\alpha \mathcal{V}_\alpha$. The constants c_α are of order $g_{\text{st}}^{2g_\ell}$.

Note that \mathcal{V} itself is a linear combination of the \mathcal{V}_α . It is helpful to generalize the problem slightly by inserting on Σ_ℓ an arbitrary linear combination of the \mathcal{V}_α , rather than making a particular choice as we did in the above presentation. Then the dependence on the choice of cutoff would give us a general $s \times s$ wavefunction renormalization matrix, as one might expect in field theory. The S -matrix is natural, but there is no natural notion in general of computing it using vertex operators that are independent of the string coupling constant and the other moduli.

7.7. More detail on the anomaly

In section 7.3, we gave a somewhat heuristic explanation of an important result. The result concerned a separating degeneration in string theory, in which a worldsheet Σ decomposes to two components Σ_ℓ and Σ_r , joined at a double point, with a BRST-trivial vertex operator $\{Q_B, \mathcal{W}_1\}$ inserted on Σ_ℓ . The claim is that any BRST anomaly arising in this situation is always proportional to the amplitude obtained by inserting at the double point on Σ_r a physical state vertex operator $\mathcal{O}(\mathcal{W}_1)$ that depends linearly on \mathcal{W}_1 . To be more precise, this is supposed to be true in an inductive sense: assuming that there are no BRST anomalies in genus less than g , any BRST anomaly in genus g should have the property just stated.

The relevant ideas can be explained somewhat more simply for the case of an open-string degeneration. Moreover, supersymmetry will not be important and we will use the language of bosonic strings. Gluing for open strings is described by the usual gluing formula

$$(7.47) \quad xy = q,$$

where q is real and positive and x and y are local parameters on Σ_ℓ and Σ_r that are real along the boundary. For $q \rightarrow 0$, we glue the boundary point $x = 0$ in Σ_ℓ to the boundary point $y = 0$ in Σ_r . In section 7.3, we used the parameter s defined by $q = e^{-s}$, which parametrizes the length of a long strip joining the two branches, but here it will be more convenient to work with q , since this makes it easier to describe what is happening at $q = 0$ or $s = \infty$. The condition $q = 0$ defines a component \mathcal{B} of the boundary of $\widehat{\mathcal{M}}_{g,n}$. The

last statement, which is explained in more detail in section 7.4 of [16] (and also in section 9.2.1 below), makes open-string degenerations slightly simpler than closed-string degenerations for our present purposes, and that is why we consider this case. \mathcal{B} is a product of moduli spaces:

$$(7.48) \quad \mathcal{B} \cong \widehat{\mathcal{M}}_\ell \times \widehat{\mathcal{M}}_r.$$

This statement is an open-string analog of the closed-string statement (6.65). For open superstrings, q is replaced by the analogous parameter q_{NS} or q_{R} .

In section 7.3, motivated by eqn. (6.89), we claimed that the anomaly can be evaluated by inserting at the double point

$$(7.49) \quad \sum_i \left(c\partial c U_i \otimes c U^i + c U_i \otimes c\partial c U^i \right).$$

The sum runs over a complete set of physical open-string states with momentum and spin that match those of \mathcal{W}_1 . We have omitted the BRST-trivial term in (6.89), which can be dropped because we assume that there are no BRST anomalies in lower orders. The formula (7.49) suggests that, with $\mathcal{V}_i = cU_i$, the anomaly can be computed by a path integral on Σ_r with insertion of

$$(7.50) \quad \mathcal{O}(\mathcal{W}_1) = \sum_i a_i \mathcal{V}_i,$$

where the coefficient a_i is to be computed by a path integral on Σ_ℓ with an insertion of $c\partial c U_i$ (along with \mathcal{W}_1 and possibly other vertex operators).

Although it is true that the anomaly can be computed from an insertion of an operator $\mathcal{O}(\mathcal{W}_1)$ of the form indicated in (7.50), the proposed formula for the coefficients a_i is oversimplified. This must be the case, as the insertion of $c\partial c U_i$ cannot arise in our formalism, for this operator is not annihilated by b_0 .

To analyze the problem more systematically, we return to the basic eqn. (7.1) for the anomaly, which we repeat for convenience:

$$(7.51) \quad \int_{\widehat{\mathcal{M}}_{\mathbf{g},\mathbf{n}}} F_{\{Q_B, \mathcal{W}_1\}, \mathcal{V}_2, \dots, \mathcal{V}_n} = - \int_{\partial \widehat{\mathcal{M}}_{\mathbf{g},\mathbf{n}}} F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n}.$$

Since \mathcal{W}_1 and $\mathcal{V}_2, \dots, \mathcal{V}_n$ are all conformal vertex operators – conformal primary fields of dimension 0 annihilated by the antighost modes b_n , $n \geq 0$ (and by β_r , $r \geq 0$, in the superstring case) – the forms $F_{\{Q_B, \mathcal{W}_1\}, \mathcal{V}_2, \dots, \mathcal{V}_n}$ and $F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n}$ are both pullbacks from $\widehat{\mathcal{M}}_{\mathbf{g},\mathbf{n}}$. The former is a top form that can be integrated over $\widehat{\mathcal{M}}_{\mathbf{g},\mathbf{n}}$, and the latter is a form of codimension 1 that

can be integrated over a codimension 1 submanifold, such as the boundary of $\widehat{\mathcal{M}}_{g,n}$. In particular, we will study the behavior near the component \mathcal{B} of the boundary.

If $F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n}$ were non-singular along \mathcal{B} , then we could literally restrict it to $\mathcal{B} \cong \widehat{\mathcal{M}}_\ell \times \widehat{\mathcal{M}}_r$. The restriction could be expressed in terms of insertion at the double point of some bilinear expression in local operators, schematically $\mathcal{S} = \sum_\alpha \mathcal{R}_\alpha \otimes \mathcal{R}'_\alpha$, where \mathcal{R}_α and \mathcal{R}'_α are local operators on Σ_ℓ and Σ_r , respectively. But \mathcal{R}_α and \mathcal{R}'_α must be conformal vertex operators because of the conformal invariance of our formalism, and \mathcal{S} must have total ghost number 3. These conditions are incompatible, as the maximum possible ghost number of a conformal vertex operator is 1 (corresponding to conformal vertex operators $\mathcal{V} = cU$, where U is a matter primary of dimension 1), and $1 + 1 < 3$. We conclude that if $F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n}$ is nonsingular along \mathcal{B} , so that its restriction to \mathcal{B} can be defined, then this restriction is actually 0, and in particular there is no anomaly. (Essentially the same argument applies to open superstring theory, using the constraints on superconformal vertex operators explained in sections 4 and 5. The extension to closed-string degenerations is also straightforward.)

In general, $F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n}$ is singular at $q = 0$. In fact, the singularity can be computed via the insertion

$$(7.52) \quad \frac{dq}{q} \sum_i cU_i \otimes cU^i$$

that is familiar from section 6 and especially from eqn. (6.84). (The operators cU_i that appear here are the most general conformal vertex operators of ghost number 1. The conformal invariance of the formalism ensures that other operators cannot arise.) At first sight, it may seem that we are not interested in this contribution to $F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n}$, since we are interested in setting q to a small “constant,” rather than integrating over q . However, this is not correct because q is really a section of a real line bundle, rather than a real number, so setting q to a constant value is not a natural operation. (The dependence of q on the choice of local parameters was described explicitly in eqn. (6.111).) It is meaningful to multiply q by a small positive constant or equivalently to add a large constant to s , but there is no natural notion of setting q to a constant value. We will see that the dq/q term has to be included to write a conformally invariant formula for the coefficients a_i of eqn. (7.50).

In writing eqn. (7.52), we are using a sort of hybrid formalism. The expression dq/q is an explicit one-form on the normal direction to \mathcal{B} in $\widehat{\mathcal{M}}$. In addition, the path integral on $\Sigma = \Sigma_\ell \cup \Sigma_r$ with insertion of $\sum_i cU_i \otimes cU^i$ will

generate a codimension 1 form on $\mathcal{M}_\ell \times \mathcal{M}_r$. (This form has codimension 1 because one of the operators inserted on Σ_ℓ , namely \mathcal{W}_1 , is a gauge parameter rather than the vertex operator for a physical state.) So eqn. (7.52) describes the singularity along \mathcal{B} of the codimension 1 form $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$ on $\widehat{\mathcal{M}}$.

In addition to the singular terms (7.52), $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$ has the contributions of eqn. (7.49) that are nonsingular for $q \rightarrow 0$. Combining them gives

$$(7.53) \quad \frac{dq}{q} \sum cU_i \otimes cU^i + \sum_i \left(c\partial cU_i \otimes cU^i + cU_i \otimes c\partial cU^i \right),$$

a formula that suffices for evaluating $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$ modulo terms that vanish along \mathcal{B} . The fact that the same matter primary fields U_i appear in these singular and nonsingular contributions to $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$ can be explained as follows. To evaluate the codimension 1 differential form $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$, we are supposed to make all possible insertions of antighost modes. To evaluate a term in $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$ that is proportional to $ds = dq/q$, one of the insertions should be a b_0 insertion in the narrow neck. This removes ∂c from the operators. To evaluate contributions to $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$ with no dq/q , we omit the b_0 insertion (making instead an extra antighost insertion somewhere else), in which case we are left with ∂c in the operators. But the matter primaries that appear are the same.

Now let us show that the combined formula (7.53) behaves correctly with respect to reparametrizations of Σ_ℓ and Σ_r . It suffices to consider reparametrizations of Σ_ℓ . Consider a reparametrization $x \rightarrow \hat{x}(x)$ that leaves fixed the point at which the gluing occurs, so that $\hat{x} = 0$ at $x = 0$. The operators cU_i and $c\partial cU_i$ are primary fields of dimension 0, so they are unaffected by this reparametrization. The reparametrization multiplies q by the constant $\partial\hat{x}/\partial x|_{x=0}$ (see eqn. (6.111)). The differential form dq/q is invariant under this rescaling. So eqn. (7.53) is reparametrization-invariant.

Just as in section 2.4.2, we need an additional condition beyond reparametrization invariance to ensure that $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$ is a pullback from moduli space. The additional condition says that $F_{\mathcal{W}_1, \nu_2, \dots, \nu_n}$ should vanish if contracted with a vector field induced from a diffeomorphism of Σ . In the present context, such a vector field is induced from a vector field that generates a reparametrization of Σ_ℓ that leaves fixed the point $x = 0$ along with the corresponding rescaling of q :

$$(7.54) \quad v(x)\partial_x + \left. \frac{\partial v(x)}{\partial x} \right|_{x=0} q\partial_q, \quad v(0) = 0.$$

(Of course, we also consider in a similar way reparametrizations of Σ_r .) As we know from section 2.4.2, contraction with the vector field on the space of

metrics that is induced from the vector field $v(x)\partial_x$ on Σ_ℓ has the effect of $c \rightarrow c + v$. On an operator supported at $x = 0$, since $v(0) = 0$, this leaves $c(0)$ unchanged and this acts by $\partial c(0) \rightarrow v'(0)$. On the other hand, contraction with $v'(0)q\partial_q$ maps dq/q to $v'(0)$. The two contributions cancel (a minus sign appears because the contraction operation $\partial c(0) \rightarrow v'(0)$ anticommutes with the fermionic field c), so the expression in (7.53) does have the desired property to make $F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n}$ a pullback.

To evaluate the anomaly, we observe the following. First of all, an insertion of $c\partial c U^i$ on Σ_r will vanish because the ghost number of this operator is too large by 1. So the anomaly will come from an insertion of $\sum_i a_i c U^i$ on Σ_ℓ , where the coefficients a_i can be computed by integration on Σ_ℓ . To compute the a_i , we cannot set $q = 0$ and integrate over \mathcal{M}_ℓ , since the form we want to integrate is singular at $q = 0$. Rather, we define an integration cycle \mathcal{M}_ℓ^* that is isomorphic to \mathcal{M}_ℓ by taking q to be small (for example, we could fix any metric on the real line bundle where q takes values and define \mathcal{M}_ℓ^* by $|q| = \epsilon$, with ϵ a small positive constant). If we set

$$(7.55) \quad \mathcal{X}_i = \frac{dq}{q} c U_i + c \partial c U_i,$$

then the coefficients a_i are

$$(7.56) \quad a_i = \int_{\mathcal{M}_\ell^*} F_{\mathcal{W}_1, \dots, \mathcal{X}_i}.$$

In other words, they are obtained by a path integral in which we insert on Σ_ℓ the expression \mathcal{X}_i as well as \mathcal{W}_1 and any other operators (indicated by the ellipses in (7.56)) that were present on Σ_ℓ at the beginning, and then integrate over \mathcal{M}_ℓ^* . It is hopefully now clear that the dq/q term in \mathcal{X}_i is necessary here.

For open-string degenerations in superstring theory, everything is almost the same, with q replaced by q_{NS} or q_{R} , and with $c U_i$ replaced by its superconformal analog. For Ramond degenerations, one has to include the integration over the fermionic gluing parameter. For closed bosonic strings, the gluing parameter is a complex variable $q = \exp(-(s + i\alpha))$. To evaluate the anomaly, we want to integrate over α while fixing $|q|$. As in the derivation of the propagator in section 6.2.1, the integration over α is associated to an insertion of $b_0 - \tilde{b}_0$ and leads to a factor of $2\pi\delta_{L_0 - \tilde{L}_0}$. Just as in the open-string case, there is no natural operation of setting $|q|$ to a constant and instead one must develop a formalism with terms proportional to $d|q|/|q|$. The anomaly form $F_{\mathcal{W}_1, \mathcal{V}_2, \dots, \mathcal{V}_n}$ has terms proportional to $d|q|/|q|$ that can be computed by making a $b_0 + \tilde{b}_0$ insertion and regular terms that can be computed by instead

making a different antighost insertion. The analog of eqn. (7.53) is

$$(7.57) \quad \frac{d(|q|^2)}{|q|^2} \sum_i \tilde{c}cU_i \otimes \tilde{c}cU^i + \sum_i \left(\tilde{c}c(\tilde{\partial}\tilde{c} + \partial c)U_i \otimes \tilde{c}cU^i + \tilde{c}cU_i \otimes \tilde{c}c(\tilde{\partial}\tilde{c} + \partial c)U^i \right).$$

This leads to an obvious analog of the formula (7.56) for the anomaly coefficients. For closed superstrings, one has left and right gluing parameters q_ℓ and q_r (as in section 6.6, one can define the integration cycle so that they are complex conjugates modulo the odd variables). Eqn. (7.57) has an immediate analog, with $|q|^2$ replaced by $q_\ell q_r$, and the vertex operators $\tilde{c}cU_i$ replaced by their superconformal analogs. In Ramond sectors, one also integrates over the fermionic gluing parameters.

The main conclusions of this analysis – such as eqns. (7.53) and (7.57) – must be supplemented at zero momentum with exceptional terms whose origin we first saw in eqns. (6.90) and (6.98).

8. Spacetime supersymmetry and its consequences

We focus here on the spacetime supersymmetry of the S -matrix. In fact, in any supersymmetric compactification of string theory, there is at tree level a massless field of spin $3/2$, known as the gravitino. Its tree-level couplings are constrained by spacetime supersymmetry and in particular at zero momentum these couplings are proportional to the matrix elements of the supercurrent. Perturbative corrections cannot alter the fact that the gravitino couplings are non-zero at zero momentum.

Assuming that the perturbative S -matrix exists and the gravitino remains massless in perturbation theory, spacetime supersymmetry of the perturbative S -matrix follows just from these facts. Indeed, just as the existence of a massless spin 1 particle with nonvanishing couplings at zero momentum implies conservation of electric charge, and a massless spin 2 particle with nonvanishing couplings at zero momentum must interact like a graviton [74], the existence of a massless spin $3/2$ particle with nonvanishing couplings at zero momentum implies spacetime supersymmetry [75]. (Such arguments were originally applied to string theory in [76, 77].)

We elaborate on this point in section 8.1. However, the information one can gain from such arguments appears to be not quite adequate for our purposes. To address the tadpole problem, we seem to need an argument that can be formulated directly at zero momentum in spacetime, not by taking a limit from non-zero momentum. Also, we should not assume *a priori* that the

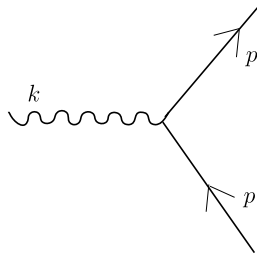


Figure 21: The vertex coupling a massless “photon” to a massive charged scalar.

perturbative S -matrix exists; spacetime supersymmetry is supposed to be an ingredient in proving this.

So in section 8.2.2, we describe a more precise and stringy proof of spacetime supersymmetry. The main tool is what one might call the spacetime supercurrent. In the standard conformal field theory language, this is simply the holomorphic (or antiholomorphic) fermion vertex operator of [1], at zero spacetime momentum. We then go on to show in what sense this operator generates spacetime supersymmetry, and how it can be used to show – under appropriate conditions – that massless tadpoles vanish and hence the S -matrix exists.

8.1. Massless particles and conserved charges

Here we will review the arguments [74, 75] showing that massless particles of spin ≥ 1 couple at zero momentum to conserved charges. The couplings of a given massless field might vanish at zero momentum (this is the case in perturbative superstring theory for massless Ramond-Ramond gauge fields in the absence of D-branes), but if a massless field of spin ≥ 1 has nonvanishing couplings at zero momentum, then it couples to a conserved charge. We will be brief since the considerations here are not novel, and in any event we will introduce a more explicit approach to conserved charges in string theory in section 8.2.2. The purpose of this section is to explain what can be understood based on general arguments that do not involve the details of string theory.

8.1.1. Gauge theory Let us consider an S -matrix element describing the coupling of a massless spin 1 “photon” of momentum k and polarization vector ε (with $k^2 = k \cdot \varepsilon = 0$) to n additional particles of masses m_i , momenta p_i , and charges e_i , with $i = 1, \dots, n$. Gauge invariance means that S -matrix

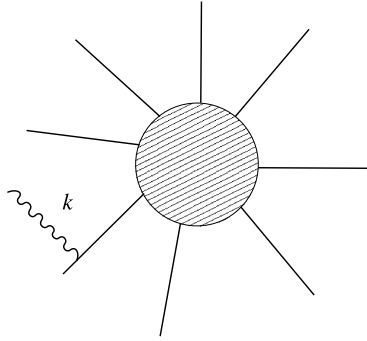


Figure 22: The most singular contributions to scattering of a soft photon come from processes in which the photon is attached to an external charged line, as sketched here.

elements must vanish if $\varepsilon = k$. We consider the photon to be soft, meaning that we will study the limit $k \rightarrow 0$, and we consider the other particles to be non-soft, meaning that their momenta will have non-zero limits for $k \rightarrow 0$. For simplicity, we take the non-soft particles to have spin 0, in which case, their propagators (in Lorentz signature) are

$$(8.1) \quad \frac{i}{p_i^2 - m_i^2}.$$

For small k , the vertex by which the photon couples to the i^{th} non-soft particle is (fig. 21)

$$(8.2) \quad -ie_i \varepsilon \cdot (p_i + p'_i), \quad p'_i = p_i + k.$$

The specific form of the trilinear vertex assumed in (8.2) follows from⁶⁷ gauge invariance and Lorentz invariance applied to the on-shell three-point function.

The condition of gauge invariance can be usefully analyzed for small k . For $k \rightarrow 0$, the most singular contributions to the scattering amplitude \mathcal{A} come from processes in which the soft photon is attached to one of the external lines (fig. 22). In perturbative field theory, one can understand this picture in terms of Feynman diagrams, but more generally one can understand it simply

⁶⁷This statement holds above four dimensions. In four dimensions, there is a more general possibility involving magnetic charge. This is not relevant to superstring perturbation theory, as there are no magnetic monopoles in superstring perturbation theory.

as a pictorial description of a certain singularity of the S -matrix, associated to an on-shell particle in a particular channel. The contribution to \mathcal{A} with the soft photon attached to the i^{th} external line factors as

$$(8.3) \quad -ie_i \varepsilon \cdot (p_i + p'_i) \frac{i}{(p'_i)^2 - m_i^2} = \frac{e_i \varepsilon \cdot p_i}{k \cdot p_i},$$

times an amplitude \mathcal{A}' with no external photon and with the i^{th} external momentum shifted slightly from p_i to p'_i . (We used $p_i^2 = m_i^2$, $k^2 = 0$, and $p'_i = p_i + k$; we also dropped a term proportional to k in the numerator.) The shift in p_i is unimportant in the soft limit, by which we mean the limit $k \rightarrow 0$ with the other momenta fixed. Adding all contributions with a soft photon attached to an external line, the scattering amplitude \mathcal{A} with the soft photon behaves for $k \rightarrow 0$ as

$$(8.4) \quad \mathcal{A} \sim \sum_{i=1}^n \frac{e_i \varepsilon \cdot p_i}{k \cdot p_i} \mathcal{A}'.$$

Therefore, in the soft limit, the condition for \mathcal{A} to vanish if $\varepsilon = k$ is

$$(8.5) \quad \sum_i e_i = 0,$$

or in other words conservation of electric charge. We simplified the reasoning slightly by assuming that the non-soft particles were all of spin 0; for the general case, see [74].

It is instructive to generalize this to the case of several massless fields of spin 1. In that case, it is important to consider the possibility that there are several non-soft particles of the same mass, so rather than speaking of the i^{th} non-soft particle in a given amplitude, we speak of the i^{th} mass level of non-soft particle in that amplitude. A massless field close to zero momentum can only couple a non-soft particle to another particle of the same mass. If there are s massless “photons,” then in general the coupling of the a^{th} one to the non-soft particles will take the form

$$(8.6) \quad -it_{a,i} \varepsilon \cdot (p + p'),$$

where $t_{a,i}$ is a matrix that acts on the states at the i^{th} mass level.⁶⁸ First let us consider processes in which a single soft “photon” of type a is coupled to

⁶⁸We absorb the coupling constants in the definition of the $t_{a,i}$, and again we assume for simplicity that the non-soft fields have spin 0.

non-soft particles of momenta p_1, \dots, p_n . As in (8.4), the amplitude behaves for small photon momentum k as

$$(8.7) \quad \mathcal{A} \sim \sum_{i=1}^n \frac{t_{a,i} \varepsilon \cdot p_i}{k \cdot p_i} \mathcal{A}',$$

and gauge-invariance implies that

$$(8.8) \quad \sum_i t_{a,i} \mathcal{A}' = 0.$$

This is usually described by saying that the spin-one field of type a couples to the conserved charge t_a ; one interprets $t_{a,i}$ as the matrix by which the symmetry t_a acts on the i^{th} mass level. So far, this is not really a significant generalization of (8.5); as long as we consider only one massless field of spin 1, we can diagonalize the charge matrices $t_{a,i}$ and reduce to the previous result. Now consider an amplitude with two soft photons, say of types a and b , with momenta k_a, k_b and polarization vectors $\varepsilon_a, \varepsilon_b$, coupling to non-soft particles. The most singular contributions are found by attaching the soft particles on external lines. If we attach the two soft photons to two different external lines (fig. 23(a)), we get a contribution

$$(8.9) \quad \sum_{i \neq i'} \frac{t_{a,i} \varepsilon_a \cdot p_i}{k_a \cdot p_i} \frac{t_{b,i'} \varepsilon_b \cdot p_{i'}}{k_b \cdot p_{i'}} \mathcal{A}'$$

to the scattering amplitude. We test gauge invariance for particle a by setting $\varepsilon_a = k_a$, whereupon this contribution becomes

$$(8.10) \quad \sum_{i \neq i'} t_{a,i} \frac{t_{b,i'} \varepsilon_b \cdot p_{i'}}{k_b \cdot p_{i'}} \mathcal{A}' = - \sum_i \frac{t_{b,i} \varepsilon_b \cdot p_i}{k_b \cdot p_i} t_{a,i} \mathcal{A}',$$

where in the last step, we use (8.8) (and the fact that $t_{a,i}$ commutes with $t_{b,i'}$ for $i \neq i'$). Evidently, these contributions to the scattering amplitude are not gauge-invariant by themselves. Another contribution comes by attaching the two soft photons to the same external line. This can be done in two possible orderings (fig. 23(b)). The resulting contribution to the amplitude is, in the soft limit $k_a, k_b \rightarrow 0$,

$$(8.11) \quad \left(\sum_i \frac{t_{a,i} \varepsilon_a \cdot p_i}{k_a \cdot p_i} \frac{t_{b,i} \varepsilon_b \cdot p_i}{(k_a + k_b) \cdot p_i} + a \leftrightarrow b \right) \mathcal{A}'.$$

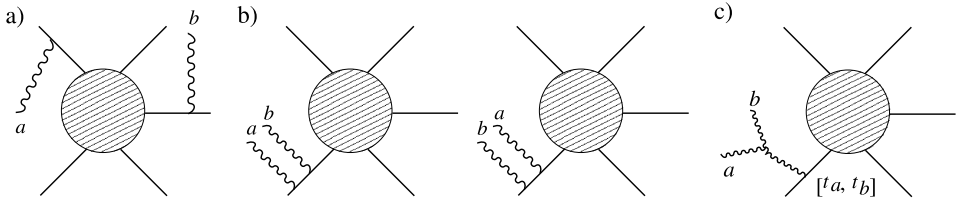


Figure 23: Couplings of two soft photons of types a and b to non-soft particles. The most singular contributions to the S -matrix arise from (a) processes in which the photons attach to distinct external lines, (b) processes in which they couple to the same external line in one of the two possible orderings, and (c) processes involving a trilinear coupling of massless fields of spin 1. The label $[a, b]$ in (c) indicates a massless spin 1 field coupling to $[t_a, t_b]$.

Upon setting $\varepsilon_a = k_a$, this becomes

$$(8.12) \quad \left(\sum_i \frac{t_{b,i} \varepsilon_b \cdot p_i}{k_b \cdot p_i} t_{a,i} + \sum_i [t_{a,i}, t_{b,i}] \frac{\varepsilon_b \cdot p_i}{(k_a + k_b) \cdot p_i} \right) \mathcal{A}'.$$

The first term in (8.12) cancels the contribution (8.10) from insertions on distinct lines. To cancel the second term, we need a new singular contribution (fig. 23(c)) involving a massless spin 1 particle that couples to $[t_a, t_b]$; this field must participate in a trilinear vertex with the external photons of types a and b . We are beginning to uncover here the basic structure of Yang-Mills theory: the conserved charges form a Lie algebra, and the structure constants of this Lie algebra determine trilinear couplings of massless spin 1 fields.

All of these considerations are valid only if the spin 1 fields under discussion are truly massless. The soft limit does not make sense for massive spin 1 particles.

8.1.2. Gravity and supergravity What we have summarized in section 8.1.1 has close analogs for theories with massless particles of spin 3/2 or 2. We refer the reader to the references and merely state the conclusions. In the context of a Poincaré invariant S -matrix, a massless particle of spin 2 that does not decouple in the zero-momentum limit must couple at zero momentum to the stress tensor, with a universal coefficient [74]. This statement can be expressed as the equality of gravitational and inertial mass, something that is usually deduced from the Principle of Equivalence. The proof is rather similar to the proof of charge conservation in the spin 1 case. Going farther

in this vein, one can deduce the Einstein equations (with possible higher derivative corrections), starting simply with a Poincaré-invariant theory of a massless spin 2 particle – the “graviton” – that does not decouple at zero momentum [78]. A derivation of this along the lines of what we explained for Yang-Mills theory would proceed by considering a scattering amplitude with two or more soft gravitons coupled to non-soft particles; as we saw for Yang-Mills theory, gauge-invariance would require nonlinear interactions among the gravitons.

There is a similar story for massless fields of spin 3/2. In the context of a Poincaré invariant S -matrix, a massless spin 3/2 field – the “gravitino” – that does not decouple at zero momentum must couple at zero momentum to a conserved supersymmetry current [75], and the S -matrix must be supersymmetric. The proof is again rather similar to the proof of charge conservation in the spin 1 case. To show that the conserved supercharges obey the standard supersymmetry algebra, one considers S -matrix elements with two soft gravitinos. From gauge-invariance of such S -matrix elements, one can deduce [75] that the supersymmetry algebra must take a standard form and also that in addition to the gravitino, there must be a massless spin 2 particle, the graviton. (If there are massless fields of spin 1, then in the right dimensions and with the right amount of unbroken supersymmetry, it is possible for the supersymmetry algebra to have central charges.) The low energy structure is that of supergravity. The reasoning involved is similar to the reasoning by which we showed in section 8.1.1 that if there are massless spin 1 fields coupled to t_a and to t_b , then there must be one coupled to $[t_a, t_b]$.

The output of this reasoning is a supersymmetric identity obeyed by the S -matrix, which can be stated as follows. Let \mathcal{A}' be an amplitude for scattering of particles with masses m_1, \dots, m_n . Let Q_α be the supercharges of the theory and let $Q_{\alpha,i}$ be the matrices by which these act at the i^{th} mass level. Then

$$(8.13) \quad \sum_{i=1}^n Q_{\alpha,i} \mathcal{A}' = 0.$$

The resemblance to eqn. (8.8) is hopefully obvious, and the derivation by taking a $k \rightarrow 0$ limit of a scattering amplitude with a soft gravitino is similar. Of course, the relation (8.13) only holds if the gravitino is truly massless; otherwise one cannot take the limit $k \rightarrow 0$. In perturbation theory around a supersymmetric classical background, the relation holds if the gravitino remains massless in perturbation theory.

8.2. Gauge symmetries in string theory and conserved charges

We will now examine the relation between gauge symmetries and conserved charges in string theory. We begin in section 8.2.1 with symmetries of bosonic string theory and symmetries coming from the NS (or NS-NS) sector of superstring theory. These are easy to understand because the conserved charges in question are manifest symmetries of string perturbation theory. Then starting in section 8.2.2, we investigate spacetime supersymmetry.

8.2.1. Manifest symmetries of the worldsheet theory Let us begin with the closed bosonic string theory in \mathbb{R}^{26} , and consider for illustration the gauge parameter $\mathcal{W} = c\varepsilon_I \partial X^I e^{ik \cdot X}$. For $k^2 = \varepsilon \cdot k = 0$, this is a conformal primary of dimension 0, and of course it obeys our usual condition of not depending on ∂c or $\tilde{\partial} \tilde{c}$. The corresponding null state is

$$(8.14) \quad \mathcal{V} = \{Q_B, \mathcal{W}\} = \tilde{c} c i k_J \varepsilon_I \tilde{\partial} X^J \partial X^I e^{ik \cdot X}$$

and is a linear combination of longitudinal graviton and B -field vertex operators. (The orientation-reversed gauge parameter $\tilde{\mathcal{W}} = \tilde{c} \varepsilon_I \tilde{\partial} X^I e^{ik \cdot X}$ generates a different linear combination of graviton and B -field gauge transformations; symmetric and antisymmetric combinations are associated to graviton or B -field gauge transformations only.)

The decoupling of the pure gauge mode \mathcal{V} can be proved using the BRST machinery, but as we have described in sections 2.5.3 and 7.4.1, for gauge transformations of massless states of the bosonic string, a more explicit approach is available. The vertex operator \mathcal{V} can be written

$$(8.15) \quad \mathcal{V} = \tilde{c} c \tilde{\partial} (\varepsilon_I \partial X^I e^{ik \cdot X}),$$

so its integrated version is the total derivative $V = \tilde{\partial} (\varepsilon_I \partial X^I e^{ik \cdot X})$. As V is a total derivative, its integral $\int_{\Sigma} V$ vanishes, and this ensures the decoupling of the null state represented by \mathcal{V} , establishing gauge-invariance for the fields in question.

The arguments reviewed in section 8.1 relating gauge-invariance and conserved charges make perfect sense in string theory as well as in field theory. These arguments involve taking a limit of S -matrix elements for $k \rightarrow 0$. However, it will become clear that one can learn more by setting $k = 0$ at the beginning, whereupon $\mathcal{V} = V = 0$, as is evident in eqn. (8.14). The relation $\mathcal{V} = \{Q_B, \mathcal{W}\}$ becomes

$$(8.16) \quad 0 = \{Q_B, \mathcal{W}\}$$

and the relation $V = \tilde{\partial}(\varepsilon_I \partial X^I e^{ik \cdot X})$ becomes

$$(8.17) \quad 0 = \tilde{\partial}J, \quad J = \varepsilon_I \partial X^I.$$

So J is a conserved current and therefore generates a manifest symmetry of the worldsheet theory. Actually, once we set k to zero, the constraint $\varepsilon \cdot k = 0$ becomes vacuous and any ε is allowed. So we get a whole family of conserved currents

$$(8.18) \quad J^I = \partial X^I.$$

For the bosonic string in \mathbb{R}^{26} , the symmetry associated to these conserved currents is simply translation invariance $X^I \rightarrow X^I + a^I$ (with constant a^I) and the associated conserved quantity is the energy-momentum (or more briefly the momentum). Conservation of J^I can be applied to closed-string amplitudes on a worldsheet Σ of any genus and therefore the associated conservation law is valid to all orders of bosonic closed-string perturbation theory.

Of course, we can also define a second conserved current $\tilde{J}^I = \tilde{\partial}X^I$. In \mathbb{R}^{26} , the currents J^I and \tilde{J}^I are associated to the same conserved quantities – the energy-momentum. But if some of the X^I are circle-valued, as is appropriate to describe strings in $\mathbb{R}^{26-m} \times T^m$, where T^m is an m -torus, then the conserved charges generated by J^I are linear combinations of momentum and winding numbers, while \tilde{J}^I are associated to different linear combinations of the same conserved quantities. The linear combinations $J^I_+ = \star dX^I$ and $J^I_- = dX^I$ are associated to momentum and winding, respectively.

In the case of a theory with open as well as closed bosonic strings, the statements of the last two paragraphs assume that the boundary conditions along $\partial\Sigma$ are invariant under the symmetry generated by the current under consideration (this is so precisely if the normal component of the current vanishes along $\partial\Sigma$). Otherwise, mixing of closed and open strings can spontaneously break a closed-string gauge symmetry, as explained in section 7.4.1. Some examples of boundary conditions that preserve or do not preserve a symmetry are as follows [80, 81]. Neumann boundary conditions for a scalar field X^I preserve J^I_+ but not J^I_- , while Dirichlet boundary conditions preserve J^I_- but not J^I_+ .

Symmetries associated to conserved currents on the string worldsheet are usually so obvious in perturbation theory that a sophisticated formalism is not really necessary. That is certainly so in the case of translation invariance. However, for comparison with what we will say about spacetime supersymmetry, it is instructive to derive a Ward identity, as follows. A vertex operator

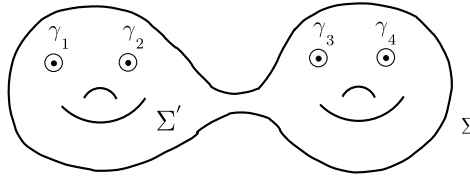


Figure 24: Σ' is made by omitting small open balls around each of the vertex operator insertions $\mathcal{V}_1, \dots, \mathcal{V}_n$ in the string worldsheet Σ . The boundary of Σ' is a union of circles $\gamma_1, \dots, \gamma_n$ that enclose the \mathcal{V}_i . In the example sketched here, $n = 4$.

\mathcal{V} is said to have momentum p^I if

$$(8.19) \quad \frac{1}{2\pi\alpha'} \oint_{\gamma} J^I \cdot \mathcal{V} = p^I \mathcal{V}.$$

The integral is taken over a contour γ that encloses the operator \mathcal{V} once in the counterclockwise direction. Now consider n vertex operators $\mathcal{V}_1, \dots, \mathcal{V}_n$ of momenta p_1, \dots, p_n , inserted at points $u_1, \dots, u_n \in \Sigma$. Let Σ' be obtained from Σ by omitting a small open ball around each of the u_i (fig. 24). The boundary of Σ' is a union of circles $\gamma_1, \dots, \gamma_n$, with γ_i wrapping once around u_i in the counterclockwise direction. Since $dJ = 0$, we have

$$(8.20) \quad 0 = \left\langle \int_{\Sigma'} dJ \cdot \mathcal{V}_1 \dots \mathcal{V}_n \right\rangle = \sum_{i=1}^n \left\langle \mathcal{V}_1 \dots \mathcal{V}_{i-1} \left(\oint_{\gamma_i} J \cdot \mathcal{V}_i \right) \mathcal{V}_{i+1} \dots \mathcal{V}_n \right\rangle.$$

No integration over moduli is relevant here; we make the appropriate antighost insertions so that the correlation function in question is not trivially zero, but we do not integrate over any moduli. The correlation function $\langle J \mathcal{V}_1 \dots \mathcal{V}_n \rangle$ is singular when J approaches any of the \mathcal{V}_i . It is only because of this singularity that the integral over a small circle γ_i surrounding one of the operators is non-zero. According to (8.19), the contribution of the singularity can be evaluated by replacing $\oint_{\gamma_i} J \cdot \mathcal{V}_i$ with $2\pi\alpha' p_i^I \mathcal{V}_i$, so that (8.20) becomes

$$(8.21) \quad 0 = \left(\sum_{i=1}^n p_i \right) \langle \mathcal{V}_1 \dots \mathcal{V}_n \rangle.$$

Thus the correlation function $\langle \mathcal{V}_1 \dots \mathcal{V}_n \rangle$ vanishes unless the charge associated

to J is conserved,

$$(8.22) \quad \sum_{i=1}^n p_i = 0.$$

This is analogous to the conservation laws that we derived in section 8.1.1.

There have been two main steps in our reasoning:

1. At $k = 0$, the gauge parameter \mathcal{W} becomes BRST-invariant; the relation $\mathcal{V} = \{Q_B, \mathcal{W}\}$ reduces to $\{Q_B, \mathcal{W}\} = 0$.
2. The integrated version of the vertex operator is a total derivative, $V = dJ$, so the fact that $V = \mathcal{V} = 0$ at $k = 0$ means that J becomes a conserved current on the string worldsheet.

The first of these two steps is completely general and applies to gauge symmetries of all massless states in all of the bosonic and supersymmetric string theories. It just reflects the fact that gauge transformations of massless fields are proportional to the derivative of the gauge parameter, so that the gauge transformations act trivially at $k = 0$.

The second step is more special. It applies to gauge symmetries of arbitrary massless states of the bosonic string. It also applies to gauge symmetries of massless superstring states that come from the NS (or NS-NS) sector, because the integrated vertex operator corresponding to a massless NS null vector is a total derivative on the worldsheet, as explained in section 4.3.1. So momentum and winding symmetry of superstring theory can be treated exactly as we have described for the bosonic string. For example, for the heterotic string, the relevant conserved currents are $\tilde{\partial}X^I$ and $D_\theta X^I$.

However, the second step listed above does not hold for spacetime supersymmetry, to which we turn next.

8.2.2. Spacetime supersymmetry In discussing spacetime supersymmetry in closed, oriented string theory, we will take the heterotic string as the basic example. As we explain in section 9, the analog for Type II superstring theory involves no essential novelty, but some new things do happen for open and/or unoriented superstrings.

We consider a supersymmetric compactification of the heterotic string to \mathbb{R}^d . The value of d , the number of unbroken supersymmetries, and the chirality (if d is of the form $4k + 2$, $k \in \mathbb{Z}$) of the supersymmetry generators will not play a major role. So we will not specify these and we will attempt to keep the notation generic.

The gravitino gauge parameter, at momentum k , is the operator

$$(8.23) \quad \mathcal{W}(k, u) = c \Theta_{-1/2} u^\alpha \Sigma_\alpha e^{ik \cdot X},$$

inserted at a Ramond divisor. Here $k^2 = 0$ and u^α is a c -number spinor that obeys the massless Dirac equation in momentum space,

$$(8.24) \quad (\gamma \cdot k)_{\alpha\beta} u^\beta = 0.$$

Also, as in section 5, $\Theta_{-1/2}$ represents the $\beta\gamma$ ground state at picture number $-1/2$ (or in other words in the presence of the Ramond divisor) and Σ_α , which transforms as a spinor under rotations of \mathbb{R}^d , is the fermion vertex operator of [1]. (We need not be concerned with the details of how Σ_α depends on the variables that describe the compact dimensions, if there are any.)

The null state

$$(8.25) \quad \mathcal{V}(k, u) = \{Q_B, \mathcal{W}(k, u)\} = i\tilde{c}ck \cdot \tilde{\partial}X \Theta_{-1/2} u^\alpha \Sigma_\alpha e^{ik \cdot X}$$

is the vertex operator for a longitudinal gravitino. Because this vertex operator is BRST-trivial, the corresponding state will decouple from the S -matrix. As in [75], and as summarized in section 8.1.2, decoupling of this state implies spacetime supersymmetry of the S -matrix, provided the S -matrix exists and the gravitino is exactly massless.

To prove the vanishing of massless tadpoles and the existence of the S -matrix, it seems best to use a formalism in which k is set to 0 from the outset. Precisely at $k = 0$, we have $\mathcal{V}(0, u) = 0$ and thus $\{Q_B, \mathcal{W}(0, u)\} = 0$. This is the first main step in section 8.2.1. However, there is no analog of the second step. We can think of $\mathcal{W}(0, u)$ as being holomorphic, in the sense that it varies holomorphically with the moduli of Σ . (What this means is explained more precisely in the next paragraph.) But we cannot view $\mathcal{W}(0, u)$ as a holomorphic function (or form) on Σ because the only place that it can be inserted on Σ is at a Ramond divisor. The only way to “move” the Ramond divisor at which $\mathcal{W}(0, u)$ is inserted is to vary the moduli of Σ , and among these moduli, there is no distinguished one that controls only the position of a given Ramond divisor. Hence, the only type of “integration” involving $\mathcal{W}(0, u)$ that is possible is integration over the moduli of Σ . Accordingly, we will have to prove spacetime supersymmetry by integration over the moduli space, not just by integration over Σ . It is because of this that it is possible [44, 39, 42], though somewhat unusual, for spacetime supersymmetry to be spontaneously broken in loops though unbroken at tree level. This is impossible for those gauge symmetries for which the second step of section 8.2.1 goes through.

Since $\mathcal{V}(0, u) = 0$, we have $\tilde{\partial}\mathcal{W}(0, u) = 0$. Does not this mean that $\mathcal{W}(0, u)$ can be viewed as a conserved current? Here it helps to remember (see section 5 of [15]) that the precise interpretation of a heterotic string worldsheet Σ is that it is a smooth cs supermanifold of dimension $2|1$ embedded in a product $\Sigma_L \times \Sigma_R$, where Σ_L is an ordinary Riemann surface and Σ_R is a super Riemann surface. The symbol $\tilde{\partial}$ is $\partial_{\tilde{z}}$, where \tilde{z} is a local holomorphic parameter on Σ_L . The equation $\tilde{\partial}\mathcal{W}(0, u) = 0$ reflects the fact that $\mathcal{W}(0, u)$ is the product of the identity operator on Σ_L (which is annihilated by $\tilde{\partial}$) times an object on Σ_R . That object, which we may as well just call $\mathcal{W}(0, u)$, can only be inserted at a Ramond puncture and there is no way to “move” it except by varying the moduli of Σ_R . The statement that $\mathcal{W}(0, u)$ varies holomorphically with the moduli of Σ just means that it varies holomorphically with the moduli of Σ_R and is independent of those of Σ_L .

Once we set $k = 0$, the condition $k^2 = 0$ is satisfied and the Dirac equation (8.24) holds for all u . So, as in the case of translation symmetry, we can drop u from the definition and take the basic object to be

$$(8.26) \quad \mathcal{S}_\alpha = c \Theta_{-1/2} \Sigma_\alpha.$$

This is the worldsheet operator that, in a sense that we will explore, generates spacetime supersymmetry.

On a heterotic string worldsheet Σ of genus g , we will consider a correlation function of \mathcal{S}_α together with n_{NS} physical state vertex operators from the NS sector and n_{R} such operators from the Ramond sector. (The total number of Ramond punctures is $n_{\text{R}} + 1$, so n_{R} must be odd.) It is convenient to set $n = n_{\text{NS}} + n_{\text{R}}$ and denote the physical state vertex operators simply as $\mathcal{V}_1, \dots, \mathcal{V}_n$, without specifying which are of which type.

By the machinery described in sections 3-5, we can define a form $F_{\mathcal{S}_\alpha \mathcal{V}_1 \dots \mathcal{V}_n}(\mathcal{J}, \delta\mathcal{J})$ on the appropriate moduli space. We can view it as a holomorphic form on $\mathcal{M}_L \times \mathcal{M}_R$ (defined in a suitable neighborhood of the “diagonal” in the reduced space), where \mathcal{M}_L and \mathcal{M}_R are the moduli spaces of Σ_L and Σ_R , respectively. However, $F_{\mathcal{S}_\alpha \mathcal{V}_1 \dots \mathcal{V}_n}$ is not an integral form of top degree that could be integrated over the usual integration cycle⁶⁹ $\Gamma \subset \mathcal{M}_L \times \mathcal{M}_R$ to compute a scattering amplitude. The reason for this is that the ghost number of \mathcal{S}_α is lower by 1 than that of a physical state vertex operator; accordingly $F_{\mathcal{S}_\alpha \mathcal{V}_1 \dots \mathcal{V}_n}$ is a form of codimension 1. After all, \mathcal{S}_α is a symmetry generator, not the vertex operator of a physical state. The exterior derivative operator d increases the degree by 1, so it would map $F_{\mathcal{S}_\alpha \mathcal{V}_1 \dots \mathcal{V}_n}$ to a form of top

⁶⁹We define this cycle in the usual way for a heterotic string worldsheet of genus g with n_{NS} NS punctures and $n_{\text{R}} + 1$ Ramond punctures, just as in section 6.6.

degree. But since \mathcal{S}_α and $\mathcal{V}_1, \dots, \mathcal{V}_n$ are all Q_B -invariant, the usual relation $dF_{\mathcal{X}} = -F_{\{Q_B, \mathcal{X}\}}$ tells us in this case that

$$(8.27) \quad dF_{\mathcal{S}_\alpha \mathcal{V}_1 \dots \mathcal{V}_n} = 0.$$

Since this is the case, we have via the supermanifold version of Stokes's theorem

$$(8.28) \quad 0 = \int_\Gamma dF_{\mathcal{S}_\alpha \mathcal{V}_1 \dots \mathcal{V}_n} = \int_{\partial\Gamma} F_{\mathcal{S}_\alpha \mathcal{V}_1 \dots \mathcal{V}_n}.$$

This formula is the Ramond sector analog of (8.20); it involves the use of Stokes's theorem on Γ , not just on Σ . We will study this formula exactly the way that we studied the general formula (7.1) for decoupling of pure gauge modes.

In fact, there are only two differences from that rather general case. First, the left hand side is now 0, since $\mathcal{V}(0, u) = 0$. This means that, rather than a relation saying that a pure gauge mode decouples, we will get, under appropriate conditions, a conservation law, saying that a certain linear combination of scattering amplitudes vanishes. Second, since \mathcal{S}_α carries zero momentum in spacetime, some considerations of section 7 will be modified.

8.2.3. The supersymmetric Ward identity Which components of $\partial\Gamma$ can contribute on the right hand side of eqn. (8.28)? The analysis of this question is very similar to the analysis in section 7.1 of which boundary components can contribute BRST anomalies. In fact, the only relevant difference between the right hand sides of (8.28) and of the general BRST anomaly formula (7.1) is that in (8.28), we have set $k = 0$.

Just as in section 7.1, we only need to consider separating degenerations, at which Σ splits into two components Σ_ℓ and Σ_r , say of genera g_ℓ and g_r , joined at a double point; we label these so that \mathcal{S}_α is contained in Σ_ℓ . And a separating degeneration can only contribute if the momentum-carrying vertex operators $\mathcal{V}_1, \dots, \mathcal{V}_n$ are divided between Σ_ℓ and Σ_r in a way that forces the total momentum flowing between Σ_ℓ and Σ_r to be on-shell.

In our previous analysis, BRST anomalies came from three types of separating degenerations, pictured in figs. 16 and 17 of section 7.3. Setting $k = 0$ adds only one novelty. A fourth type of degeneration can contribute; if Σ_ℓ contains \mathcal{S}_α and precisely one additional vertex operator \mathcal{V}_i , then the total momentum flowing between the two branches is precisely that of \mathcal{V}_i and hence is on-shell.

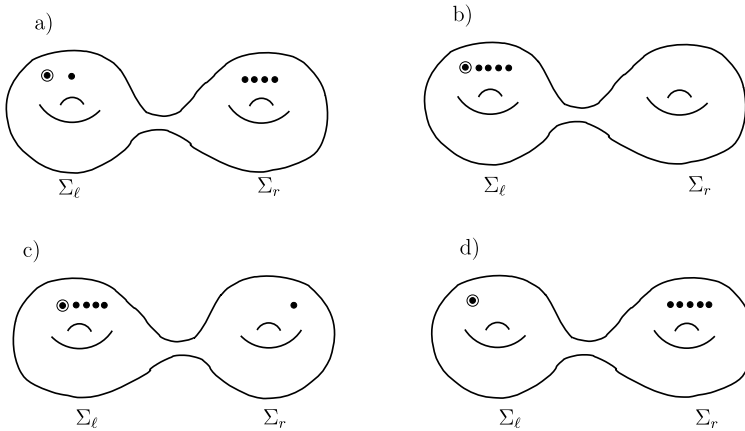


Figure 25: Four types of separating degeneration that potentially can contribute to the supersymmetric Ward identity. The dot enclosed by a circle represents the supercurrent \mathcal{S}_α ; the other dots represent vertex operators of physical states.

For convenience, we display the four potentially relevant degenerations in fig. 25. They can each be analyzed as described in sections 7.3 and 7.7. In each case, Σ_ℓ can be replaced by a conformal or superconformal vertex operator \mathcal{O} that is inserted on Σ_r . \mathcal{O} can be determined by a path integral on Σ_ℓ , as described most precisely in section 7.7.

The physical interpretation of the four interesting types of degeneration is as follows. After replacing Σ_ℓ by an insertion of \mathcal{O} on Σ_r , fig. 25(c) is equal to an on-shell two-point function $\langle \mathcal{O} \mathcal{V}_i \rangle_{g_r}$ (where \mathcal{V}_i is the vertex operator in Σ_R). This is a matrix element of mass renormalization. For reasons explained in section 7.4.3, we restrict ourselves to the S -matrix of massless particles, and we only consider compactifications in which supersymmetry prevents mass renormalization for massless particles. (This condition is satisfied in the ten-dimensional superstring theories and, because of nonrenormalization theorems for the superpotential and the Fayet-Iliopoulos terms, in most supersymmetric compactifications to 4 or more dimensions.) Thus, we will not have to analyze fig. 25(c).

In fig. 25(b), \mathcal{O} carries zero momentum and is the vertex operator of a massless boson of zero momentum. Thus the boundary contribution from fig. 25(b) is proportional to the massless tadpole $\langle \mathcal{O} \rangle_{g_r}$. We will prove in section 8.3 that massless tadpoles vanish in models in which supersymmetry is unbroken in perturbation theory. For the moment, let us simply assume that

this is true and also that we are studying a model in which supersymmetry is unbroken in perturbation theory. Given this, we need not worry about fig. 25(b).

Fig. 25(d) was analyzed in section 7.4 (without specializing to $k = 0$). If $\mathcal{O} \neq 0$, this degeneration will lead to an extra contribution to the supersymmetric Ward identity, and supersymmetry will be spontaneously broken, in the vacuum under study. Since \mathcal{O} in this degeneration is completely determined by \mathcal{S}_α , it is natural to denote it as \mathcal{O}_α . \mathcal{O}_α is the vertex operator for a physical fermion field of spin 1/2 – the Goldstino or Goldstone fermion of spontaneous supersymmetry breaking – at zero momentum. In practice, in any given model, it is possible to explicitly show that $\mathcal{O}_\alpha \neq 0$ for some g_ℓ (for example, $\mathcal{O}_\alpha \neq 0$ for $g_\ell = 1$ in the models studied in [39, 42, 44]), or else to argue that $\mathcal{O}_\alpha = 0$ on general grounds.

For example,⁷⁰ in the ten-dimensional superstring theories, $\mathcal{O}_\alpha = 0$ simply because there is no massless fermion with the same chirality as \mathcal{S}_α and the same transformation under the global symmetries of perturbation theory. To spell this out in more detail, consider first the heterotic string. This theory has a massless neutral fermion of spin 1/2, but its chirality is opposite from that of the spacetime supersymmetry generator \mathcal{S}_α , so it cannot arise as a contribution to \mathcal{O}_α . In Type I and Type IIB superstring theories, there again is no massless neutral fermion with the same chirality as \mathcal{S}_α . In Type IIA superstring theory, there is no such fermion with the same chirality and the same transformation under $(-1)^{F_L}$ as \mathcal{S}_α .⁷¹

Below ten dimensions, we need more detailed arguments. We can use low energy effective field theory to analyze the problem because we have seen in section 7.4 that in the lowest order of perturbation theory in which it occurs, spontaneous supersymmetry breaking means that the effective action is supersymmetric with a correction (7.11) to the transformation law of the Goldstino. Thus if the supersymmetric Ward identity has a contribution in g -loop order from fig. 25(b), this means that in g -loop order, the low energy effective action is spacetime supersymmetric but describes spontaneous breakdown of supersymmetry. One expects that in this case, massless tadpoles will arise in $2g$ -loop order and perturbation theory will break down, but in g -loop

⁷⁰Arguments similar to the following were used in section 7.2.1 in analyzing the implications of supersymmetry for massless tadpoles.

⁷¹ In Type IIA and Type IIB, one must consider both holomorphic and anti-holomorphic supersymmetry generators, but this does not affect what we have just stated. In Type I, there is more to the story when open and/or unoriented string worldsheets are considered; see section 9.

order one will only see an effective action that describes spontaneous supersymmetry breaking. Accordingly, in compactifications to four-dimensions with $\mathcal{N} = 1$ supersymmetry, nonrenormalization of the superpotential by loops [69] and nonrenormalization of Fayet-Iliopoulos D -terms beyond one-loop [44] means that spontaneous supersymmetry breaking by loops is limited to the models in which it is known to occur (and perhaps some of their close relatives in string theories other than the $\text{Spin}(32)/\mathbb{Z}_2$ heterotic string). In higher dimensions or with more unbroken supersymmetry, the constraints are only more severe.

At any rate, we limit ourselves here to models in which general considerations show that spacetime supersymmetry cannot be spontaneously broken by loop effects. This being so, the only boundary contributions that remain are those of fig. 25(a). These are the “new” contributions in which the vertex operators on Σ_ℓ are \mathcal{S}_α and just one of the \mathcal{V}_i . In this case, \mathcal{O} is bilinear in \mathcal{S}_α and in \mathcal{V}_i , so we will denote it as $Q_\alpha(\mathcal{V}_i)$, where Q_α is a transformation between superconformal vertex operators of the NS and R sectors. From its definition, Q_α manifestly commutes with the spacetime momentum so in particular it acts within the space of physical string states at a given mass level. It also clearly has the same spinor quantum numbers as \mathcal{S}_α . We define Q_α to be the spacetime supersymmetry charge.

Obviously, there are precisely n degenerations of the type of fig. 25(a), corresponding to n possible choices for which of the \mathcal{V}_i is on Σ_ℓ . Let us write \mathcal{D}_i for the i^{th} such boundary component. The contribution of \mathcal{D}_i to the right hand side of eqn. (8.28) is an n -particle scattering amplitude with \mathcal{V}_i replaced by $Q_\alpha(\mathcal{V}_i)$. So the vanishing of eqn. (8.28) becomes

$$(8.29) \quad \sum_{i=1}^n \langle \mathcal{V}_1 \dots \mathcal{V}_{i-1} Q_\alpha(\mathcal{V}_i) \mathcal{V}_{i+1} \dots \mathcal{V}_n \rangle = 0.$$

This is our supersymmetric Ward identity, the precise analog of the field theory relation (8.13).

In general, $\partial\Gamma$ has many different algebraic components, corresponding to different ways that the string worldsheet Σ may degenerate. But these components can intersect each other, since Σ can undergo multiple degenerations. Because of this, it is not clear *a priori* that the contributions to the right hand side of (8.28) of individual components of Γ are well-defined. However, the \mathcal{D}_i do not intersect each other (see fig. 6 in section 6.1.2 of [16] for an explanation), so that if they are the only relevant divisors (in other words, if the other configurations sketched in fig. 25 do not contribute), there is no problem in defining their individual contributions.

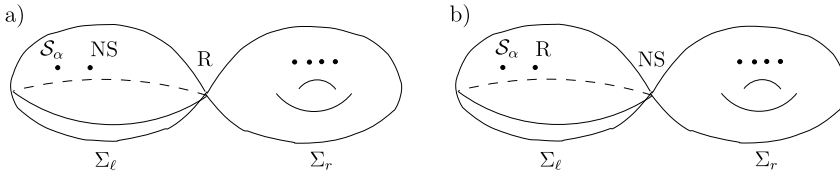


Figure 26: The $g_\ell = 0$ contribution to the action of Q_α on a vertex operator \mathcal{V} . Whether \mathcal{V} is of NS or R type, because of the \mathcal{S}_α insertion in Σ_ℓ , the vertex operator inserted at the double point is of the opposite type, as sketched in (a) and (b). So in either case, Σ_ℓ is a super Riemann surface of genus 0 with one NS puncture and two R punctures, and therefore has no even or odd moduli. However, there is an important difference between the two cases; if \mathcal{V} is of NS type as in (a), the vertex operator at the double point is of R type, so there is a fermionic gluing parameter. If \mathcal{V} is of R type as in (b), the vertex operator at the double point is of NS type and there is no fermionic gluing parameter.

8.2.4. Comparison to the standard description We would now like to compare our definition of the spacetime supersymmetry generators to the standard one [1].

In fact, we can recover the standard definition by just setting $g_\ell = 0$. In this case, Σ_ℓ is simply a three-punctured sphere. The subtleties of section 7.7 are irrelevant, since a three-punctured sphere has no bosonic moduli, and we can omit the $d|q|^2/|q|^2$ terms from eqn. (7.57). To evaluate $Q_\alpha(\mathcal{V})$, for a superconformal vertex operator \mathcal{V} , we need to compute a three-point function on Σ_ℓ (fig. 26). A genus zero three-point function is determined by the operator product expansion, which was also the basis for the definition of the supercharges given in [1], so it should come as no surprise that the two definitions agree.

When we evaluate $Q_\alpha(\mathcal{V})$, because of the \mathcal{S}_α insertion in Σ_ℓ , the string state that flows through the double point connecting Σ_ℓ to Σ_r is of opposite NS or R type from \mathcal{V} , as indicated in the figure, so regardless of whether \mathcal{V} is of NS or R type, Σ_ℓ contains one NS insertion and two R insertions. Accordingly, Σ_ℓ has no even or odd moduli (a super Riemann surface of genus 0 with one NS insertion and two R insertions has no moduli). However, there is an interesting difference between the two cases. If \mathcal{V} is of NS type, then the vertex operator at the double point is of R type and hence in the gluing of Σ_ℓ to Σ_r , there is a fermionic gluing parameter. If \mathcal{V} is of R type, there is no fermionic gluing parameter.

Eqn. (7.57), or its superconformal analog, applies uniformly for all g_ℓ , but to calculate just for $g_\ell = 0$, it is easier to remember where eqn. (7.57) came from. A degeneration in which Σ splits off a genus 0 component Σ_ℓ that contains two vertex operators (in our case \mathcal{S}_α and \mathcal{V}) is the way that in the Deligne-Mumford compactification one describes a collision between two vertex operators. So let us consider an operator product $\mathcal{S}_\alpha(z)\mathcal{V}(0)$. For the degeneration that occurs as $z \rightarrow 0$, the gluing parameter q_{NS} or q_{R} is simply equal to z . So our usual procedure for cutting off the moduli space amounts, in this example, to requiring $|z| \geq \epsilon$ for some small ϵ , and the boundary of the cutoff moduli space is the circle $|z| = \epsilon$. We wish to integrate over this circle. It is convenient to define

$$(8.30) \quad \mathcal{S}_\alpha = c\widehat{S}_\alpha, \quad \widehat{S}_\alpha = \Theta_{-1/2}\Sigma_\alpha.$$

By essentially the usual procedure of passing from a vertex operator to its integrated form, as described in sections 2.5.2 and 4.3, the integral can be accomplished by replacing c with a one-form dz , so that \mathcal{S}_α is replaced by its “integrated form” $dz\widehat{S}_\alpha$. Then we simply integrate over the circle $|z| = \epsilon$:

$$(8.31) \quad \frac{1}{2\pi i} \oint_{|z|=\epsilon} dz \widehat{S}_\alpha \cdot \mathcal{V}.$$

(We have taken the liberty of dividing the integral by $2\pi i$ to agree with standard normalizations.) The integral picks out the residue of the pole in the operator product $\widehat{S}_\alpha(z) \cdot \mathcal{V}(0)$. If \mathcal{V} is in the Ramond sector, this residue is the genus 0 approximation to $Q_\alpha(\mathcal{V})$; if \mathcal{V} is in the NS sector, we still have to integrate over the fermionic gluing parameter, which can be accomplished by multiplying by a picture-changing operator.⁷² What we have arrived at is the definition of the spacetime supercharges Q_α given in [1].

It is interesting that we have a framework to compute higher genus corrections to the supersymmetry generators Q_α , namely the contributions with $g_\ell > 0$. But to make this transparent, it really needs to be combined with an understanding of mass renormalization, presumably via an off-shell approach such as that of [19, 21]. From the point of view of low energy effective field

⁷²One integrates over the fermionic gluing parameter by acting with $\delta(\beta_0)G_0$, while the usual picture-changing recipe is to multiply with the picture-changing operator $\mathcal{Y}(z')$ (defined in eqn. (3.95)), with $z' \rightarrow 0$. In general, these operations coincide modulo $Q_B(\dots)$; on superconformal vertex operators with the usual simple dependence on the ghosts, they coincide precisely. A useful reference on such matters is [11].

theory, the representations of the supersymmetry algebra are completely determined by the particle masses, spins, and possibly the central charges (in string theory compactifications in which these are present in the supersymmetry algebra). One expects therefore that renormalizations of the masses and possibly the central charges completely determine the corrections to the Q_α .

Our definition of the Q_α is actually a very close cousin of what is briefly explained in [10]. The main difference is that we use a superconformally invariant formalism, while [10] is based on a more general formalism along the lines of section 2.4.4.

8.2.5. More on momentum conservation The definition of the spacetime supercharge in section 8.2.2 was notably more sophisticated than the treatment of bosonic symmetries such as momentum and winding in section 8.2. It does not seem that the definition of spacetime supersymmetry can be reduced to the more elementary ideas used in section 8.2, but there is no problem to go in the opposite direction and describe bosonic gauge symmetries in the same language that we used for supersymmetry.

For example, let us consider the momentum and winding symmetries of the bosonic string, either in \mathbb{R}^{26} or toroidally compactified. The gauge parameters for the sum and difference of the metric and B -field are $\mathcal{W} = c\varepsilon \cdot \partial X \exp(ip \cdot X)$, $\widetilde{\mathcal{W}} = \widetilde{c}\varepsilon \cdot \widetilde{\partial} X \exp(ip \cdot X)$, with $p^2 = p \cdot \varepsilon = 0$. Taking $p = 0$, ε becomes arbitrary and we define $\mathcal{P}^I = c\partial X^I$, $\widetilde{\mathcal{P}}^I = \widetilde{c}\widetilde{\partial} X^I$. These Q_B -invariant operators can be used to define conservation laws in the same way that we did for supersymmetry. In the case of \mathcal{P}^I , for example, we use the correlation function $\langle \mathcal{P}^I \mathcal{V}_1, \dots, \mathcal{V}_n \rangle$ to define a differential form $F_{\mathcal{P}^I, \mathcal{V}_1, \dots, \mathcal{V}_n}$ on $\widehat{\mathcal{M}}_{g, n+1}$. It is closed and of codimension 1, so as usual

$$(8.32) \quad 0 = \int_{\widehat{\mathcal{M}}_{g, n+1}} dF_{\mathcal{P}^I, \mathcal{V}_1, \dots, \mathcal{V}_n} = \int_{\partial \widehat{\mathcal{M}}_{g, n+1}} F_{\mathcal{P}^I, \mathcal{V}_1, \dots, \mathcal{V}_n}.$$

Analysis of the surface terms here leads to a conservation law, which is the same one that we deduced in section 8.2 using the conserved worldsheet current $J^I = \partial X^I$.

For the heterotic string, one defines similarly $\mathcal{P}^I = c\delta(\gamma)D_\theta X^I$, $\widetilde{\mathcal{P}}^I = \widetilde{c}\widetilde{\partial} X^I$. The Type II analog is evident. Again, the same formalism can be used to deduce momentum and winding conservation. In the language of [1], the operator \mathcal{P}^I is related by picture-changing to the conserved worldsheet current $D_\theta X^I$ that is used in the more elementary explanation. (To be more precise, \mathcal{P}^I is a picture-changed version of $\partial_z X^I = \int d\theta D_\theta X^I$.) The advantage of the more abstract approach to momentum and winding symmetry

using the operator \mathcal{P}^I rather than the current $D_\theta X^I$ is that it enables one to treat momentum conservation and spacetime supersymmetry in the same framework. When we study the spacetime supersymmetry algebra in section 8.2.6, it will be hard to avoid treating the different spacetime symmetries in the same framework.

Though it is not clear that this is useful, one can also place in the same framework the conservation laws associated to massless spin 1 gauge fields. We will just mention a few illustrative examples. In the fermionic description of the Spin(32)/ \mathbb{Z}_2 heterotic string, one describes the left-moving current algebra via fermionic primary fields Λ_a , $a = 1, \dots, 32$ of dimension $(1/2, 0)$. The associated current operators are $\mathcal{R}_{ab} = \tilde{c}\Lambda_a\Lambda_b$ and can be used in an argument along the lines of eqn. (8.32) to establish the relevant global symmetries. In Type I superstring theory, the usual massless gauge fields are open-string modes. Including the Chan-Paton factors (we usually do not make them explicit in this paper), the corresponding vertex operators are $\mathcal{V}_T = c\delta(\gamma)\varepsilon \cdot \partial X \exp(ip \cdot X)T$, where T is a group generator acting on the Chan-Paton factors. The corresponding gauge parameter (see eqn. (C.3)) is $\mathcal{W}_T = c\delta'(\gamma)\exp(ip \cdot X)T$. Setting $p = 0$, we define the current operator $\mathcal{R}_T = c\delta'(\gamma)T$, and one can use this operator, inserted on $\partial\Sigma$, to prove the global symmetries associated to Chan-Paton gauge-invariance by following the procedure of eqn. (8.32). Of course, what we have just described are rather long routes to establish symmetries that actually are manifest in the perturbative formalism.

The fields of perturbative string theory that are usually called gauge fields but that cannot be put in this framework are the r -form gauge-fields of the R-R sector of Type II superstring theory (possibly enriched with D-branes and/or orientifold planes). They are not really gauge fields in the relevant sense. There are no R sector gauge parameters at the massless level of perturbative string theory (the first R sector gauge parameter is described in eqn. C.9). Moreover, the elementary string states are all neutral under the R-R gauge symmetries, so R-R gauge symmetry does not lead to a non-trivial conservation law in scattering of these states. R-R gauge invariance comes into play when one constrains the D-branes and orientifold planes via R-R tadpole cancellation; see section 9.4.

8.2.6. The supersymmetry algebra Since we have obtained the standard supersymmetry generators, at least in the $g_\ell = 0$ approximation, it is fairly obvious that we must get the standard supersymmetry algebra, at least in that approximation. It is interesting to see how to compute this algebra, in the present formulation. First let us recall how we would do this

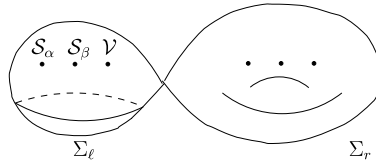


Figure 27: As a preliminary to evaluating the anticommutator $\{Q_\alpha, Q_\beta\}$, we restrict the closed integral form $F_{\mathcal{S}_\alpha \mathcal{S}_\beta \mathcal{V}_1 \dots \mathcal{V}_n}(\mathcal{J}, \delta \mathcal{J})$ to a locus $|q| = \epsilon$, where the divisor indicated here is defined by $q = 0$. After this restriction, $F_{\mathcal{S}_\alpha \mathcal{S}_\beta \mathcal{V}_1 \dots \mathcal{V}_n}(\mathcal{J}, \delta \mathcal{J})$ still has codimension 1.

at non-zero momentum. For this, we can follow a procedure described in section 8.1.2. Given any n non-soft string states represented by vertex operators $\mathcal{V}_1, \dots, \mathcal{V}_n$, we would add two vertex operators for soft gravitinos, compute a g -loop contribution by integrating a suitable integral form on the appropriate integration cycle Γ , and deduce the supersymmetry algebra by studying the limit of this integral as the gravitino momentum goes to zero.

Instead here we want to describe an analogous procedure with the soft gravitino vertex operators replaced by supersymmetry generators \mathcal{S}_α and \mathcal{S}_β , which carry zero momentum. The usual procedure will extract from the world-sheet path integral a closed integral form $F_{\mathcal{S}_\alpha \mathcal{S}_\beta \mathcal{V}_1 \dots \mathcal{V}_n}(\mathcal{J}, \delta \mathcal{J})$ on the same cycle Γ as before, but now, as two physical state vertex operators have been replaced by \mathcal{S}_α or \mathcal{S}_β (whose ghost number is smaller by 1), $F(\mathcal{J}, \delta \mathcal{J})$ is a form of codimension 2. So it certainly cannot be integrated over Γ .

We can reduce the gap in dimension from 2 to 1 by restricting $F(\mathcal{J}, \delta \mathcal{J})$ from Γ to one of its boundary components \mathcal{B} . (This is done in the usual way by a relation such as $|q| = \epsilon$, where q is a complex gluing parameter; such a relation reduces the real dimension by 1.) We pick \mathcal{B} to correspond (fig. 27) to a separating degeneration in which Σ splits into two components Σ_ℓ and Σ_r , where Σ_ℓ contains precisely \mathcal{S}_α , \mathcal{S}_β , and one more vertex operator that we will call simply \mathcal{V} ; all other vertex operators are in Σ_r .

We still cannot integrate $F(\mathcal{J}, \delta \mathcal{J})$ over \mathcal{B} since even after restriction to \mathcal{B} , it is of codimension 1. However, \mathcal{B} itself has boundaries associated to further degenerations. Since $dF(\mathcal{J}, \delta \mathcal{J}) = 0$, we have a relation just analogous to (8.28) but with Γ replaced by \mathcal{B} :

$$(8.33) \quad 0 = \int_{\mathcal{B}} dF(\mathcal{J}, \delta \mathcal{J}) = \int_{\partial \mathcal{B}} F(\mathcal{J}, \delta \mathcal{J}).$$

Among the boundary component of \mathcal{B} , those that arise by degeneration of Σ_r are not relevant, since all vertex operators on Σ_r are those of physical states;

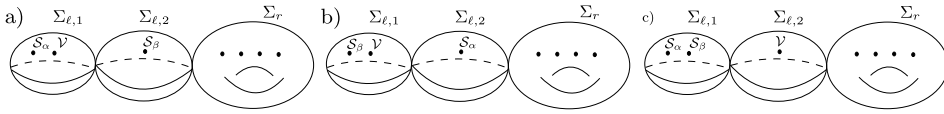


Figure 28: Stokes’s theorem applied to the configuration of fig. 27 says that the sum of these three contributions vanishes. This yields the spacetime supersymmetry algebra.

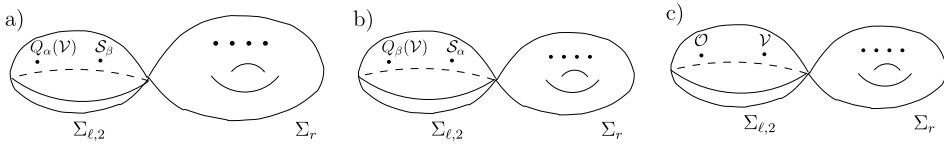


Figure 29: As described in the text, these three pictures represent a step in evaluating the contributions from the three pictures in fig. 28. The operator labeled \mathcal{O} in (c) is bilinear in \mathcal{S}_α and \mathcal{S}_β .

the path integral on Σ_r is already generating a form of top degree. We only care about degenerations of Σ_ℓ .

To keep things simple,⁷³ we will only analyze the genus 0 contribution to the supersymmetry algebra. If Σ_ℓ has genus 0, it can degenerate in precisely three ways, sketched in fig. 28. In each of these cases, Σ_ℓ splits into two components $\Sigma_{\ell,1}$ and $\Sigma_{\ell,2}$ where only the second intersects Σ_r . We can compute the corresponding amplitudes by iterating ideas we have already explained. In each case, we can replace $\Sigma_{\ell,1}$ by some operator insertion on $\Sigma_{\ell,2}$. In fig. 28(a), this operator is by definition $Q_\alpha(\mathcal{V})$. Replacing $\Sigma_{\ell,1}$ by an insertion of $Q_\alpha(\mathcal{V})$ on $\Sigma_{\ell,2}$, we arrive at fig. 29(a), where now we can replace $\Sigma_{\ell,2}$ by an insertion of $Q_\beta(Q_\alpha(\mathcal{V}))$ on Σ_r . A similar analysis of fig. 28(b), leads via fig. 29(b) to an insertion of $Q_\alpha(Q_\beta(\mathcal{V}))$ on Σ_r . Finally, in fig. 28(c), we can replace $\Sigma_{\ell,1}$ by an insertion of an operator that is bilinear in \mathcal{S}_α and \mathcal{S}_β ; we call this operator $\mathcal{O}_{\mathcal{S}_\alpha, \mathcal{S}_\beta}$ or simply \mathcal{O} . And then from fig. 29(c), we can replace $\Sigma_{\ell,2}$ by an insertion on Σ_r of an operator bilinear in \mathcal{O} and \mathcal{V} , which we call $\mathcal{O}(\mathcal{V})$.

⁷³And also because of not fully understanding the exceptional zero-momentum contributions that will enter momentarily. One expects no higher loop corrections to the supersymmetry algebra, probably because there are no contributions in which the genus of $\Sigma_{\ell,1}$ in fig. 28(c) is positive. To show this, one needs to analyze the exceptional contributions.

The identity (8.33) becomes

$$(8.34) \quad Q_\alpha(Q_\beta(\mathcal{V})) + Q_\beta(Q_\alpha(\mathcal{V})) + \mathcal{O}(\mathcal{V}) = 0.$$

Hence

$$(8.35) \quad \mathcal{O} = -\{Q_\alpha, Q_\beta\},$$

and we can determine the supersymmetry algebra by a path integral on $\Sigma_{\ell,1}$ in fig. 28(c).

In this figure, the momentum flowing between $\Sigma_{\ell,1}$ and $\Sigma_{\ell,2}$ is zero, so to understand the path integral on $\Sigma_{\ell,1}$, we need the exceptional zero-momentum contribution whose origin we have first seen in formulas such as (6.90) and (6.98) for bosonic strings. The relevant contribution is

$$(8.36) \quad \tilde{c} \tilde{\partial} \tilde{c} \tilde{\partial}^2 \tilde{c} c \delta(\gamma) D_\theta X^I \otimes c \delta(\gamma) D_\theta X_I + \dots$$

where we omit similar terms (with the two factors of the tensor product exchanged, or holomorphic factors exchanged with analogous antiholomorphic ones) that do not contribute to evaluating the supersymmetry algebra. In eqn. (8.36), we see the operator $\mathcal{P}^I = c \delta(\gamma) D_\theta X^I$ whose relation to energy-momentum conservation was described in section 8.2.5.

However, as in section 8.2.4, it is easier, for $\Sigma_{\ell,1}$ of genus 0, to recognize that what we are trying to calculate is simply a term in the operator product expansion. The same reasoning that led to (8.31) gives

$$(8.37) \quad \mathcal{O}_{\mathcal{S}_\alpha, \mathcal{S}_\beta} = \frac{1}{2\pi i} \oint_{|z|=\epsilon} dz \hat{S}_\alpha(z) \mathcal{S}_\beta(0).$$

This is the standard answer, in the sense that the supersymmetry algebra is computed in [1, 2] from just this operator product. For superstrings in \mathbb{R}^{10} , the right hand side of (8.37) can be evaluated to give $\gamma_{I\alpha\beta} \mathcal{P}^I = c \delta(\gamma) \Gamma_{I\alpha\beta} D_\theta X^I$. (Not coincidentally, the same operator $c \delta(\gamma) D_\theta X^I$ appears in (8.36), giving another route to the same result for $\{Q_\alpha, Q_\beta\}$.) Here, as we explained in section 8.2.5, the operator \mathcal{P}^I is related to energy-momentum conservation in precisely the same way that \mathcal{S}_α is related to spacetime supersymmetry. So we have arrived at the usual spacetime supersymmetry algebra.

In the language of [1], the operator $\mathcal{P}^I = c \delta(\gamma) D_\theta X^I$ is related by picture-changing to the holomorphic current that generates translation symmetry (or after toroidal compactification, a linear combination of translation and winding symmetry). The picture-changing is a way to describe the integration over the fermionic gluing parameter in fig. 29(c).

A brief explanation of a calculation of the supersymmetry algebra somewhat similar to what we have explained but in a rather different language can be found in section 6.3 of [10].

For superstring theory compactified to $\mathbb{R}^d \times Z$ with $d < 10$ for some space Z , the operator product in eqn. (8.37) may involve, in addition to the \mathcal{P}^I , also operators associated to symmetries of Z . In that case, generators of these symmetries appear as central charges in the supersymmetry algebra. If not prevented by arguments of holomorphy or other considerations of low energy effective field theory, there may be loop corrections to the central charges, which would appear as loop corrections to the supersymmetry algebra. This may be most likely for $d \leq 3$, where constraints of low energy field theory can be less powerful.

8.3. Vanishing of massless tadpoles

Finally we come to the question of proving that perturbative massless tadpoles vanish in those supersymmetric compactifications in which supersymmetry is not spontaneously broken in perturbation theory.

The one fact that we need to know is that if \mathcal{V}_ϕ is the vertex operator of a massless neutral spin-zero field ϕ at zero momentum, then there is always a fermion vertex operator $\mathcal{V}_{\psi_\alpha}$ of a zero-momentum neutral fermion field ψ_α that satisfies a relation

$$(8.38) \quad \mathcal{V}_\phi = \sum_\alpha Q_\alpha(\mathcal{V}_{\psi_\alpha}).$$

One can show this by explicitly exhibiting $\mathcal{V}_{\psi_\alpha}$ in each case. For example, in ten-dimensional heterotic string theory, the only \mathcal{V}_ϕ is

$$(8.39) \quad \mathcal{V}_\phi = \tilde{c}c\delta(\gamma)\tilde{\partial}X_I DX^I,$$

and the corresponding $\mathcal{V}_{\psi_\alpha}$ is

$$(8.40) \quad \mathcal{V}_{\psi_\alpha} = \Gamma_I^{\alpha\beta}\tilde{c}c\tilde{\partial}X^I\Theta_{-1/2}\Sigma_\beta.$$

(The Γ_I are spacetime gamma matrices and Σ_β is a spin field of the matter system.)

Rather than exhibit such formulas in all the cases, we prefer to observe that the result actually follows from the general form of the low energy supersymmetry transformations. Under spacetime supersymmetry, ϕ always transforms into a fermion field, $Q_\alpha(\phi) = \psi_\alpha$, and ψ_α is non-zero at zero momentum.

On the other hand, at zero momentum we have $Q_\alpha(\psi_\beta) = 0$, assuming that supersymmetry is not spontaneously broken.⁷⁴ Now consider turning on ϕ and ψ_α at zero momentum, perturbing the worldsheet action by $\phi\mathcal{V}_\phi + \psi_\alpha\mathcal{V}_{\psi_\alpha}$. For this class of perturbations to be closed under spacetime supersymmetry, since ϕ transforms into ψ_α while ψ_α is invariant, $\mathcal{V}_{\psi_\alpha}$ must transform into \mathcal{V}_ϕ , giving a relation (8.38).

Supersymmetric invariance of $\phi\mathcal{V}_\phi + \psi_\alpha\mathcal{V}_{\psi_\alpha}$ gives one more interesting condition: at zero momentum, \mathcal{V}_ϕ must be invariant under supersymmetry. This simply means that the perturbation by \mathcal{V}_ϕ preserves spacetime supersymmetry, so the expectation value of ϕ is a modulus of the supersymmetric theory, at least to first order.

Going back to eqns. (8.38) and (8.40), there is actually an interesting detail here, which echoes a comment in section 7.2.2. Just as \mathcal{V}_ϕ is the vertex operator for a linear combination of the dilaton and the trace of the metric, so $\mathcal{V}_{\psi_\alpha}$ is the vertex operator for a linear combination of gravitino and dilatino fields. (By the dilatino, we mean the spin 1/2 field in the ten-dimensional $\mathcal{N} = 1$ supergravity multiplet.) At non-zero momentum, there are separate superconformal vertex operators for dilatons and gravitons and likewise for dilatinos and gravitinos. At zero momentum, in a superconformal formalism, there is only one \mathcal{V}_ϕ , so only one linear combination of the tadpoles of the dilaton and the trace of the metric is a potential obstruction to the validity of perturbation theory, and there is only one $\mathcal{V}_{\psi_\alpha}$ that we can use to prove the vanishing of this one obstruction. In a more general formalism as summarized in section 2.4.4, we could potentially encounter separate tadpoles for the dilaton and the trace of the graviton, and we would have two separate $\mathcal{V}_{\psi_\alpha}$'s to deal with these two tadpoles.

Given the relation (8.38), the proof that massless tadpoles vanish in g -loop order is very short. We simply consider the two-point function

$$(8.41) \quad \sum_{\alpha} \langle \mathcal{S}_{\alpha} \mathcal{V}_{\psi_{\alpha}} \rangle$$

in genus g . As usual, we use this correlator to define a closed integral form

⁷⁴In supersymmetric field theories with unbroken supersymmetry (which excludes a constant term in the transformation law of a fermion field), massless neutral spin 1/2 fields transform under supersymmetry into derivatives of scalars and also into the field strengths of abelian gauge fields. These vanish at zero momentum. Note as well that the derivative of a scalar and the field strength of a gauge field are not Lorentz scalars, so for that reason alone, they could not contribute to $\sum_{\alpha} Q_{\alpha}(\mathcal{V}_{\psi_{\alpha}})$.

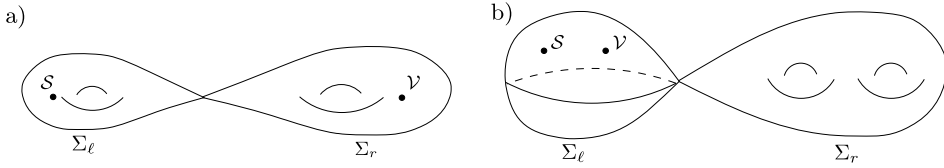


Figure 30: With only two vertex operators, only two types of separating degeneration are relevant to the supersymmetric Ward identity. They are related to (a) spontaneous supersymmetry breaking or (b) the massless tadpole.

$F_{S,V}$ on the relevant integration cycle Γ , leading to the usual relation

$$(8.42) \quad 0 = \int_{\Gamma} dF_{S V} = \int_{\partial\Gamma} F_{S V}.$$

With only two vertex operators, there are only two types of boundary contribution that may contribute to the relation (8.42); the two vertex operators are contained in opposite components or the same component of Σ (fig. 30). The two cases correspond respectively to spontaneous supersymmetry breaking (fig. 30(a)) and massless tadpoles (fig. 30(b)).

In fig. 30(a), the two branches Σ_{ℓ} and Σ_r both have positive genus, or else this degeneration does not arise in the Deligne-Mumford compactification. Since $g_{\ell} + g_r = g$, it follows that $g_{\ell}, g_r \leq g - 1$. In fig. 30(b), $g_r > 0$ but g_{ℓ} may vanish, so g_r can equal g .

A very short argument now shows the vanishing of massless tadpoles. Inductively, suppose that we know that massless tadpoles vanish and spacetime supersymmetry is unbroken up to and including genus $g - 1$. Then almost all boundary contributions to the relation (8.42) vanish. The only contribution that is not ruled out by the inductive hypothesis comes from the degeneration of fig. 30(b) with $g_{\ell} = 0, g_r = g$. Using (8.38), this is simply the genus g tadpole $\langle \mathcal{V}_{\phi} \rangle_g$. So that tadpole vanishes.

Once we know that the massless tadpoles vanish in genus g , the g -loop amplitudes make sense and the arguments of section 8.2.3 show that these amplitudes have spacetime supersymmetry, in general possibly with spontaneous supersymmetry breaking in g -loop order. If (as in most cases) general arguments show that spacetime supersymmetry is not spontaneously broken at g -loop order, then the inductive step is complete and we can repeat the argument for genus $g + 1$.

Just one more comment is perhaps called for here. It was crucial for the inductive argument that if spacetime supersymmetry is unbroken up to and

including g -loop order, there are no massless tadpoles below $g + 1$ -loop order. We deduced this from general properties of the Deligne–Mumford compactification. But in fact, low energy effective field theory implies that if supersymmetry is unbroken up to g -loop order, then massless tadpoles appear only in order $2g$.

9. Open and/or unoriented strings

9.1. Overview

In section 8, we took the heterotic string as the basic example of a string theory with spacetime supersymmetry. In that string theory, the supersymmetry generator \mathcal{S}_α is a right-moving or holomorphic field, and the distinction between NS and R sectors only exists for right-moving and not for left-moving degrees of freedom on the string worldsheet.

The analysis of section 8 extends almost at once to oriented closed Type II superstring theory. In this theory, there are two types of spin 0 field, arising either in the NS-NS sector or the R-R sector. In the superconformal framework, expectation values of R-R fields cannot be conveniently incorporated, so they are assumed to vanish; as a result, there is a global symmetry $(-1)^{FL}$ that distinguishes between the left-moving NS and R sectors and ensures the vanishing of R-R tadpoles. So only NS-NS tadpoles need to be considered.

In oriented closed Type II superstring theory, there are separate spaces \mathcal{S}_L and \mathcal{S}_R of left- and right-moving supersymmetry generators. (In ten dimensions and in the obvious compactifications, \mathcal{S}_L and \mathcal{S}_R have the same dimension, but this is not true in general; for example [79], it is not always true for asymmetric orbifolds.) The left- and right-moving symmetries can each be treated exactly as we treated the supersymmetry generators of the heterotic string. For example, every left- or right-moving supersymmetry generator $\mathcal{S}'_\alpha \in \mathcal{S}_L$ or $\mathcal{S}''_\beta \in \mathcal{S}_R$ is associated to a spacetime supersymmetry generator Q'_α or Q''_β , each defined exactly by the procedure of section 8.2.3. Q'_α exchanges the left-moving NS and R sectors, and Q''_β does the same for right-moving ones. Q'_α anticommutes with Q''_β , since there is no short distance singularity between \mathcal{S}'_α and \mathcal{S}''_β . If $\mathcal{V}_{\text{NS,NS}}$ is a spin-zero and momentum zero superconformal vertex operator from the NS-NS sector, then eqn. (8.38) bifurcates into two separate formulas, one involving Q'_α and one involving Q''_β :

$$(9.1) \quad \mathcal{V}_{\text{NS,NS}} = \sum_{\alpha} \{Q'_\alpha, \mathcal{V}_{\text{R,NS}}^\alpha\} = \sum_{\beta} \{Q''_\beta, \mathcal{V}_{\text{NS,R}}^\beta\},$$

where $\mathcal{V}_{\text{R,NS}}^\alpha$ and $\mathcal{V}_{\text{NS,R}}^\beta$ are vertex operators of the indicated types. Each of these relations holds for the same reasons as (8.38), and either one of them can be used, independent of the other, as input for the proof of vanishing of massless tadpoles given in section 8.3.

There is much more to say if we generalize Type II superstring theory to include open and/or unoriented string worldsheets, by including either orientifold planes or D-branes. In this paper,⁷⁵ orientifold planes and D-branes are assumed to preserve the full d -dimensional Poincaré symmetry of an underlying oriented closed-string compactification to $\mathbb{R}^d \times Z$, for some Z (here $d \geq 2$). In the presence of D-branes and/or orientifolds, the symmetry $(-1)^{F_L}$ is lost and no longer prevents R-R tadpoles. Such tadpoles are possible and play an important role. Also, separate conservation of Q'_α and Q''_β is not possible when the worldsheet can be open and/or unoriented. One must form linear combinations, for reasons we explain momentarily. This ends up leading to a subtle interplay between NS-NS and R-R fields.

9.1.1. Orientifolds and D-branes In practice, the string worldsheet Σ can be unorientable in an orientifold theory. Such a theory is constructed starting from an underlying Type II superstring theory (possibly with D-branes included, as discussed shortly), by projecting the string states onto states that are invariant under an orientifold projection Ω . Here Ω is defined by combining a diffeomorphism of the string that reverses its orientation with some symmetry of the target spacetime (or more generally some σ -model symmetry). In the orientifold theory, in traversing a loop in Σ around which its orientation is reversed, every left-moving field and in particular every left-moving supersymmetry generator is exchanged with some right-moving one. This exchange corresponds to an invertible map $\phi : \mathcal{S}_L \rightarrow \mathcal{S}_R$; in particular, in this situation, \mathcal{S}_L and \mathcal{S}_R always have the same dimension. Unorientability means that we cannot distinguish the spacetime supersymmetry associated to $\mathcal{S}'_\alpha \in \mathcal{S}_L$ from that associated to $\phi(\mathcal{S}'_\alpha) \in \mathcal{S}_R$. So for unoriented superstrings, the unbroken supersymmetries are associated to the sums of left- and right-moving supercharges, $\mathcal{S}'_\alpha + \phi(\mathcal{S}'_\alpha)$. A standard way to express this reasoning is to say that these sums are Ω -invariant and so make sense after projecting to the orientifold theory.

If Σ can have a boundary, which is the case when D-branes are present, then the purely left- and/or right-moving bulk supersymmetries are broken;

⁷⁵More general cases can be studied similarly. In all cases, one aims to prove that the low energy behavior in string theory is consistent with what one would expect from the appropriate low energy effective field theory.

any unbroken supersymmetries are linear combinations of left- and right-moving ones. This is usually proved by considering the conditions for a worldsheet supercurrent to be conserved in the presence of a boundary; we give a slightly more precise explanation in section 9.2.3. We are mainly interested in the case that there are some unbroken supersymmetries, since unbroken supersymmetry will be an ingredient in proving the vanishing of massless tadpoles.

A prototype is Type I superstring theory. To construct this theory, one starts with Type IIB superstring theory, in which the spaces \mathcal{S}_L and \mathcal{S}_R are isomorphic; they are 16-dimensional, and transform under the ten-dimensional Lorentz group $SO(1,9)$ as a spinor of, say, positive chirality. A spacetime-filling orientifold plane and associated D-branes are then introduced in a way that preserves half of the supersymmetry. We can consider ϕ to be 1. The unbroken supersymmetries are $Q_\alpha = Q'_\alpha + Q''_\alpha$, where α is a positive chirality spinor index. They correspond to the supersymmetry generators $\mathcal{S}_\alpha = \mathcal{S}'_\alpha + \mathcal{S}''_\alpha$.

9.1.2. What we will learn After establishing some fundamentals in section 9.2, we analyze tadpoles in open and/or unoriented superstring theory in section 9.3. What emerges is more complicated than for closed oriented superstring theory: instead of a proof that NS-NS tadpoles vanish, we get a formula that in a sense relates NS-NS tadpoles to R-R tadpoles. The precise statement is rather delicate and leads to a severe constraint on R-R tadpoles: they arise only when the worldsheet Σ has the topology of a disc or of $\mathbb{R}P^2$. Since only finitely many topologies can contribute to R-R tadpoles, it is possible to completely evaluate those tadpoles in a given superstring theory. Moreover, the disc and $\mathbb{R}P^2$ both have Euler characteristic 1 and hence contribute in the same order of superstring perturbation theory – the lowest order in which open and/or unoriented worldsheets appear at all. Accordingly it is possible for their contributions to cancel. A celebrated calculation [82] that triggered what is sometimes called the first superstring revolution showed, in effect, that this cancellation occurs in Type I superstring theory precisely if the Chan-Paton gauge group of the open strings is $SO(32)$. More generally, R-R tadpole cancellation is a standard and important ingredient in constructing superstring models with orientifolds and/or D-branes.

From general considerations such as those of section 7.2, one might expect that R-R tadpoles would be associated to infrared divergences, just like other massless tadpoles. There is a certain important sense in which this is not true: R-R tadpoles lead to gauge and gravitational anomalies in spacetime, not directly to infrared divergences. It is true [83] that infrared divergences

cancel in (for example) Type I superstring theory precisely when R-R tadpoles cancel, but the reason for this is that spacetime supersymmetry relates R-R tadpoles to NS-NS tadpoles, which in turn are associated to infrared divergences. Our explanation of these facts relies on the fermionic gluing parameter associated to a Ramond degeneration; for an earlier approach, see [84]. The relation of anomaly cancellation (cancellation of R-R tadpoles) to the cancellation of infrared divergences (cancellation of NS-NS tadpoles) is important in superstring model-building.

9.2. Fundamentals

9.2.1. Geometry The Deligne-Mumford compactification of the moduli space of oriented closed Riemann surfaces Σ is a manifold without boundary.⁷⁶ Compactification is achieved by adjoining to the ordinary moduli space \mathcal{M} a divisor \mathcal{D} at infinity; \mathcal{D} is a union of irreducible components \mathcal{D}_α , each defined by the vanishing of an appropriate complex gluing parameter q_α . As usual, we write $\widehat{\mathcal{M}}$ for the compactification.

When we speak loosely of the “boundary” of $\widehat{\mathcal{M}}$, we really mean the following. Because the integration measures of interest are typically singular at $q_\alpha = 0$, we introduce an infrared cutoff by requiring $|q_\alpha| \geq \epsilon$ for some small positive ϵ . (As usual, the definition of $|q_\alpha|$ depends on a suitable choice of hermitian metric.) This inequality gives a cutoff version of $\widehat{\mathcal{M}}$ that we might call $\widehat{\mathcal{M}}_\epsilon$. It is really $\widehat{\mathcal{M}}_\epsilon$ that has a boundary, namely the boundary at $|q_\alpha| = \epsilon$. The “boundary” contributions to Ward identities of closed oriented string theories, as studied in sections 7 and 8, are always boundary contributions in this sense. They are more precisely contributions from the divisors at infinity in $\widehat{\mathcal{M}}$. For short, we call them the contributions of virtual boundaries.

For open and/or unoriented Riemann surfaces, things are different; the relevant moduli spaces have boundaries in the naive sense. Technically, the details are a little easier to describe for ordinary Riemann surfaces, so we consider that case first. (What follows is a summary of matters explained more fully in section 7.4 of [16].) For open and/or unoriented Riemann surfaces, $\widehat{\mathcal{M}}$ has three types of boundary component. Two of them arise from closed-string degenerations in which, exceptionally, the gluing parameter q is real and nonnegative rather than being, as usual, a complex parameter. This happens (fig. 31) when Σ decomposes to a union of components Σ_ℓ and Σ_r , with one of them, say Σ_r , being a disc \mathbb{D} with one puncture or a copy of \mathbb{RP}^2 with

⁷⁶To be more precise, it is an orbifold or in fancier language a stack without boundary, rather than a manifold. This refinement will not be important and we will loosely refer to the moduli space as a manifold.

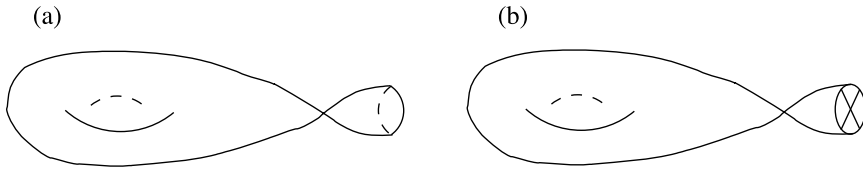


Figure 31: These closed-string degenerations, in which the right-hand component Σ_r is a disc, as in (a), or a copy of $\mathbb{R}P^2$, as in (b), are the exceptional cases in which the gluing parameter q is real and positive. Typically, for closed-string degenerations, gluing depends on a complex parameter. In these particular cases, the argument of the gluing parameter can be absorbed in a rotation of Σ_r around its intersection with Σ_ℓ .

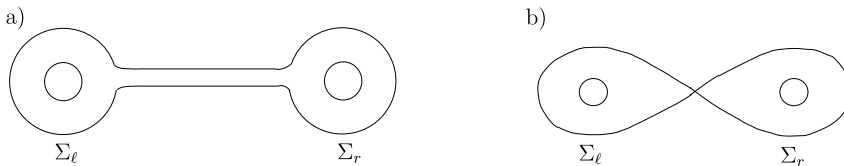


Figure 32: (a) An open-string worldsheet with a long strip connecting two branches Σ_ℓ and Σ_r . (b) Compactification is achieved by allowing an open-string degeneration, shown here, to represent the limit that the length of the strip becomes infinite.

one puncture. Precisely in those two cases, Σ_r has a $U(1)$ symmetry group (consisting of rotations around the puncture) that can be used to eliminate the argument of the gluing parameter; accordingly, one can take q to be real and nonnegative. The condition $q = 0$ then defines a boundary component of the compactified moduli space. The third type of boundary component arises from what one might call an open-string degeneration, in which (fig. 32) Σ is a union of two components joined by a long strip rather than a long tube. In other words, this degeneration is associated to an on-shell open string rather than an on-shell closed string. If s is the length of the strip, one defines the gluing parameter $q = e^{-s}$; it is real and positive, and compactification of the moduli space is achieved by including the degeneration at $q = 0$. Again, the condition $q = 0$ defines a boundary component of the compactified moduli space.

All this is for open and/or unoriented bosonic strings. The supersymmetric case involves some further details. For open and/or unoriented superstring

worldsheets Σ , it seems that there is no natural definition of a moduli space, but instead one can define an integration cycle Γ suitable for superstring perturbation theory. (This has been discussed in sections 3.3.2 and 6.6.3, as well as section 7.4 of [16].) The compactification $\widehat{\Gamma}$ is a smooth supermanifold (or more precisely superorbifold) with boundary. Its boundaries again correspond⁷⁷ to the degenerations of figs. 31 and 32. In superstring theory, each of these degenerations is of two types. In fig. 31, the closed-string state that propagates through the double point may be of either NS-NS or R-R type. In fig. 32, the open-string state that propagates through the double point may be of either NS or R type.

In each of these cases, the gluing parameter q is real and nonnegative modulo the odd variables. Moreover, it is defined up to

$$(9.2) \quad q \rightarrow e^f q,$$

where f is real modulo the odd variables. The existence of such a function q , vanishing on the boundary, is part of the definition of a smooth supermanifold with boundary; see for instance section 3.5 of [15]. With this definition, integration of a smooth measure is a well-defined operation on a supermanifold with boundary; there is no integration ambiguity.

In general, of course, in superstring perturbation theory, one is not dealing with smooth measures. The thorniest singularity that can arise at $q = 0$ is the dq/q singularity associated to a massless tadpole (this will be a closed-string tadpole in fig. 31, or an open-string tadpole in fig. 32). When the integrated tadpole vanishes, there is a natural procedure described in section 7.6 to regularize the singularity and evaluate the resulting integral.

As usual, we can be a little more precise than (9.2); we can restrict to $f = f_\ell + f_r$, where f_ℓ depends only on the moduli of Σ_ℓ and f_r depends only on the moduli of Σ_r . The indeterminacy of q is really only of the form

$$(9.3) \quad q \rightarrow e^{f_\ell + f_r} q.$$

This has an important implication for the exceptional closed-string degenerations of fig. 31. In those cases, Σ_r , being a disc or a copy of \mathbb{RP}^2 with only one puncture, has no moduli at all. So f_r is a function on a point, in other words a constant. Now suppose that Σ_ℓ is the same in figs. 31(a) and (b), and denote the gluing parameter as $q_{\Sigma_\ell, D}$ or $q_{\Sigma_\ell, \mathbb{RP}^2}$ depending on whether Σ_r is a disc D or a copy of \mathbb{RP}^2 . Since Σ_ℓ is the same in the two cases, we can use

⁷⁷In addition to these ordinary boundaries, $\widehat{\Gamma}$ has virtual boundaries associated to generic closed-string degenerations, just as in the case of oriented closed strings.

the same local parameters on Σ_ℓ in defining $q_{\Sigma_\ell, \mathbb{D}}$ and $q_{\Sigma_\ell, \mathbb{RP}^2}$. If we do this, a change in those local parameters multiplies $q_{\Sigma_\ell, \mathbb{D}}$ and $q_{\Sigma_\ell, \mathbb{RP}^2}$ by the same factor e^{f_ℓ} , and this cancels out of the ratio $q_{\Sigma_\ell, \mathbb{D}}/q_{\Sigma_\ell, \mathbb{RP}^2}$. What about f_r ? If one has what one regards as a preferred way to fix a local coordinate at the puncture of a once-punctured disc or \mathbb{RP}^2 , then one can also eliminate the indeterminacy in this ratio due to f_r . But even if one does not wish to select preferred local coordinates in these two special cases, the ambiguity due to f_r only affects the ratio by a multiplicative constant, since f_r is itself a constant in each of the two cases. So the ratio of gluing parameters is well-defined up to a positive multiplicative constant:

$$(9.4) \quad \frac{q_{\Sigma_\ell, \mathbb{D}}}{q_{\Sigma_\ell, \mathbb{RP}^2}} \rightarrow e^\kappa \frac{q_{\Sigma_\ell, \mathbb{D}}}{q_{\Sigma_\ell, \mathbb{RP}^2}}, \quad \kappa \in \mathbb{R}.$$

Here κ is entirely independent of Σ_ℓ . This is ultimately important in canceling infrared divergences.

9.2.2. BRST anomalies Obviously, once we consider superstring theories with open as well as closed strings, we have to consider gauge transformations $\mathcal{V} \rightarrow \mathcal{V} + \{Q_B, \mathcal{W}\}$, where now \mathcal{V} may be an open-string vertex operator. We also have to allow for the fact that in open and/or unoriented string theory, the integration cycle \widehat{I} may have actual boundaries, as well as virtual boundaries associated to closed-string degenerations.

Still, our analysis in section 7 carries over, with only a few changes. For example, just as before, BRST anomalies in a scattering amplitude can come only from degenerations of a worldsheet Σ to two branches Σ_ℓ and Σ_r , of such a type that the momentum flowing between the two branches is automatically on-shell. The anomalies can therefore only come from the obvious generalizations of figs. 16 and 17 of section 7.3, as follows: in general, Σ_ℓ and Σ_r may be open and/or unoriented surfaces, they may be joined at an open-string degeneration rather than a closed-string degeneration, and some or all of the external vertex operators may be open-string vertex operators. These generalizations do not affect much that we said previously, and the conclusion relating anomalies to mass renormalization and tadpoles is almost unaffected.

Clearly, mass renormalization for open-string states is now relevant. We avoid this as usual by restricting to the massless S -matrix in models in which supersymmetry prevents mass renormalization for massless particles.

Also, tadpoles may now arise for either open- or closed-string states, and closed-string tadpoles may be of either NS-NS or R-R type. The open-string tadpoles do not lead to much novelty, as will be clear in section 9.2.3. The

R-R tadpoles do involve novelties, related to the peculiar fermionic gluing parameter that appears at a Ramond degeneration; see section 9.3. In particular, hexagon anomalies that appear at one-loop order in Type I superstring theory if the gauge group is not $SO(32)$ [82] have their origin in R-R tadpoles.

As we will see, the real importance of the exceptional closed-string degenerations of fig. 31 is their relation to R-R tadpoles. They also represent potential new contributions to NS-NS tadpoles, but in this they are not unique. Open and/or unoriented superstring theory has more options for the worldsheet topology than closed oriented superstring theory, and all of them are equally relevant in analyzing NS-NS tadpoles.

9.2.3. Spacetime supersymmetry for open and/or unoriented superstrings The analysis of spacetime supersymmetry from section 8.2.3 must likewise be generalized to open and/or unoriented superstring theory. The basic idea as before is to analyze the identity

$$(9.5) \quad 0 = \int_{\partial\widehat{F}} F_{\mathcal{S}_\alpha \mathcal{V}_1 \dots \mathcal{V}_n}$$

where now $\mathcal{V}_1 \dots \mathcal{V}_n$ may be open- or closed-string vertex operators, and we must consider real as well as virtual boundary components of the compactified integration cycle \widehat{F} . (For the moment, \mathcal{S}_α may be a left- or right-moving supersymmetry generator or a linear combination.) For the same reasons as before, contributions to this formula can only come from the four types of degeneration sketched in fig. 25 of section 8.2.3, generalized in the obvious way to allow for open and/or unoriented string worldsheets and open-string vertex operators.

In particular, we use the open-string analog of the degeneration of fig. 25(a) to define how the supercharge Q_α associated to \mathcal{S}_α acts on an open-string vertex operator \mathcal{V}_i . The contribution for the case that Σ_ℓ is a disc is sketched in fig. 33(a). (Just as for closed strings, there may also be loop corrections to $Q_\alpha(\mathcal{V}_i)$, presumably associated to mass renormalization.) As before, as long as spontaneous supersymmetry breaking, mass renormalization, and tadpoles do not come into play, the analogs of the other degenerations in fig. 25 do not contribute. Under this hypothesis, one gets a linear supersymmetric Ward identity among scattering amplitudes, with the standard form of eqn. (8.29). In particular, under these conditions, massless open-string tadpoles vanish (fig. 33(b)).

However, the question of spontaneous supersymmetry breaking does have some special features in the presence of open strings. What we will now describe is the analog for spacetime supersymmetry of spontaneous breaking of

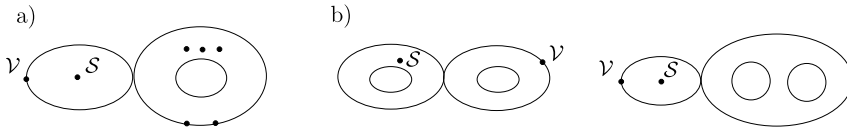


Figure 33: (a) The action of a supercharge Q_α associated to a supercurrent \mathcal{S}_α on an open-string vertex operator \mathcal{V} is obtained by evaluating the contribution to the Ward identity (9.5) of degenerations of this type, with only \mathcal{S}_α and \mathcal{V} on the left, and any collection of open and/or closed-string vertex operators on the right. (b) The proof that massless open-string tadpoles vanish comes by considering the correlator $\langle \mathcal{S}_\alpha \mathcal{V} \rangle$, where \mathcal{V} is an open-string vertex operator. The Ward identity has only the two contributions sketched here; they are associated respectively to spontaneous breaking of the supersymmetry Q_α and to the tadpole of the vertex operator $\{Q_\alpha, \mathcal{V}\}$. If there is unbroken supersymmetry in perturbation theory, then the vanishing of massless open-string tadpoles follows.

B -field gauge-invariance by mixing of open and closed strings, as described in [72] and in section 7.4.1.

Matters are particularly simple for a theory of oriented open and closed superstrings, in other words an extension of Type II superstring theory with D-branes but no orientifold plane. It is instructive to examine this case even though in higher orders such theories are frequently anomalous. (The anomaly, which we study in section 9.4, arises in a higher order of perturbation theory than the effect we will describe now, since it comes from an annulus diagram rather than a disc.)

Let us consider the supersymmetry-breaking degeneration of fig. 25(d), in which a surface Σ splits into two components Σ_ℓ and Σ_r , with only a supercurrent inserted on Σ_ℓ . But now we focus on the special case that Σ_ℓ is a disc (fig. 34(a)); and for definiteness, we consider the supercurrent to be a left-moving one \mathcal{S}'_α . By familiar reasoning, the contribution to the path integral from the degeneration indicated in the figure can be computed by replacing Σ_ℓ by an insertion of an operator $\mathcal{O}(\mathcal{S}'_\alpha)$ on Σ_r (fig. 34(b)). $\mathcal{O}(\mathcal{S}'_\alpha)$ is the vertex operator of a physical open-string state of spin 1/2 (and momentum zero) – a Goldstone fermion. Specifically for the case that Σ_ℓ is a disc, the operator $\mathcal{O}(\mathcal{S}'_\alpha)$ is always non-zero. Indeed, for Σ_ℓ a disc, the degeneration of fig. 34(a) is conformally equivalent to a limit in which the operator \mathcal{S}'_α approaches the boundary of Σ from the interior (as in fig. 34(c)). Thus, for Σ_ℓ a disc, $\mathcal{O}(\mathcal{S}'_\alpha)$ is simply a matrix element of the bulk-to-boundary analog

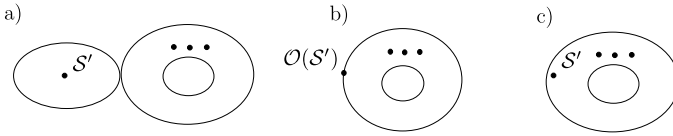


Figure 34: (a) Supersymmetry breaking by mixing of open and closed strings is reflected in a non-zero contribution to the Ward identity from the degeneration shown here, in which the left branch is a disc and contains only the supercurrent \mathcal{S}' . The right branch in the example shown is an annulus with several closed-string vertex operators. (b) As usual, this contribution can be evaluated via an insertion on the right branch of an operator $\mathcal{O}(\mathcal{S}')$ – the vertex operator for the Goldstone fermion, which in this mechanism for supersymmetry breaking is an open-string mode. (c) At tree level (that is, when the left branch is a disc), the operator $\mathcal{O}(\mathcal{S}')$ is simply the limit of the supercurrent \mathcal{S}' as the point at which it is inserted approaches the boundary; this limit is conformally equivalent to the degeneration in (a). Hence, $\mathcal{O}(\mathcal{S}')$ at tree level is a matrix element of the bulk-to-boundary analog of the operator product expansion.

of the OPE [85]. A holomorphic field such as \mathcal{S}'_α always has a non-zero limit as it approaches the boundary. It can in general be oversimplified to think of a Ramond vertex operator as a conformal field, since in this viewpoint it is difficult to incorporate the odd moduli of a super Riemann surface. However, the behavior as \mathcal{S}'_α approaches the boundary can be computed by a local bulk-to-boundary computation in which odd moduli, which are global in nature, play no important role. So for this purpose, one may view \mathcal{S}'_α as just another type of holomorphic conformal field, and it is inevitably non-zero along the boundary of Σ .

The fact that $\mathcal{O}(\mathcal{S}'_\alpha) \neq 0$ means that the presence of a boundary definitely breaks the supersymmetry generated by a left-moving supercurrent \mathcal{S}'_α . Similarly, if \mathcal{S}''_α is a right-moving supercurrent, then $\mathcal{O}(\mathcal{S}''_\alpha)$ is non-zero and the supersymmetry generated by \mathcal{S}''_α is definitely broken by the boundary condition. The only way to find a spacetime supersymmetry that is not broken by the boundary condition is to find \mathcal{S}'_α and \mathcal{S}''_α such that

$$(9.6) \quad \mathcal{O}(\mathcal{S}'_\alpha) + \mathcal{O}(\mathcal{S}''_\alpha) = 0.$$

We write \mathcal{S}_L for the space of left-moving supercurrents \mathcal{S}'_α and \mathcal{S}_L^* for the subspace of \mathcal{S}_L such that this eqn. (9.6) is satisfied for some right-moving

supercurrent \mathcal{S}''_α . For $\mathcal{S}'_\alpha \in \mathcal{S}_L^*$, the solution \mathcal{S}''_α is unique (since $\mathcal{O}(\mathcal{S}'')$ is non-zero for all non-zero right-moving supercurrents \mathcal{S}'') and of course it is linear in \mathcal{S}'_α . We denote \mathcal{S}''_α as $\phi_D(\mathcal{S}'_\alpha)$, where the notation is meant to remind us that ϕ_D reflects the effects of D-branes. Thus a generator of unbroken supersymmetry is a linear combination

$$(9.7) \quad \mathcal{S}_\alpha = \mathcal{S}'_\alpha + \phi_D(\mathcal{S}'_\alpha)$$

of left and right-moving supercurrents.

A simple example [87] that illustrates these ideas is Type IIB superstring theory in \mathbb{R}^{10} , with D9-branes. Half of the bulk supersymmetry is spontaneously broken by the coupling to the branes. The Goldstone fermion is an open-string mode.

In unoriented superstring theory, the starting point is simpler. Even in bulk, the only supercurrents that can be defined without a choice of orientation are mixtures of left- and right-moving supercurrents of an underlying Type II theory. In the notation of section 9.1, such mixtures take the form $\mathcal{S}_\alpha = \mathcal{S}'_\alpha + \phi(\mathcal{S}'_\alpha)$, where ϕ is the orientifold projection viewed as a map from left-moving to right-moving supercurrents. In a theory with D-branes as well as an orientifold projection, the condition that the mixing of open and closed strings at the level of the disc amplitude does not spontaneously break supersymmetry is that $\mathcal{O}(\mathcal{S}_\alpha) = 0$. Supersymmetries that are unbroken in the presence of both the D-branes and the orientifold projection are derived from left-moving supercurrents \mathcal{S}'_α such that $\phi(\mathcal{S}'_\alpha) = \phi_D(\mathcal{S}'_\alpha)$. Henceforth we restrict to such supercurrents and drop the distinction between ϕ and ϕ_D .

After analyzing the disc amplitudes, one can ask whether higher order perturbative corrections trigger further spontaneous breaking of supersymmetry. Just as for closed oriented strings, simple arguments can usually be given to show that this does not occur (or, more exceptionally, that it does). For example, in section 8.2.3, we gave an elementary argument of spacetime chirality to show that in ten-dimensional Type I superstring theory, the supersymmetry that is allowed by the orientifold projection is not spontaneously broken in perturbation theory. Given such an argument, one has the right ingredients for the considerations that we will present next.

9.3. NS-NS and R-R tadpoles

9.3.1. Preview We now come to a point at which open and/or unoriented superstring theories are really different. For closed oriented superstrings, we were able in section 8.3 to use spacetime supersymmetry to prove the vanishing of all massless tadpoles. But for open and/or unoriented superstrings,

spacetime supersymmetry does not lead to a result as strong as this. Given our usual assumptions, supersymmetry leads to a relation that expresses NS-NS tadpoles in terms of R-R tadpoles, schematically

$$(9.8) \quad \langle \mathcal{V}_{\text{NS-NS}} \rangle + \langle \mathcal{V}_{\text{R-R}} \rangle = 0,$$

but not to a general proof that the tadpoles vanish.

There is a surprise here. Physical R-R fields decouple at zero momentum, even in open and/or unoriented superstring theory. This is a consequence of the fact that R-R fields are r -form gauge fields (for various values of r , depending on the model), which couple only via their $r + 1$ -form field strengths. For example, the field strength of a physical R-R scalar field is a closed one-form, which vanishes at zero spacetime momentum. In section 6.2.2, we essentially deduced this vanishing as a consequence of integration over fermionic gluing parameters. That being so, one might not expect R-R tadpoles.

It is indeed true that there are no tadpoles for physical R-R fields. What we will call an R-R tadpole is the expectation value of a certain R-R vertex operator, but this is *not* the superconformal vertex operator of a physical R-R field. Why do we have to worry about R-R fields outside of the usual pantheon of superconformal vertex operators? The answer is subtle and involves the fact that at certain exceptional R-R degenerations, the usual fermionic gluing parameter is absent. Explaining this is our main goal.

The net effect is that R-R tadpoles arise from string worldsheets of only two possible topologies: a disc or $\mathbb{R}P^2$. As already noted in section 9.1.2, it follows that in a given superstring compactification, R-R tadpoles can be completely calculated.

To convince oneself without any technicalities that the R-R operators that lead to tadpoles cannot be vertex operators of physical states, it suffices to consider the most classic example of a superstring theory in which R-R tadpoles are important. This is Type I superstring theory in ten dimensions, where R-R tadpoles underlie the classic computations [82, 83] of anomalies and NS-NS tadpoles. (This is explained and the key points are analyzed in [84], in a language somewhat different from ours.) This theory has no physical R-R field of spin zero, so the tadpole must be associated to some R-R vertex operator that is not the vertex operator of a physical field.

9.3.2. Ward identity with R-R vertex operators We start with a simplified explanation of why R-R vertex operators appear in studying NS-NS tadpoles for open and/or unoriented superstrings. First we consider the illustrative case of a Type II superstring compactification modified by adding D-branes but not orientifold planes.

In a Type II model, let $\mathcal{V}_{\text{NS-NS}}$ be the vertex operator for a spin 0 field of zero momentum whose tadpole we wish to study. As already observed in the discussion of eqn. (9.1), in the absence of D-branes and orientifold planes, we could establish the vanishing of the tadpole using spacetime supersymmetries associated to either a left- or right-moving supercurrent. For example, we can take a left-moving supercurrent, and find a relation at zero momentum

$$(9.9) \quad \mathcal{V}_{\text{NS-NS}} = \sum_{\alpha} \{Q'_{\alpha}, \mathcal{V}_{\text{R-NS}}^{\alpha}\},$$

where $\mathcal{V}_{\text{R-NS}}^{\alpha}$ is a fermion vertex operator of R-NS type. From this we deduce the vanishing of the NS-NS tadpole as in section 8.3, assuming that supersymmetry is not spontaneously broken.

When we add D-branes, the supersymmetry generated by the left-moving supercurrent \mathcal{S}'_{α} is always broken by the boundary conditions and instead we must use a linear combination $\mathcal{S}'_{\alpha} + \mathcal{S}''_{\alpha}$ (or $\mathcal{S}'_{\alpha} + \phi_{\text{D}}(\mathcal{S}'_{\alpha})$, in the notation of eqn. (9.7)) that generates an unbroken supersymmetry. What happens now, roughly speaking, is that the supersymmetry Q''_{α} generated by \mathcal{S}''_{α} acts on the right-moving part of the vertex operator $\mathcal{V}_{\text{R-NS}}^{\alpha}$, transforming it into a vertex operator of R-R type that we will schematically call $\mathcal{V}_{\text{R-R}}$. So in the presence of D-branes, eqn. (9.9) must be replaced by something more like

$$(9.10) \quad \mathcal{V}_{\text{NS-NS}} + \mathcal{V}_{\text{R-R}} = \sum_{\alpha} \{Q'_{\alpha} + Q''_{\alpha}, \mathcal{V}_{\text{R-NS}}^{\alpha}\}.$$

This formula is very schematic and we will see that it needs some corrections.

If we consider an orientifold plane as well as (or instead of) D-branes, everything is much the same except that we must use operators invariant under the orientifold projection. So in the starting point, the operator $\mathcal{V}_{\text{R-NS}}^{\alpha}$ must be replaced by a linear combination of operators from the R-NS and NS-R sectors. With orientifold planes, a better (though still schematic) statement of the identity that constrains R-R tadpoles is

$$(9.11) \quad \mathcal{V}_{\text{NS-NS}} + \mathcal{V}_{\text{R-R}} = \sum_{\alpha} \{Q'_{\alpha} + Q''_{\alpha}, \mathcal{V}_{\text{R-NS}}^{\alpha} + \mathcal{V}_{\text{NS-R}}^{\alpha}\}.$$

These formulas suggest that spacetime supersymmetry for open and/or unoriented superstring theory will only tell us that NS-NS tadpoles can be expressed in terms of R-R tadpoles, not that they each vanish separately. In a sense, that is the right answer, but we are still several steps removed from a correct explanation.

A more careful derivation, as we explain shortly, shows that if eqn. (9.11) is understood as a relation between physical state vertex operators, then $\mathcal{V}_{\text{R-R}}$ must be multiplied by a factor of the spacetime momentum k_I . This factor is related to the spacetime supersymmetry algebra $\{Q_\alpha, Q_\beta\} = \Gamma_{\alpha\beta}^I P_I$; by virtue of this formula, when we act with Q_α on the vertex operators, a factor of momentum has to appear somewhere. The purpose of section 9.3.3 is to explain that this factor actually appears multiplying $\mathcal{V}_{\text{R-R}}$. From this it seems that upon setting $k_I = 0$ to study the tadpoles, the R-R contribution will disappear and we will learn, just as for Type II, that the NS-NS tadpoles vanish. That conclusion is mistaken, but only for the case that the superstring worldsheet Σ is a disc or a copy of $\mathbb{R}P^2$. It seems difficult to understand the exceptions through a corrected version of a formula such as (9.11); we will really have to go back to the identity (9.5) that was the starting point in defining the spacetime supercharges Q_α .

9.3.3. A factor of momentum To proceed, instead of being abstract, we consider the illustrative example of Type I superstring theory in ten dimensions. For the more general case of a compactification to $d < 10$ dimensions, one essentially repeats the analysis once for every relevant zero-mode wavefunction in the internal space [86]. These wavefunctions behave as constants in the following analysis, so keeping track of them would mostly modify only our notation.

The action of spacetime supersymmetry on vertex operators can be computed [1] using the operator product expansion of superconformal field theory. In doing so, one can treat the relevant vertex operators as products of holomorphic and antiholomorphic factors.⁷⁸ The relevant building blocks of physical state vertex operators are the right-moving bosonic and fermionic massless vertex operators

$$(9.12) \quad \begin{aligned} \mathcal{Y}^I &= c\delta(\gamma)D_\theta X^I \exp(ik \cdot X) \\ \mathcal{Z}_\alpha &= c\Theta_{-1/2}\Sigma_\alpha \exp(ik \cdot X), \end{aligned}$$

⁷⁸The vertex operator of a string state of momentum k contains a factor $\exp(ik \cdot X)$ that only factorizes locally, but this is good enough for computing OPE's. The factors $\exp(ik \cdot X)$ in eqns. (9.12) and (9.13) below are understood as functions of the right-moving and left-moving parts of X , respectively. In compactifications, the zero-mode wavefunctions mentioned in the last paragraph do not factorize, but they behave as constants in the specific OPE's we will need and do not affect the analysis.

and their left-moving counterparts

$$(9.13) \quad \begin{aligned} \tilde{\mathcal{Y}}^I &= \tilde{c}\delta(\tilde{\gamma})D_{\tilde{\theta}}X^I \exp(ik \cdot X) \\ \tilde{\mathcal{Z}}_\alpha &= \tilde{c}\tilde{\Theta}_{-1/2}\tilde{\Sigma}_\alpha \exp(ik \cdot X). \end{aligned}$$

The expressions we have written here have the canonical picture numbers -1 for bosons and $-1/2$ for fermions; as we know from sections 4.1 and 5.1, superconformal vertex operators appropriate for evaluating the scattering amplitudes by integration over the usual moduli spaces exist only at those picture numbers. $\Theta_{-1/2}$ represents the $\beta\gamma$ ground state at picture number $-1/2$, and Σ_α is the spin field of the matter system, projected to the positive chirality part so that \mathcal{Z}_α is invariant under the GSO projection Π_{GSO} . For $k \neq 0$, the physical state conditions require us to take certain linear combinations of the above operators (such as $\varepsilon_I \mathcal{Y}^I$ with $\varepsilon \cdot k = 0$ or $u^\alpha \mathcal{Z}_\alpha$, with $(\Gamma \cdot k)_{\alpha\beta} u^\alpha = 0$). At $k = 0$, this is unnecessary. Physical state vertex operators are constructed by multiplying left- and right-moving factors of this type. For example, the NS-NS vertex operator whose tadpole we need to analyze is the operator $\mathcal{V}^\phi = \tilde{\mathcal{Y}}_I \mathcal{Y}^I$ at zero momentum.

Finally, we will also need the fermion vertex operators at picture number $-3/2$:

$$(9.14) \quad \begin{aligned} \mathcal{Z}_*^\alpha &= c\Theta_{-3/2}\Sigma^\alpha \exp(ik \cdot X) \\ \tilde{\mathcal{Z}}_*^\alpha &= \tilde{c}\tilde{\Theta}_{-3/2}\tilde{\Sigma}^\alpha \exp(ik \cdot \tilde{X}). \end{aligned}$$

Here $\Theta_{-3/2}$ represents the $\beta\gamma$ ground state now at picture number $-3/2$, and Σ^α is the chirality -1 part of the spin field of the matter system. $\Theta_{-3/2}$ and Σ^α are both GSO-odd, so their product is even (see section 5.5.1). Thus \mathcal{Z}_*^α is GSO-invariant, and so similarly is $\tilde{\mathcal{Z}}_*^\alpha$. Importantly, they have opposite space-time chirality from \mathcal{Z}_α and $\tilde{\mathcal{Z}}_\alpha$. The right-moving picture-changing operator \mathcal{Y} maps \mathcal{Z}_*^α to \mathcal{Z}_β , modulo a BRST-exact operator, with an important factor of the momentum:

$$(9.15) \quad \mathcal{Y} \cdot \mathcal{Z}_*^\alpha = (\Gamma \cdot k)^{\alpha\beta} \mathcal{Z}_\beta + \{Q_B, \cdot\}.$$

There is of course a similar formula for left-movers.

Since \mathcal{Z}_*^α and $\tilde{\mathcal{Z}}_*^\alpha$ are not superconformal vertex operators, they cannot be used in computing scattering amplitudes by integration over the usual moduli space of super Riemann surfaces. But it will turn out that it is difficult to discuss R-R tadpoles without considering these operators.

As we have already discussed in section 8.2.4 in the context of oriented closed superstrings, the usual procedure [1] to construct the spacetime supersymmetry generators is as follows. The basic ingredient is the spin 1 holomorphic field $\widehat{\mathcal{S}}_\alpha = \Theta_{-1/2}\Sigma_\alpha$, which we have written at the canonical and most convenient picture number $-1/2$. This field may loosely speaking be regarded as a conserved current on the string worldsheet, and as such it generates a symmetry – spacetime supersymmetry. This interpretation is valid locally, though globally it somewhat obscures the super Riemann surface geometry. (What we call the supercurrent in the present paper is not $\widehat{\mathcal{S}}_\alpha$ but $\mathcal{S}_\alpha = c\widehat{\mathcal{S}}_\alpha$.) The action of spacetime supersymmetry on a vertex operator \mathcal{V} is usually defined by extracting a pole in the OPE $\widehat{\mathcal{S}}(z)\mathcal{V}(w)$, and if necessary applying a picture-changing operation so as to return to a canonical value of the picture number. For example

$$(9.16) \quad \widehat{\mathcal{S}}_\alpha(z)\mathcal{Z}_\beta(w) \sim \frac{1}{z-w}\Gamma_{\alpha\beta}^I\mathcal{Y}_I.$$

In this case, picture-changing is unnecessary, and one interprets the formula to mean that

$$(9.17) \quad \{Q_\alpha, \mathcal{Z}_\beta\} = \Gamma_{\alpha\beta}^I\mathcal{Y}_I.$$

On the other hand, since $\widehat{\mathcal{S}}$ has picture number $-1/2$ and \mathcal{Y}_I has picture number -1 , the product $\widehat{\mathcal{S}} \cdot \mathcal{Y}_I$ has picture number $-3/2$, so in this case we will want to apply a picture-changing operation to map back to the canonical picture number, which for Ramond vertex operators is $-1/2$. The pole in the OPE is

$$(9.18) \quad \widehat{\mathcal{S}}_\alpha(z)\mathcal{Y}_I(w) \sim \frac{1}{z-w}\Gamma_{I\alpha\beta}\mathcal{Z}_*^\beta.$$

After picture-changing via eqn. (9.15), one therefore defines

$$(9.19) \quad \{Q_\alpha, \mathcal{Y}_I\} = (\Gamma_I\Gamma \cdot k)_\alpha{}^\beta\mathcal{Z}_\beta.$$

Eqns. (9.17) and (9.19) are compatible⁷⁹ with the expected supersymmetry algebra $\{Q_\alpha, Q_\beta\} = \Gamma_{\alpha\beta}^I P_I$. We will also want later a variant of (9.18); with

⁷⁹To demonstrate this is a little tricky. One must use the fact that \mathcal{Y}^I only appears in a combination $\varepsilon_I\mathcal{Y}^I$ where $\varepsilon \cdot k = 0$, and one must also use the fact that $k_I\mathcal{Y}^I$ is Q_B -trivial. And similarly, one must use the fact that \mathcal{Z}_α only appears in a combination $u^\alpha\mathcal{Z}_\alpha$ where $k \cdot \Gamma u = 0$.

$\mathcal{S}_\alpha = c\widehat{\mathcal{S}}_\alpha$, we have

$$(9.20) \quad \mathcal{S}_\alpha(z)\mathcal{Y}_I(w) \sim \Gamma_{I\alpha\beta}\partial c\mathcal{Z}_*^\beta.$$

The purpose of this explanation has been to show that the action of Q_α on a massless NS vertex operator to produce a Ramond vertex operator has no factor of the momentum k_I (eqn. (9.17)), while its action on a massless Ramond vertex operator to produce a vertex operator in the NS sector is proportional to k_I (eqn. (9.19)). Accordingly, in a more careful derivation, the R-R vertex operator $\mathcal{V}_{\text{R-R}}$ on the left hand side of eqns. (9.10) or (9.11) would be multiplied by a factor of k_I .

Setting $k = 0$ to study tadpoles, this contribution would therefore disappear. This result is consistent with the intuition that R-R fields decouple at zero momentum, but it leaves one wondering: How can R-R tadpoles possibly appear?

9.3.4. When the fermionic gluing parameter disappears The answer to the last question is that under certain conditions, there is no picture-changing in the derivation of the supersymmetric Ward identity.

Let us return to our definition of the supercharge, based on the process sketched in fig. 26 of section 8.2.3. What in the usual description is called picture-changing in the action of the supercharge Q_α on an NS vertex operator arises in our point of view from integration over the fermionic gluing parameter in fig. 26(a). If we want the usual picture-changing not to appear, we need to find a reason that the usual fermionic gluing parameter is missing. The way to eliminate it is the same as the way to eliminate the argument of the closed-string gluing parameter q , as described in section 9.2.1 and fig. 31. If either Σ_ℓ or Σ_r has a fermionic symmetry that acts by shifting the fermionic divisor at which Σ_ℓ and Σ_r are glued together, then the fermionic gluing parameter can be absorbed in an automorphism of Σ_ℓ or Σ_r .

For convenience, we repeat the relevant picture here (fig. 35) for the case of a right-moving supercurrent \mathcal{S}_α'' acting on a vertex operator $\mathcal{V}_{\text{R-NS}}$ from the R-NS sector. An R-R state is propagating between Σ_ℓ and Σ_r . If Σ_ℓ and Σ_r were completely generic, then to evaluate the matrix element represented by this picture, we would have to integrate over both a left-moving and right-moving (antiholomorphic and holomorphic) fermionic gluing parameter.

However, we will take Σ_ℓ to have genus 0, and in this case something special happens. From an antiholomorphic point of view, the operator \mathcal{S}_α'' is the identity and hence Σ_ℓ is a genus 0 surface with only two punctures, associated to the insertion of $\mathcal{V}_{\text{R-NS}}$ and the intersection of Σ_ℓ and Σ_r . These are

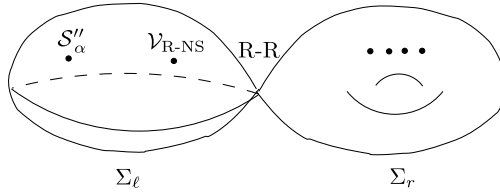


Figure 35: A right-moving supercurrent acting on a vertex operator of R-NS type to produce an R-R vertex operator.

both Ramond punctures. A genus 0 super Riemann surface with only two Ramond punctures has a fermionic automorphism which can be used to remove the antiholomorphic gluing parameter. One can describe such a surface Σ_ℓ by coordinates $\tilde{z}|\tilde{\theta}$ (including a divisor at $\tilde{z} = \infty$), with the superconformal structure being defined by $D_\theta^* = \partial_{\tilde{\theta}} + \tilde{\theta}\tilde{z}\partial_{\tilde{z}}$. The Ramond divisors are defined by $\tilde{z} = 0$ and $\tilde{z} = \infty$; we call them $\tilde{\mathcal{F}}_0$ and $\tilde{\mathcal{F}}_\infty$, respectively. Σ_ℓ admits the odd superconformal vector field

$$(9.21) \quad \tilde{\nu} = \partial_{\tilde{\theta}} - \tilde{\theta}\tilde{z}\partial_{\tilde{z}}.$$

For future reference, we note that

$$(9.22) \quad \tilde{\nu}^2 = -\tilde{z}\partial_{\tilde{z}}.$$

When restricted to either $\tilde{\mathcal{F}}_0$ or $\tilde{\mathcal{F}}_\infty$, $\tilde{\nu}$ reduces to $\partial_{\tilde{\theta}}$ (to see this for $\tilde{\mathcal{F}}_\infty$, it helps to transform $\tilde{z} \rightarrow 1/\tilde{z}$), and generates the symmetry

$$(9.23) \quad \tilde{\theta} \rightarrow \tilde{\theta} + \alpha,$$

with an anticommuting parameter α . For more on this, see sections 5.1.4 and 7.4.4 of [16]. The symmetry of Σ_ℓ that we have just described, since it shifts $\tilde{\theta}$ by an arbitrary constant, can be used to transform away the gluing parameter that arises when $\tilde{\mathcal{F}}_0$ or $\tilde{\mathcal{F}}_\infty$ is glued to a Ramond divisor in Σ_r .

From a holomorphic point of view, Σ_ℓ has a third puncture and no fermionic automorphism. According, if Σ_r is generic, then to evaluate the action of the supercharge, we will have to integrate over one fermionic gluing parameter, the holomorphic one. This will reproduce what in the conventional approach comes from picture-changing.

To eliminate the holomorphic fermionic gluing parameter from fig. 35, Σ_r will have to have a fermionic automorphism. A closed, oriented super Riemann surface Σ_r with at least one puncture (the point at which Σ_r intersects

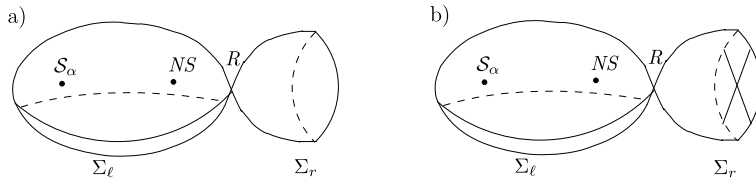


Figure 36: Contributions to the supersymmetric Ward identity that give an R-R operator without a fermionic gluing parameter. Σ_r in fig. 35 must be either (a) a disc or (b) a copy of $\mathbb{R}P^2$, in either case with no vertex operator insertions.

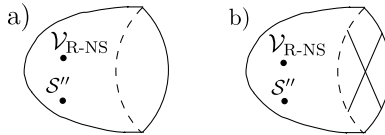


Figure 37: Correlation function $\langle S'' \mathcal{V}_{R-NS} \rangle$ on (a) a disc or (b) a copy of $\mathbb{R}P^2$.

Σ_ℓ) appearing in the Deligne-Mumford compactification⁸⁰ never has such an automorphism, so for closed oriented superstrings, it is never possible to get rid of the fermionic gluing parameter. If Σ_r is open and/or unoriented, it can have a fermionic automorphism, but only in two special cases: Σ_r must be a disc or a copy of $\mathbb{R}P^2$, with precisely one puncture which must be of R-R type. In either of these cases, the closed oriented double cover $\widehat{\Sigma}_r$ of Σ_r is a genus 0 super Riemann surface with two punctures of R type, and its fermionic symmetry can again be described as in eqn. (9.21). Moreover, just as before, this symmetry generates a shift of the Ramond divisors at which gluing occurs and hence can be used to transform away the fermionic gluing parameter.

It is no coincidence that the two cases in which Σ_r has a fermionic automorphism are the two cases in which Σ_r has a continuous bosonic symmetry group, whose role was discussed in section 9.2.1. If Σ_r has an odd superconformal vector field ν , then (as illustrated in eqn. (9.22)), ν^2 is an even superconformal vector field that generates a one-parameter bosonic symmetry group of Σ_r .

In short, the cases in which there is no fermionic gluing parameter are the cases that Σ_r is a disc or a copy of $\mathbb{R}P^2$ (fig. 36). Since the absence of a

⁸⁰We exclude the case that Σ_r is a genus 0 surface with at most two punctures, since this case does not arise in the Deligne-Mumford compactification.

fermionic gluing parameter in those two cases is an important result, we will explain it in another way. The worldsheets in fig. 36 arise by degeneration (as the two vertex operators approach each other) from the smooth worldsheets Σ sketched in fig. 37, consisting of a disc or a copy of \mathbb{RP}^2 with insertions of the two vertex operators \mathcal{S}'_α and $\mathcal{V}_{\text{R-NS}}$. In each of these cases, the closed oriented double cover of Σ is a genus zero surface with one NS puncture and two R punctures. (Since the supercurrent \mathcal{S}'_α is the identity operator from an antiholomorphic point of view, its insertion point lifts on the double cover to a single R puncture. But the insertion point of $\mathcal{V}_{\text{R-NS}}$ lifts on the closed oriented double cover to a pair of punctures, one of NS type and one of R type.) A surface of genus 0 with one NS puncture and two R punctures has no fermionic moduli, and there are still none when one degenerates from the smooth worldsheets of fig. 37 to the singular ones of fig. 36. So in particular, there are no fermionic gluing parameters in that figure.

Since picture-changing does not come into play, the R-R vertex operator $\mathcal{V}_{\text{R-R}}$ that propagates between Σ_ℓ and Σ_r in fig. 36 has picture numbers $(-1/2, -3/2)$. Actually, in open and/or unoriented string theory, we should use a mixture of left- and right-moving supercurrents, such as $\mathcal{S}_\alpha = \mathcal{S}'_\alpha + \phi(\mathcal{S}'_\alpha)$, rather than the right-moving one that we considered for brevity in explaining how the fermionic gluing parameter can disappear. Also, in the unoriented case, a closed-string fermion vertex operator cannot simply come from the R-NS sector; it must be a linear combination of R-NS and NS-R vertex operators. We denote such a linear combination by $\mathcal{V}_{\text{NS/R}}$. These refinements do not affect the essence of what we have said.

9.3.5. The tadpoles We are finally ready to analyze the tadpoles of open and/or unoriented superstring theory. We consider on a string worldsheet Σ of genus g the Ward identity associated to a two-point function $\langle \mathcal{S}_\alpha \mathcal{V}_{\text{NS/R}}^\alpha \rangle$, with \mathcal{S}_α chosen as the generator of a supersymmetry that is unbroken in perturbation theory, and a suitable NS/R vertex operator $\mathcal{V}_{\text{NS/R}}^\alpha$.

Since supersymmetry is unbroken, the Ward identity

$$(9.24) \quad 0 = \int_{\partial\widehat{R}} F_{\mathcal{S}_\alpha \mathcal{V}_{\text{NS/R}}^\alpha}$$

will receive contributions only from one-point functions or in other words tadpoles. Just as explained in section 9.3.2, the commutator of a supercharge with a vertex operator of mixed NS/R type will in general be a linear combination of NS-NS and R-R vertex operators. However, if Σ (or more precisely its reduced space) is anything other than a disc or \mathbb{RP}^2 , the R-R contribution vanishes at zero momentum after integrating over the fermionic gluing

parameter in fig. 35. So for generic Σ , the Ward identity just tells us that the contribution of Σ to an NS-NS tadpole vanishes:

$$(9.25) \quad \langle \mathcal{V}_{\text{NS-NS}} \rangle_{\Sigma} = 0.$$

Hence in supersymmetric compactifications of open and/or unoriented superstrings that obey the usual mild conditions ensuring that supersymmetry is not spontaneously broken in perturbation theory, there are no NS-NS tadpoles except possibly for the case that Σ is a disc or a copy of $\mathbb{R}P^2$.

Precisely in those two cases, there is no fermionic gluing parameter that would kill the R-R contribution. Such a contribution may appear and the identity becomes schematically

$$(9.26) \quad \langle \mathcal{V}_{\text{NS-NS}} \rangle_{\Sigma} + \langle V_{\text{R-R}} \rangle_{\Sigma} = 0.$$

Accordingly, supersymmetry tells us not that NS-NS tadpoles vanish, but that they vanish if and only if the R-R contribution also vanishes. But simplicity persists because the R-R contributions arise only from a disc or $\mathbb{R}P^2$. Since those Riemann surfaces both have Euler characteristic 1, they arise in the same order of perturbation theory and their contributions should be added together. The tadpoles of any given NS vertex operator $\mathcal{V}_{\text{NS-NS}}$ vanish to all orders of perturbation theory if and only if precisely one number vanishes, namely the sum of the disc and $\mathbb{R}P^2$ contributions to the corresponding R-R tadpole.

Let us make all this more concrete for the important example of Type I superstring theory in ten dimensions. In this case, in the notation of section 9.3.3, the only NS-NS vertex operator that might have a tadpole is $\mathcal{V}_{\text{NS-NS}} = \tilde{\mathcal{Y}}^I \mathcal{Y}^I$. The supercurrent of interest is the usual Type I supercurrent $\mathcal{S}_{\alpha} = \mathcal{S}'_{\alpha} + \mathcal{S}''_{\alpha}$. The NS/R vertex operator that is related by supersymmetry to $\mathcal{V}_{\text{NS-NS}}$ is $\mathcal{V}_{\text{NS/R}}^{\alpha} = \Gamma_I^{\alpha\beta} (\tilde{\mathcal{Y}}^I \mathcal{Z}_{\beta} + \tilde{\mathcal{Z}}_{\beta} \mathcal{Y}^I)$. In deriving the Ward identity, since there is no fermionic gluing parameter, we simply use (9.20) to evaluate the R-R contribution. So the R-R vertex operator that contributes in the Ward identity is the operator

$$(9.27) \quad \mathcal{V}_{\text{R-R}} = \tilde{\partial} \tilde{c} \tilde{\mathcal{Z}}_*^{\alpha} \mathcal{Z}_{\alpha} + \tilde{\mathcal{Z}}_{\alpha} \partial c \mathcal{Z}_*^{\alpha}.$$

Although this operator is not a superconformal vertex operator that we could use to calculate scattering amplitudes in the usual way, it does have a well-defined one-point function on a disc or $\mathbb{R}P^2$. A naive way to explain why is that a disc or $\mathbb{R}P^2$ with only one puncture has no even or odd moduli.

Therefore, in computing the one-point function in question, we do not need to know how to integrate over moduli. We give a better explanation momentarily.

The one-point function of $\mathcal{V}_{\text{R-R}}$ on a worldsheet Σ whose reduced space is a disc or $\mathbb{R}\mathbb{P}^2$ is indeed non-zero. This can actually be demonstrated by a simple computation on the closed oriented double cover $\widehat{\Sigma}$ of Σ . $\widehat{\Sigma}$ is a genus 0 super Riemann surface in the purely holomorphic sense, with two Ramond punctures. Passing to $\widehat{\Sigma}$ separates the holomorphic and antiholomorphic factors in $\mathcal{V}_{\text{R-R}}$, which lift to holomorphic operators at the two distinct Ramond punctures of $\widehat{\Sigma}$. So the one-point function $\langle \mathcal{V}_{\text{R-R}} \rangle_{\Sigma}$ is proportional to the purely holomorphic two-point function $\langle \partial c \mathcal{Z}_*^{\alpha}(z) \mathcal{Z}_{\alpha}(z') \rangle_{\widehat{\Sigma}}$. This two-point function is non-zero (see eqn. (5.71)).

It is tricky to correctly normalize the contributions of a disc and $\mathbb{R}\mathbb{P}^2$ and thereby show (for example) that the R-R tadpole cancels in Type I superstring theory precisely for gauge group $SO(32)$. Factorization from an annulus and a Mobius strip is a convenient way to do this [82]; annulus and Mobius strip amplitudes can be conveniently normalized using their Hamiltonian interpretation in the open string channel. The boundary state formalism for superstrings [88] is also useful in constraining and understanding R-R tadpoles. For instance, the operator $\mathcal{V}_{\text{R-R}}$ was related to anomalies using this formalism [89].

The geometrical meaning of the one-point function of $\mathcal{V}_{\text{R-R}}$ on a disc or $\mathbb{R}\mathbb{P}^2$ is as follows. As explained in sections 4.1.1 and 5.1.1 of this paper and in more detail in section 4.3 and Appendix C of [16], in superstring perturbation theory, one can calculate systematically with vertex operators of picture number more negative than the canonical values, provided that one suitably modifies the definition of supermoduli space in a way that increases its odd dimension. Usually this does not add much, since one can immediately integrate over the extra odd moduli and reduce to vertex operators of canonical picture number. We are discussing here the one case in which that step is not possible. As explained in section 5.1.1 above and in [16], a Ramond operator of picture number $-1/2$ is associated to a Ramond divisor, while one of picture number $-3/2$ is associated to a Ramond divisor together with a choice of a point on the divisor; the choice of point usually adds an odd modulus. On a purely holomorphic super Riemann surface, a Ramond divisor is a subvariety \mathcal{F} of dimension $0|1$. For Type II superstrings (possibly enriched with D-branes or orientifold planes), an R-R vertex operator of canonical picture numbers $(-1/2, -1/2)$ is associated to a product $\widetilde{\mathcal{F}} \times \mathcal{F}$ of antiholomorphic and holomorphic submanifolds of dimension $0|1$. An R-R vertex operator of picture numbers $(-3/2, -1/2)$ or $(-1/2, -3/2)$ is associated to $\widetilde{\mathcal{F}} \times \mathcal{F}$ with a choice of a point on $\widetilde{\mathcal{F}}$ or on \mathcal{F} , respectively. For a generic choice of the string

worldsheet Σ , the choice of a point on $\tilde{\mathcal{F}}$ or \mathcal{F} adds one odd modulus, and by integrating over this odd modulus, we can reduce to the case of a vertex operator of picture number $(-1/2, -1/2)$. However, by now we know that a disc or \mathbb{RP}^2 with only one R-R puncture has a fermionic automorphism that acts by shifting $\tilde{\mathcal{F}}$ and \mathcal{F} and hence can be used to gauge away the choice of a point on $\tilde{\mathcal{F}}$ or \mathcal{F} (just as earlier we used it to gauge away a gluing parameter). Thus the choice of this point does not constitute an odd modulus that one could integrate over to reduce to the canonical picture numbers $(-1/2, -1/2)$. So the R-R one-point function on a disc or \mathbb{RP}^2 is the only computation that one can perform with vertex operators of picture $(-3/2, -1/2)$ and $(-1/2, -3/2)$ but which one cannot reduce to a computation with vertex operators of canonical picture numbers.

We are actually here running into precisely the exceptional case [90] in which there is no isomorphism between vertex operators of different picture numbers. At non-zero momentum, there are always picture-changing isomorphisms between vertex operators of different picture number, but this fails at zero momentum in precisely the case that we have just encountered.

9.3.6. What happens when R-R tadpoles cancel? It remains to discuss the following two questions. What happens when R-R tadpoles cancel? And what happens when they do not cancel? The second question is tricky and is reserved to section 9.4. The first question is more straightforward; all we have to do is to restate what we have said in section 7.6 in a context in which vanishing of tadpoles depends on a cancellation between worldsheets of different topologies.

Let us consider a supersymmetric model (for example Type I superstring theory in ten dimensions with gauge group $SO(32)$) in which the R-R tadpoles on a disc \mathbb{D} and on \mathbb{RP}^2 are separately non-zero, but add up to zero. Then likewise, the same is true of the NS-NS tadpoles on \mathbb{D} and on \mathbb{RP}^2 . We write $\mathcal{V}_{\text{NS-NS}}$ for the zero-momentum NS-NS vertex operator that has a tadpole (we can always pick a basis of operators so that there is just one such operator), and let $\langle \mathcal{V}_{\text{NS-NS}} \rangle_{\mathbb{D}}$ and $\langle \mathcal{V}_{\text{NS-NS}} \rangle_{\mathbb{RP}^2}$ be its tadpoles. Tadpole cancellation means that

$$(9.28) \quad \langle \mathcal{V}_{\text{NS-NS}} \rangle_{\mathbb{D}} + \langle \mathcal{V}_{\text{NS-NS}} \rangle_{\mathbb{RP}^2} = 0.$$

Consider a scattering amplitude $\mathcal{A}_{\mathcal{V}_1 \dots \mathcal{V}_n}$ with vertex operators $\mathcal{V}_1 \dots \mathcal{V}_n$. An NS-NS degeneration in which one component is a disc or \mathbb{RP}^2 with no vertex operator insertions (fig. 38) will make an infrared-divergent contribution to this scattering amplitude. (There are no infrared divergences associated to

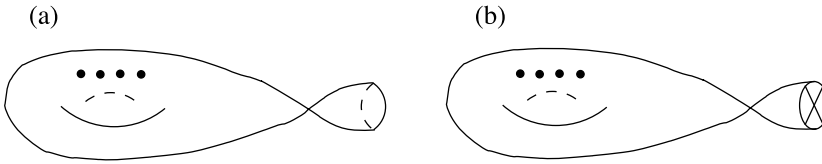


Figure 38: Infrared divergences in scattering amplitudes due to the NS-NS tadpole on (a) a disc, and (b) $\mathbb{R}P^2$. When R-R tadpoles cancel, these infrared divergences also cancel in a natural way.

R-R tadpoles. This will be explained in section 9.4.) To get physically sensible scattering amplitudes, we will need to cancel these infrared divergences in a natural way.

We let Σ_1 be a string worldsheet that can degenerate as in fig. 38(a), to an intersection of an arbitrary worldsheet Σ_ℓ with a disc D . Similarly, let Σ_2 be a string worldsheet that has the degeneration of fig. 38(b), with branches Σ_ℓ and $\mathbb{R}P^2$. The contributions of Σ_1 and Σ_2 to the scattering amplitude $\mathcal{A}_{\nu_1 \dots \nu_n}$ are infrared-divergent; we will give a natural recipe to cancel the divergence and get a finite sum. As in section 9.2.1, we write $q_{\Sigma_\ell, D}$ and $q_{\Sigma_\ell, \mathbb{R}P^2}$ for the gluing parameters in the two cases. Let $m_1 \dots | \dots \eta_s$ be the even and odd moduli of Σ_ℓ (including the point at which it meets Σ_r) and let \mathcal{M}_ℓ be the corresponding moduli space. Near $q_{\Sigma_\ell, D} = 0$, the integral that gives the contribution of Σ_1 to the scattering amplitude looks like

$$(9.29) \quad I_1 \sim \int [dm_1 \dots | \dots d\eta_s] \mathcal{G}_1(m_1 \dots; q_{\Sigma_\ell, D} | \dots \eta_s) \frac{dq_{\Sigma_\ell, D}}{q_{\Sigma_\ell, D}},$$

where the $dq_{\Sigma_\ell, D}/q_{\Sigma_\ell, D}$ singularity reflects the tadpole, and the function $\mathcal{G}_1(m_1 \dots; q_{\Sigma_\ell, D} | \dots \eta_s)$ in general depends on all even and odd moduli, including $q_{\Sigma_\ell, D}$. Similarly, the contribution of Σ_2 to the same scattering amplitude looks near $q_{\Sigma_\ell, \mathbb{R}P^2} = 0$ like

$$(9.30) \quad I_2 \sim \int [dm_1 \dots | \dots d\eta_s] \mathcal{G}_2(m_1 \dots; q_{\Sigma_\ell, \mathbb{R}P^2} | \dots \eta_s) \frac{dq_{\Sigma_\ell, \mathbb{R}P^2}}{q_{\Sigma_\ell, \mathbb{R}P^2}},$$

with some function \mathcal{G}_2 that depends on all moduli, including $q_{\Sigma_\ell, \mathbb{R}P^2}$. Both integrals have logarithmic divergences whose coefficients can be extracted by setting $q_{\Sigma, D}$ to zero in \mathcal{G}_1 and setting $q_{\Sigma_\ell, \mathbb{R}P^2}$ to zero in \mathcal{G}_2 . The coefficients of

the divergences are

$$(9.31) \quad \begin{aligned} I_{1,\log} &= \int_{\mathcal{M}_\ell} [dm_1 \dots | \dots d\eta_s] \mathcal{G}_1(m_1 \dots; 0 | \dots \eta_s) \\ I_{2,\log} &= \int_{\mathcal{M}_\ell} [dm_1 \dots | \dots d\eta_s] \mathcal{G}_2(m_1 \dots; 0 | \dots \eta_s). \end{aligned}$$

As usual, $\mathcal{G}_1(m_1 \dots; q_{\Sigma_\ell; \mathbb{D}} | \dots \eta_s)$ and $\mathcal{G}_s(m_1 \dots; q_{\Sigma_\ell; \mathbb{R}P^2} | \dots \eta_s)$ factor at $q_{\Sigma_\ell; \mathbb{D}} = q_{\Sigma_\ell; \mathbb{R}P^2} = 0$:

$$(9.32) \quad \begin{aligned} \mathcal{G}_1(m_1 \dots; 0 | \dots \eta_s) &= \mathcal{G}_0(m_1 \dots | \dots \eta_s) \langle \mathcal{V}_{\text{NS-NS}} \rangle_{\mathbb{D}} \\ \mathcal{G}_2(m_1 \dots; 0 | \dots \eta_s) &= \mathcal{G}_0(m_1 \dots | \dots \eta_s) \langle \mathcal{V}_{\text{NS-NS}} \rangle_{\mathbb{R}P^2}, \end{aligned}$$

where $\mathcal{G}_0(m_1 \dots | \dots \eta_s)$ depends only on Σ_ℓ . In fact, integration of $[dm_1 \dots | \dots d\eta_s] \mathcal{G}_0$ over \mathcal{M}_ℓ gives the contribution of Σ_ℓ to a scattering amplitude $\mathcal{A}_{\mathcal{V}_1 \dots \mathcal{V}_n; \mathcal{V}_{\text{NS-NS}}}$ with an insertion of $\mathcal{V}_{\text{NS-NS}}$ at zero momentum, as well as insertions of $\mathcal{V}_1, \dots, \mathcal{V}_n$:

$$(9.33) \quad \mathcal{A}_{\mathcal{V}_1 \dots \mathcal{V}_n; \mathcal{V}_{\text{NS-NS}}} = \int_{\mathcal{M}_\ell} [dm_1 \dots | \dots d\eta_s] \mathcal{G}_0.$$

The reasoning here should be familiar from section 7.6.

In view of (9.32), tadpole cancellation (9.28) ensures that logarithmic divergences cancel

$$(9.34) \quad I_{1,\log} + I_{2,\log} = 0.$$

This condition is enough to eliminate the divergent part of $I_1 + I_2$, but not enough by itself to give meaning to the finite remainder. We might introduce infrared cutoffs in I_1 and I_2 by restricting to $|q_{\Sigma_\ell, \mathbb{D}}|, |q_{\Sigma_\ell, \mathbb{R}P^2}| \geq \epsilon$, where ϵ is a small positive constant and the definitions of $|q_{\Sigma_\ell, \mathbb{D}}|$ and $|q_{\Sigma_\ell, \mathbb{R}P^2}|$ depend on arbitrary choices of metric on the relevant line bundles. Tadpole cancellation is enough to ensure that the sum $I_1 + I_2$ has a limit for $\epsilon \rightarrow 0$, but in general this limit will depend on the choices of metric.

This problem is very similar to the one that we grappled with in section 7.6, and its resolution is similar. There is a natural ratio $q_{\Sigma_\ell, \mathbb{D}}/q_{\Sigma_\ell, \mathbb{R}P^2}$ up to a multiplicative constant, as we asserted in eqn. (9.4). So once we make an arbitrary choice of infrared cutoff $|q_{\Sigma_\ell, \mathbb{D}}| \geq \epsilon$ in I_1 , the corresponding choice $|q_{\Sigma_\ell, \mathbb{R}P^2}| \geq \epsilon$ in I_2 is naturally determined, up to a multiplicative constant $q_{\Sigma_\ell, \mathbb{R}P^2} \rightarrow e^{-\kappa} q_{\Sigma_\ell, \mathbb{R}P^2}$.

If we transform $q_{\Sigma_\ell, D}$ and $q_{\Sigma_\ell, \mathbb{R}P^2}$ by common factors of e^h , for some function h , then the limits of I_1 and I_2 for $\epsilon \rightarrow 0$ are both shifted. For example, similarly to the derivation of (7.27), the limit of I_1 for $\epsilon \rightarrow 0$ is shifted by

$$\begin{aligned}
 I_1 &\rightarrow I_1 - \int_{\mathcal{M}_\ell} [dm_1 \dots | \dots d\eta_s] h \mathcal{G}_1(m_1 \dots; 0 | \dots \eta_s) \\
 (9.35) \quad &= I_1 - \langle \mathcal{V}_{\text{NS-NS}} \rangle_D \int_{\mathcal{M}_\ell} [dm_1 \dots | \dots d\eta_s] h \mathcal{G}_0(m_1, \dots | \dots \eta_s),
 \end{aligned}$$

where we used the factorization condition (9.32). Similarly

$$(9.36) \quad I_2 \rightarrow I_2 - \langle \mathcal{V}_{\text{NS-NS}} \rangle_{\mathbb{R}P^2} \int_{\mathcal{M}_\ell} [dm_1 \dots | \dots d\eta_s] h \mathcal{G}_0(m_1, \dots | \dots \eta_s).$$

So the tadpole cancellation condition (9.28) ensures that the sum $I_1 + I_2$ does not depend on h . On the other hand, a transformation $q_{\Sigma_\ell, \mathbb{R}P^2} \rightarrow e^{-\kappa} q_{\Sigma_\ell, \mathbb{R}P^2}$ with constant κ (and with no such change in $q_{\Sigma_\ell, D}$) transforms I_2 , in the limit $\epsilon \rightarrow 0$, by

$$(9.37) \quad I_2 \rightarrow I_2 + \kappa \langle \mathcal{V}_{\text{NS-NS}} \rangle_{\mathbb{R}P^2} \int_{\mathcal{M}_\ell} [dm_1 \dots | \dots d\eta_s] \mathcal{G}_0(m_1, \dots | \dots \eta_s),$$

which is just the special case of (9.36) with $h = -\kappa$. So in view of (9.33), the effect of the rescaling by $e^{-\kappa}$ on the scattering amplitude $\mathcal{A}_{\mathcal{V}_1 \dots \mathcal{V}_n}$ is

$$(9.38) \quad \mathcal{A}_{\mathcal{V}_1 \dots \mathcal{V}_n} \rightarrow \mathcal{A}_{\mathcal{V}_1 \dots \mathcal{V}_n} + \kappa \langle \mathcal{V}_{\text{NS-NS}} \rangle_{\mathbb{R}P^2} \mathcal{A}_{\mathcal{V}_1 \dots \mathcal{V}_n; \mathcal{V}_{\text{NS-NS}}}.$$

This shift in the scattering amplitude can be interpreted as the result of shifting the scalar field that couples to $\mathcal{V}_{\text{NS-NS}}$ by the constant $\kappa \langle \mathcal{V}_{\text{NS-NS}} \rangle_{\mathbb{R}P^2}$. So the change in the S -matrix under the shift by κ can be absorbed in a field redefinition. This is the answer one should expect from section 7.6.

9.4. What happens when R-R tadpoles do not cancel?

Now let us ask what happens if the R-R tadpoles on a disc and $\mathbb{R}P^2$ do not add to zero.

In this case, in a supersymmetric model, there will also be a nonvanishing sum of the NS-NS tadpoles on a disc and $\mathbb{R}P^2$, leading to an infrared divergence and spoiling the validity of perturbation theory.

But what is the significance of the R-R tadpoles themselves? At first, one might assume that R-R tadpoles will lead to an infrared divergence in the R-R channel, but a little reflection might make one skeptical. After all, infrared

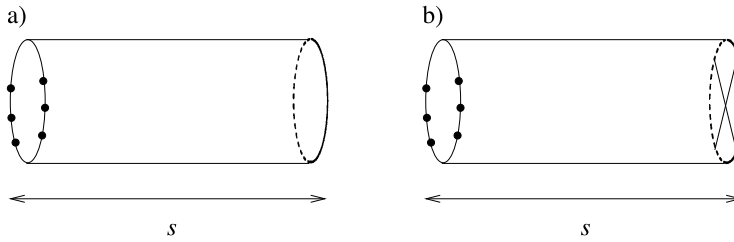


Figure 39: (a) A cylinder Σ of width s with n open-string vertex operators attached to its left boundary Σ_ℓ (pictured for $n = 6$). (b) A Mobius strip of width s with n open-string vertex operators attached to its boundary.

divergences should be associated with propagation of on-shell physical fields, and the R-R vertex operator $\mathcal{V}_{\text{R-R}}$ whose one-point function is the tadpole is not the vertex operator of a physical field.

What happens is rather [82, 84] that an R-R tadpole leads to an anomaly – a failure of gauge invariance. This is a more serious deficiency than an infrared divergence. An infrared divergence in perturbation theory can possibly be eliminated by expanding around a different – possibly time-dependent – classical solution. But an anomaly not related to any infrared divergence is an overall inconsistency of a theory, independent of the choice of a particular classical state or quantum solution. To be more exact, it is clear in field theory that anomalies are not affected by the choice of a quantum state, or of a classical solution around which one expands in order to do perturbation theory. One expects the same in string theory, though string theory is not well enough understood to make it possible to make this statement completely clear.

The classic example [82, 83] of a scattering amplitude that is affected by R-R tadpoles is the one-loop contribution to scattering of massless bosonic open-string states. We aim here to study this example in the framework of the present paper. The worldsheets Σ that contribute to open-string scattering at one-loop order are a cylinder or a Mobius strip. We focus on the “single trace” contribution to the scattering amplitude. This means that if Σ is a cylinder, we take all open-string vertex operator insertions $\mathcal{V}_1, \dots, \mathcal{V}_n$ on its left boundary component, which we call $\partial_\ell \Sigma$ (fig. 39). It will become clear that the anomaly arises precisely in this case. If Σ is a Mobius strip, it has only one boundary component, which we again call $\partial_\ell \Sigma$. As the width s of Σ goes to infinity, Σ undergoes a closed-string degeneration (fig. 40) with Σ_ℓ a disc with n open-string insertions and Σ_r a disc or a copy of $\mathbb{R}\mathbb{P}^2$. For the case that the

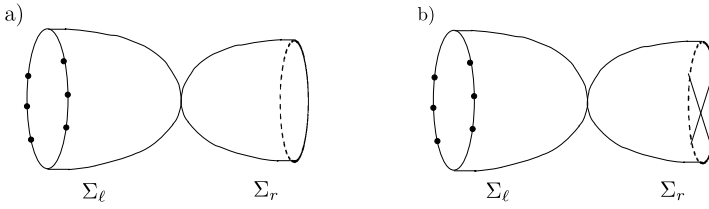


Figure 40: For $s \rightarrow \infty$, the string worldsheets in fig. 39 undergo a closed-string degeneration, as pictured here. Σ_ℓ is a disc with vertex operator insertions, while Σ_r is a disc in (a), or a copy of $\mathbb{R}P^2$ in (b). These particular closed-string degenerations are the ones that produces R-R tadpoles and other exceptional behavior.

closed string propagating from left to right in fig. 39 is in the R-R sector, we want to investigate whether this type of closed-string degeneration, which is associated to R-R tadpoles, leads to infrared divergences and/or anomalies.

Whether Σ is a cylinder or a Mobius strip, its closed oriented double cover $\widehat{\Sigma}$ has genus 1. A spin structure on $\widehat{\Sigma}$ is specified by a pair of binary choices. We fix one of these choices by asking that the closed-string state propagating from left to right in fig. 39 should be in the R-R sector. This still leaves another binary choice to determine the spin structure of $\widehat{\Sigma}$. This second binary choice will determine whether the spin structure of $\widehat{\Sigma}$ is either odd or even. From a spacetime point of view, the resulting scattering amplitudes are parity-violating (odd spin structure on $\widehat{\Sigma}$) or parity-conserving (even spin structure). We will consider both cases.

If the spin structure of $\widehat{\Sigma}$ is odd, then both β and γ have a single zero-mode. This being so, the vacuum path integral $\langle 1 \rangle$ of the $\beta\gamma$ system is not well-defined. The simplest well-defined and non-zero path integral is obtained by inserting a single factor $\delta(\beta)$ and a single factor $\delta(\gamma)$ to remove the zero-modes:

$$(9.39) \quad \langle \delta(\gamma(z_1)) \delta(\beta(z_2)) \rangle \neq 0.$$

Instead of local operators $\delta(\gamma(z_1))$ and $\delta(\beta(z_2))$, we could similarly consider delta functions of integrated modes of γ or β . Any generic delta functions will remove the zero-modes and give a sensible and non-zero path integral. The same goes for other formulas below. If we replace $\delta(\gamma)$ in (9.39) by $\delta'(\gamma)$, the path integral vanishes:

$$(9.40) \quad \langle \delta'(\gamma(z_1)) \delta(\beta(z_2)) \rangle = 0.$$

One way to explain why is that the operator $\delta'(\gamma)$ has ghost number 1 less than that of $\delta(\gamma)$ (it transforms as λ^{-2} rather than λ^{-1} under $\gamma \rightarrow \lambda\gamma$) so (9.40) vanishes because the integrand has the wrong $\beta\gamma$ ghost number. We can restore the ghost number balance by adding an insertion of γ ; this gives another non-zero path integral:

$$(9.41) \quad \langle \delta'(\gamma(z_1)) \delta(\beta(z_2)) \gamma(z_3) \rangle \neq 0.$$

If the assertions (9.40) and (9.41) are not already clear, they will hopefully become so in section 10.

If the spin structure of $\widehat{\Sigma}$ is even, then neither β nor γ has a zero-mode. This being so, the $\beta\gamma$ vacuum amplitude $\langle 1 \rangle$ is well-defined and non-zero. But that will not really be relevant as we will always have $\delta(\beta)$ and $\delta(\gamma)$ insertions. The other statements in the last paragraph concerning which correlation functions are nonvanishing remain valid with an even spin structure.

The anticommuting ghost and antighost fields b and c are not sensitive to the spin structure. They both have one zero-mode, so the vacuum correlation function $\langle 1 \rangle$ vanishes, but the bc path integral with one b insertion and one c insertion is non-zero:

$$(9.42) \quad \langle b(z_1)c(z_2) \rangle \neq 0.$$

Overall, then, typical non-zero correlation functions for the $bc\beta\gamma$ ghost system are

$$(9.43) \quad \begin{aligned} \langle bc \delta(\beta) \delta(\gamma) \rangle &\neq 0 \\ \langle bc \delta(\beta) \delta'(\gamma) \gamma \rangle &\neq 0. \end{aligned}$$

We have written the formulas of the last few paragraphs in terms of holomorphic ghost fields $bc\beta\gamma$, but in the context of open and/or unoriented superstrings, one should remember that upon lifting to $\widehat{\Sigma}$, there is no difference between $bc\beta\gamma$ and their antiholomorphic counterparts $\widetilde{b}\widetilde{c}\widetilde{\beta}\widetilde{\gamma}$. Both holomorphic and antiholomorphic ghosts on Σ descend from holomorphic variables on $\widehat{\Sigma}$. So in (9.43), to write the formulas on Σ rather than $\widehat{\Sigma}$, one can for example replace b with \widetilde{b} or β with $\beta + \widetilde{\beta}$.

We will look for an infrared divergence (and later for an anomaly) in the scattering of n massless bosonic open-string states, with momenta $k^{(i)}$ and polarization vectors $\varepsilon^{(i)}$. It is convenient to represent $n - 1$ of the external

vertex operators, say $\mathcal{V}_2, \dots, \mathcal{V}_n$, in their integrated form.⁸¹ For example, in Type I superstring theory, the corresponding factor in the worldsheet path integral is

$$(9.44) \quad \prod_{i=2}^n \oint_{\partial_\ell \Sigma} [dz|d\theta] \varepsilon_I^{(i)} D_\theta X^I \exp(ik^{(i)} \cdot X).$$

(More generally, one should include Chan-Paton factors and restrict the integral to a region in which the \mathcal{V}_i are inserted on $\partial_\ell \Sigma$ in a definite cyclic order.) The only important property of this formula for what follows is that it does not depend on the ghosts. It is not convenient to try to represent the last vertex operator \mathcal{V}_1 in integrated form; because of the symmetries of Σ , the position at which \mathcal{V}_1 is inserted is not a modulus. So we use the unintegrated form of the vertex operator, which for Type I is

$$(9.45) \quad \mathcal{V}_1 = c\delta(\gamma)\varepsilon^{(1)} \cdot D_\theta X \exp(ik^{(1)} \cdot X).$$

Now we are ready to integrate over the moduli of Σ , to look for an infrared divergence and/or an anomaly. We have to integrate over the proper time s that the closed string propagates from left to right in fig. 39. The closed-string gluing parameter is $q = e^{-s}$. We also have to integrate over the sum of holomorphic and antiholomorphic gluing parameters. We do not have to integrate over either the argument of q or the difference of holomorphic and antiholomorphic gluing parameters, since these can be absorbed in automorphisms of Σ_r , as we have learned in sections 9.2.1 and 9.3.4. So we modify the usual construction of the closed string propagator by integrating over only s and the sum of fermionic gluing parameters. This gives⁸²

$$(9.46) \quad (b_0 + \tilde{b}_0) \int_{s_0}^\infty ds \exp(-s(L_0 + \tilde{L}_0)) \delta(\beta_0 + \tilde{\beta}_0)(G_0 + \tilde{G}_0).$$

⁸¹In general, the use of integrated vertex operators has a drawback that we have explained in sections 2.5.1 and 4.4: it makes it difficult to describe correctly the compactification of the moduli space. For our purposes here, there is no problem, because the use of integrated open-string vertex operators does not cause any trouble in understanding the closed-string R-R degeneration of interest. As long as the momenta $k^{(i)}$ are generic, the only problem that might be difficult to properly understand in terms of integrated open-string vertex operators is the open-string tadpole that may arise (in some models) when the insertion points of all n open-string vertex operators become coincident.

⁸²In the following, s_0 is an arbitrary lower cutoff in the integral over s . We are really only interested in a possible infrared divergence for $s \rightarrow \infty$.

Including also (9.45), the factors in the worldsheet path integral that depend on the ghosts are

$$(9.47) \quad c\delta(\gamma)\varepsilon^{(1)}\cdot D_\theta X \exp(ik^{(1)}\cdot X)(b_0+\tilde{b}_0)\int_{s_0}^{\infty} ds \exp(-s(L_0+\tilde{L}_0))\delta(\beta_0+\tilde{\beta}_0)(G_0+\tilde{G}_0).$$

Explicitly written in eqn. (9.47), though somewhat scattered, are the factors $bc\delta(\beta)\delta(\gamma)$ that we need to get a non-zero path integral (eqn. (9.43)). If we could drop the $G_0+\tilde{G}_0$ factor in eqn. (9.47), the integral over s would give the infrared divergence that one would naively expect from an R-R tadpole. However the factor of $G_0+\tilde{G}_0$ eliminates the apparent infrared divergence. We write

$$(9.48) \quad G_0 = G_0^X + G_0^{\text{gh}},$$

where G_0^X and G_0^{gh} are respectively the G_0 operators of the matter and ghost systems. Similarly

$$(9.49) \quad \tilde{G}_0 = \tilde{G}_0^X + \tilde{G}_0^{\text{gh}}.$$

The massless on-shell R-R state that contributes the tadpole⁸³ is annihilated by G_0^X and \tilde{G}_0^X . Nor do G_0^{gh} and \tilde{G}_0^{gh} help. We may as well consider G_0^{gh} , which has terms $b\gamma$ and $\beta\partial c$. Insertion of a single factor of either of these in (9.47) causes the path integral to vanish. For instance, an insertion of $b\gamma$ lowers the bc ghost number by 1 and raises the $\beta\gamma$ ghost number by 1; either of these shifts causes the path integral to vanish.

So there is no infrared divergence from a closed-string degeneration of R-R type. Now let us instead look for an anomaly. For this, we should replace one of the open-string vertex operators, say \mathcal{V}_1 , by $\{Q, \mathcal{W}_1\}$, where the gauge parameter \mathcal{W}_1 obeys the usual superconformal conditions. There is an anomaly if the resulting amplitude does not vanish. For massless open-string states from the NS sector, the appropriate gauge parameter⁸⁴ is

$$(9.50) \quad \mathcal{W}_1 = c\delta'(\gamma)\exp(ik^{(1)}\cdot X).$$

To evaluate the anomaly, according to section 7.1, we consider an amplitude with insertion of \mathcal{W}_1 instead of $\{Q, \mathcal{W}_1\}$, and we dispense with the integration

⁸³For Type I superstring theory, it was described explicitly in eqn. (9.27). It is annihilated by G_0^X and by \tilde{G}_0^X because these are some of the defining conditions of a superconformal vertex operator.

⁸⁴This comes from eqn. (C.3), with $\Phi_0 = \exp(ik^{(1)}\cdot X)$.

over s . Instead of integrating over s , we set s to a large value, and omit the factor $(b_0 + \tilde{b}_0)$ that is associated to the integral over s . So for computing anomalies rather than infrared divergences, eqn. (9.47) must be replaced by

$$(9.51) \quad c\delta'(\gamma) \exp(ik^{(1)} \cdot X) \exp(-s(L_0 + \tilde{L}_0))\delta(\beta_0 + \tilde{\beta}_0)(G_0 + \tilde{G}_0).$$

The anomaly is computed from a path integral with this insertion (multiplied by the factor (9.44) that does not involve the ghosts) in the limit of large s .

Now we are in the opposite situation from before. If we drop the $G_0 + \tilde{G}_0$ factor, the path integral will vanish since the bc and $\beta\gamma$ ghost numbers are both wrong. Comparing to the second line of (9.43), we see that to get a non-zero path integral, we need a b (or \tilde{b}) insertion and a γ (or $\tilde{\gamma}$) insertion. But G_0^{gh} contains a term $b\gamma$, and similarly \tilde{G}_0^{gh} contains a term $\tilde{b}\tilde{\gamma}$. So it is possible to get an anomaly.

Whether we actually get an anomaly depends on whether the contributions from G_0^{gh} and \tilde{G}_0^{gh} add or cancel. This depends on whether the spin structure of Σ is even or odd (which determines the sign in the mixing between $\tilde{\gamma}$ and γ). A more careful study of the boundary states [89], [84] shows that the anomaly comes from the odd spin structure. Hence, as one would expect from field theory, at one-loop order, the anomaly affects only the parity-violating amplitudes. The contribution from the odd spin structure vanishes if \mathfrak{n} is too small because of the zero-modes of the matter fermions $\psi^I = D_\theta X^I$, $I = 1, \dots, d$. Each of the $\mathfrak{n} - 1$ integrated vertex operators in (9.44) can absorb two of these d zero-modes, so an anomaly can only occur if $\mathfrak{n} - 1 \geq d/2$. In $d = 10$, the condition is $\mathfrak{n} \geq 6$ and the anomaly is traditionally called a hexagon anomaly.

10. More on the $\beta\gamma$ system

10.1. Preliminaries

The goal of the present section is to increase our comfort level with the commuting $\beta\gamma$ ghosts of superstring theory. As in [7, 46, 47] and as explained in section 3.2.2 above, we treat the $\beta\gamma$ path integral algebraically, using the rules of Gaussian integration, rather than literally trying to interpret it as an integral.

We particularly want to gain more experience with the delta function operators that are ubiquitous in superstring perturbation theory. Super Riemann surfaces will play no role; we work simply on an ordinary compact Riemann surface Σ of genus g . We assume that Σ is closed and oriented. The

generalization to open and/or unoriented surfaces is immediate, since the combined $\tilde{\beta}\tilde{\gamma}\beta\gamma$ system on an open and/or unoriented surface Σ is equivalent to a holomorphic $\beta\gamma$ system on the closed oriented double cover $\hat{\Sigma}$ of Σ .

Our basic approach is to compare the $\beta\gamma$ system to a system of fields β^* and γ^* that are identical in all respects except that they have the opposite statistics. So β and γ are governed by the action

$$(10.1) \quad I_{\beta\gamma} = \frac{1}{2\pi} \int_{\Sigma} d^2z \beta \partial_{\bar{z}} \gamma$$

and β^* and γ^* are governed by an identical-looking action

$$(10.2) \quad I_{\beta^*\gamma^*} = \frac{1}{2\pi} \int_{\Sigma} d^2z \beta^* \partial_{\bar{z}} \gamma^*.$$

We define an anomalous ghost number symmetry that assigns the values 1 and -1 to γ and β , respectively, and similarly for γ^* and β^* .

In superstring theory, to begin with β is a section of $K^{3/2}$ and γ a section of $K^{-1/2}$, where $K^{1/2}$ is a square root of the canonical line bundle K of Σ . However, the $\beta\gamma$ system makes sense more generally with γ a section of an arbitrary holomorphic line bundle \mathcal{L} , and β a section of $K \otimes \mathcal{L}^{-1}$. For example, if $\mathcal{L} = K^s$, then γ is a primary field of spin or dimension s and β is a primary of dimension $1 - s$. In superstring perturbation theory, once one includes Ramond punctures, one has to study the $\beta\gamma$ system with a variety of choices of \mathcal{L} , so we do not want to restrict to the case $\mathcal{L} = K^{-1/2}$.

The simplest case is actually $\mathcal{L} = K^{1/2}$, so that β and γ have spin $1/2$. This being so, if we chose $K^{1/2}$ to define an even spin structure, then for generic complex structure of Σ , neither β nor γ has zero-modes. This allows a minor simplification in the presentation, since it means that we do not need to introduce delta function operators (which can remove the zero-modes) at the very start. Ultimately, however, we certainly want to introduce and study them.

Let $M = \partial_{\bar{z}}/2\pi$ be the kinetic operator of the $\beta\gamma$ or $\beta^*\gamma^*$ system. (One often writes $\bar{\partial}$ for what we call $\partial_{\bar{z}}$.) The path integral of the $\beta^*\gamma^*$ system is a determinant:

$$(10.3) \quad \int \mathcal{D}\beta^* \mathcal{D}\gamma^* \exp(-I_{\beta^*\gamma^*}) = \det M.$$

Similarly, the path integral of the $\beta\gamma$ system is an inverse determinant:

$$(10.4) \quad \int \mathcal{D}\beta \mathcal{D}\gamma \exp(-I_{\beta\gamma}) = \frac{1}{\det M}.$$

The theory of determinants of differential operators, and their anomalies, is rather subtle, but we will assume that it is known, so we will not discuss the properties of the holomorphic object $\det M$. We focus here only on understanding the correlation functions of the $\beta\gamma$ system.

For $\mathcal{L} = K^{1/2}$, there generically are no β or γ zero-modes, but such modes may occur (for $g \geq 3$) as the complex structure of Σ is varied. When this happens, $\det M = 0$, so the partition function of the $\beta^*\gamma^*$ system vanishes. On the other hand, the partition function of the $\beta\gamma$ system acquires a pole. In superstring perturbation theory, such poles are called spurious singularities; they result from an incorrect gauge-fixing procedure. Superstring perturbation theory with a correct gauge-fixing procedure always leads to a sensible and finite $\beta\gamma$ path integral; when there are zero modes, there always are delta function operators that remove them.

It will be convenient to write $\langle 1 \rangle$ for the vacuum amplitude, or in other words the path integral with no operator insertion. We write $\langle \mathcal{O} \rangle$ for an unnormalized path integral with insertion of an operator (or product of operators) \mathcal{O} . The normalized expectation value $\langle \mathcal{O} \rangle_N$ is defined as the ratio

$$(10.5) \quad \langle \mathcal{O} \rangle_N = \frac{\langle \mathcal{O} \rangle}{\langle 1 \rangle}.$$

10.1.1. The importance of ghost number The algebraic treatment that we will give for the $\beta\gamma$ system would not work well for a bosonic system without a conserved ghost number symmetry. To see why, let us consider a bosonic Gaussian integral with finitely many variables x_1, \dots, x_n and a quadratic form $(x, Nx) = \sum_{a,b} N_{ab} x_a x_b$. With a suitable normalization of the measure $d^n x$, the Gaussian integral gives

$$(10.6) \quad \int d^n x \exp\left(-\frac{1}{2}(x, Nx)\right) = \frac{1}{\sqrt{\det N}}.$$

The square root poses an immediate problem. To pick the correct sign of the square root requires some use of calculus, not just algebra.

A bigger problem is that our main idea is of comparing a bosonic system to a fermionic one does not work well for an abstract Gaussian integral with no ghost number symmetry. Suppose that we replace the bosonic variables x_1, \dots, x_n with fermionic ones x_1^*, \dots, x_n^* . A quadratic action would have to be $(x^*, \tilde{N}x^*) = \sum_{a,b} \tilde{N}_{ab} x_a^* x_b^*$ where now \tilde{N} is an antisymmetric, rather than symmetric, bilinear form. But there is no natural way to convert a symmetric bilinear form to an antisymmetric one, so there is no natural map in general from a bosonic Gaussian integral to a fermionic one.

Both problems are removed if we assume the existence of a ghost number symmetry, with the x_a split into variables β_i of $N_{\text{gh}} = -1$ and variables γ_i of $N_{\text{gh}} = +1$. This corresponds to taking

$$(10.7) \quad N = \begin{pmatrix} 0 & M \\ M^t & 0 \end{pmatrix},$$

whereupon

$$(10.8) \quad \sqrt{\det N} = \det M,$$

and there is no problem with the square root. Moreover, there is now a natural map from a symmetric form N to an antisymmetric one \tilde{N} , namely

$$(10.9) \quad \tilde{N} = \begin{pmatrix} 0 & M \\ -M^t & 0 \end{pmatrix}.$$

Finally, there is no difficulty in this general context to understand the case that the ghost number symmetry is anomalous. This means that the number of variables with $N_{\text{gh}} = -1$ differs from the number of variables with $N_{\text{gh}} = 1$, so that the measure $d\beta_1 \dots d\gamma_s$ (or $d\beta_1^* \dots d\gamma_s^*$) transforms nontrivially under ghost number. That being so, only operator insertions that transform in an appropriate fashion under ghost number can have non-zero expectation values. In the bosonic case, as well as balancing the ghost number, the operator insertions will also have to remove the zero-modes of M that are inescapably present if the number of β 's and γ 's are unequal. Within the class of operators that we will consider in our algebraic treatment – polynomials and/or delta functions of the β 's and γ 's – the ones that can remove the zero-modes are the delta function operators. That is one reason that they are a crucial part of the formalism. Removing the correct number of zero-modes amounts to balancing the picture number as well as the ghost number.

10.2. Correlation functions for spin 1/2

10.2.1. Elementary fields As explained in section 10.1, we begin with the case that all fields have spin 1/2. The rules of Gaussian integration tell us how to compute the expectation value of a product of elementary fields β^* and γ^* or β and γ . The normalized two-point function is

$$(10.10) \quad \langle \gamma(u)\beta(w) \rangle_N = \langle \gamma^*(u)\beta^*(w) \rangle_N = S(u, w),$$

where $S(z, z')$, which is called the propagator, is the inverse of the kinetic operator $M = \partial_z/2\pi$. $S(z, z')$ can be understood as a section of the line bundle $K^{1/2} \boxtimes K^{1/2} \rightarrow \Sigma \times \Sigma$ (that is, $S(z, z')$ is a section of $K^{1/2}$ in each variable) with a pole of unit residue on the diagonal. It is antisymmetric under $z \leftrightarrow z'$ and is holomorphic away from the diagonal:

$$(10.11) \quad \frac{\partial_z}{2\pi} S(z, z') = \delta^2(z, z').$$

Our assumption that β and γ have no zero-modes ensures that $S(z, z')$ exists and is unique.

The expectation value of any product of elementary fields is directly constructed in the usual way from the propagator. For fermions

$$(10.12) \quad \begin{aligned} & \langle \gamma^*(u_1) \gamma^*(u_2) \dots \gamma^*(u_s) \beta^*(w_s) \beta^*(w_{s-1}) \dots \beta^*(w_1) \rangle_N \\ &= \sum_{\pi} (-1)^{\pi} \prod_{i=1}^s S(u_i, w_{\pi(i)}), \end{aligned}$$

where the sum runs over all permutations π of s objects, and $(-1)^{\pi}$ is 1 or -1 for even or odd permutations. More succinctly, the right hand side of (10.12) is $\det P_{(s)}$, where $P_{(s)}$ is the $s \times s$ matrix whose ij matrix element is $P_{(s)ij} = S(u_i, w_j)$. For bosons, we need not worry about the ordering of the factors. The correlation function is given by a formula like (10.12), but without the factor $(-1)^{\pi}$:

$$(10.13) \quad \left\langle \prod_{i=1}^s \gamma(u_i) \prod_{j=1}^s \beta(w_j) \right\rangle_N = \sum_{\pi} \prod_{i=1}^s S(u_i, w_{\pi(i)}).$$

10.2.2. Delta function operators Now we consider delta function operators. For fermions this is straightforward, since $\delta(\gamma^*) = \gamma^*$ and $\delta(\beta^*) = \beta^*$. Hence $\langle \delta(\gamma^*(u)) \delta(\beta^*(w)) \rangle_N = S(u, w)$. It will be more convenient to express this in terms of the unnormalized path integral:

$$(10.14) \quad \langle \delta(\gamma^*(u)) \delta(\beta^*(w)) \rangle = (\det M) S(u, w).$$

How can we go from here to a corresponding formula for $\langle \delta(\beta) \delta(\gamma) \rangle$ in the bosonic case? The idea is to express the unnormalized path integral $\langle \delta(\gamma^*(u)) \delta(\beta^*(w)) \rangle$ as a Gaussian integral. We introduce an anticommuting variable σ^* that we consider to have ghost number -1 , like β^* , and a variable

τ^* that has ghost number 1 like γ^* . Then we have⁸⁵

$$(10.15) \quad \int d\tau^* d\sigma^* \exp(-\sigma^* \gamma^*(u) - \beta^*(w) \tau^*) = \gamma^*(u) \beta^*(w) = \delta(\gamma^*(u)) \delta(\beta^*(w)).$$

Given this, we see that the unnormalized path integral (10.14) can actually be represented as a Gaussian integral

$$(10.16) \quad \langle \delta(\gamma^*(u)) \delta(\beta^*(w)) \rangle = \int \mathcal{D}\beta^* \mathcal{D}\gamma^* d\tau^* d\sigma^* \exp(-I_{\beta^* \gamma^*} - \sigma^* \gamma^*(u) - \beta^*(w) \tau^*).$$

It is convenient to combine all variables of ghost number 1 to $\widehat{\gamma}^* = (\gamma^*(z), \tau^*)$ and all variables of ghost number -1 to $\widehat{\beta}^* = (\beta^*(z), \sigma^*)$. We also define the extended action

$$(10.17) \quad \widehat{I}_{\widehat{\beta}^* \widehat{\gamma}^*} = I_{\beta^* \gamma^*} + \sigma^* \gamma^*(u) + \beta^*(w) \tau^*.$$

This function is a homogeneous and quadratic function of the full set of variables. It is convenient to view $\widehat{\beta}^*$ and $\widehat{\gamma}^*$ as row and column vectors

$$(10.18) \quad \widehat{\beta}^* = (\beta^*(z) \ \sigma^*), \quad \widehat{\gamma}^* = \begin{pmatrix} \gamma^*(z) \\ \tau^* \end{pmatrix},$$

and to view the extended action as $(\widehat{\beta}^*, \widehat{M} \widehat{\gamma}^*)$, where \widehat{M} is an extended version of M . The integral in (10.16) is simply a fermionic Gaussian integral with kinetic operator \widehat{M} , so

$$(10.19) \quad \langle \delta(\gamma^*(u)) \delta(\beta^*(w)) \rangle = \det \widehat{M}.$$

Comparing to eqn. (10.14), we see that we must have

$$(10.20) \quad \det \widehat{M} = (\det M) S(u, w).$$

The enterprising reader can verify this directly, thinking of \widehat{M} as a matrix that is obtained by adding one row and one column to the $\infty \times \infty$ matrix M .

It is now fairly obvious how to imitate this for bosons. We just drop the $*$'s everywhere and repeat all the steps with bosonic variables. We introduce

⁸⁵Signs are most simple if in writing the exponent, we place fields of ghost number -1 to the left of fields of ghost number $+1$, just as we did in defining the classical action $I_{\beta^* \gamma^*}$.

bosonic variables σ, τ of ghost numbers -1 and 1 and write

$$(10.21) \quad \delta(\gamma(u))\delta(\beta(w)) = \int d\tau d\sigma \exp(-\sigma\gamma(u) - \beta(w)\tau).$$

This representation of the delta function in the context of an algebraic treatment of Gaussian integrals was explained in the discussion of eqn. (3.43). So if as in the bosonic case we define the extended action

$$(10.22) \quad \widehat{I}_{\beta\gamma} = I_{\beta\gamma} + \sigma\gamma(u) + \beta(w)\tau,$$

then we can write $\langle \delta(\gamma(u))\delta(\beta(w)) \rangle$ as a Gaussian integral:

$$(10.23) \quad \langle \delta(\gamma(u))\delta(\beta(w)) \rangle = \int \mathcal{D}\widehat{\beta} \mathcal{D}\widehat{\gamma} \exp(-\widehat{I}_{\beta\gamma}) = \frac{1}{\det \widehat{M}}.$$

In view of (10.20), this is the same as $1/(\det M \cdot S(u, w))$. The normalized two-point function $\langle \delta(\gamma(u))\delta(\beta(w)) \rangle_N$ is therefore

$$(10.24) \quad \langle \delta(\gamma(u))\delta(\beta(w)) \rangle_N = \frac{1}{S(u, w)}.$$

This formula has many interesting consequences and generalizations. For one thing, we can immediately determine the dimensions of the operators $\delta(\beta)$ and $\delta(\gamma)$. The propagator $S(u, w)$ is the two-point function of operators β^* and γ^* of dimension $1/2$ (since β^* and γ^* are sections of $\mathcal{L} = K^{1/2}$); this is encoded in the way that it transforms under changes of coordinates. The inverse function $1/S(u, w)$ transforms oppositely, so if it is a two-point function, then it is the two-point function of operators of dimension $-1/2$. This should not be a surprise. Our algebraic treatment of the delta functions ensures that under scaling, $\delta(\beta)$ and $\delta(\gamma)$ transform oppositely to β and γ . But β and γ have conformal dimension $1/2$ so naturally $\delta(\beta)$ and $\delta(\gamma)$ have dimension $-1/2$. Similarly, we will see in section 10.2.4 that $\delta(\beta)$ and $\delta(\gamma)$ have opposite ghost numbers from β and γ .

The short distance behavior of $S(u, w)$ is $S(u, w) \sim 1/(u - w)$ for $u \rightarrow w$, corresponding to the OPE $\gamma^*(u)\beta^*(w) \sim 1/(u - w)$. So $1/S(u, w) \sim u - w$, giving the OPE

$$(10.25) \quad \delta(\gamma(u))\delta(\beta(w)) \sim u - w.$$

In general, a fermionic Gaussian integral vanishes if and only if there is a fermion zero-mode. Let us investigate this condition in the context of the

extended action $\widehat{I}_{\widehat{\beta}^* \widehat{\gamma}^*}$. We will look for a zero-mode of $\widehat{\gamma}^*$. (With $\mathcal{L} = K^{1/2}$, the index of the operator \widehat{M} vanishes, and there is a $\widehat{\gamma}^*$ zero-mode if and only if there is a $\widehat{\beta}^*$ zero-mode. We will write an explicit formula shortly.) The classical equations for $\widehat{\gamma}^*$ are

$$(10.26) \quad \begin{aligned} \gamma^*(u) &= 0 \\ \frac{\partial_z \gamma^*}{2\pi} + \delta^2(z, w) \tau^* &= 0. \end{aligned}$$

The second condition says that $\gamma^*(z)$ is holomorphic away from $z = w$, with at most a simple pole (of residue τ^*) at $z = w$. Taking account also of the first condition, we see that $\gamma^*(z)$ is a section of $K^{1/2}$ that vanishes at $z = u$ and may have a pole at $z = w$, but is otherwise holomorphic.

We can summarize all this by saying that $\gamma^*(z)$ is a holomorphic section of $K^{1/2} \otimes \mathcal{O}(w) \otimes \mathcal{O}(u)^{-1}$. Under what conditions does such a section exist? We have assumed that $K^{1/2}$ has no holomorphic sections, so upon allowing a pole at w , there is at most a one-dimensional space of holomorphic sections. Explicitly, such a section is the propagator $S(z, w)$, regarded as a function of z for fixed w . This section (with τ^* given by eqn. (10.26)) gives a $\widehat{\gamma}^*$ zero-mode if and only if $\gamma^*(z)$ vanishes at $z = u$, that is if and only if $S(u, w) = 0$. So $S(u, w) = 0$ is the condition under which the Gaussian integral of the $\widehat{\beta}^* \widehat{\gamma}^*$ system should vanish, and that is indeed what we see in eqn. (10.20). If and only if there is a $\widehat{\gamma}^*$ zero-mode, there is also a $\widehat{\beta}^*$ zero-mode, with $\beta^*(z)$ now given by $S(u, z)$ as a function of z for fixed u .

By the same token, in the case of bosons, the $\widehat{\beta} \widehat{\gamma}$ system has a zero-mode if and only if $S(u, w) = 0$, and that should be the condition under which the $\widehat{\beta} \widehat{\gamma}$ path integral is not well-defined. And this of course is what we see in eqn. (10.24). The pole in that formula when there are zero-modes is a typical example of what in superstring perturbation theory is called a spurious singularity. The $\beta \gamma$ path integral with $\delta(\beta)$ and $\delta(\gamma)$ insertions can in general develop a pole, but in superstring perturbation theory with a correct gauge fixing, the $\delta(\beta)$ and $\delta(\gamma)$ insertions are always such as to keep one away from the poles.

10.2.3. Multiple delta function insertions More generally, we would like to understand how to calculate a $\beta \gamma$ path integral with an arbitrary collection of delta function insertions $\langle \prod_{i=1}^s \delta(\gamma(u_i)) \prod_{j=1}^s \delta(\beta(w_j)) \rangle$. We can do this by the same reasoning as before. The corresponding fermionic correlation function $\langle \delta(\gamma^*(u_1)) \dots \delta(\gamma^*(u_s)) \delta(\beta^*(w_s)) \dots \delta(\beta^*(w_1)) \rangle$ is completely equivalent to the correlator (10.12) of a product of elementary fields. On the

other hand, by adding auxiliary variables σ_i^* , τ_i^* , $i = 1, \dots, s$, this fermionic correlation function can be expressed as a fermionic Gaussian integral:

(10.27)

$$\int \mathcal{D}\beta^* \mathcal{D}\gamma^* \prod_{i=1}^s d\tau_i^* d\sigma_i^* \exp \left(-I_{\beta^* \gamma^*} - \sum_i \sigma_i^* \gamma^*(u_i) - \sum_j \beta^*(w_j) \tau_j^* \right).$$

Dropping all the $*$'s, we get a bosonic Gaussian integral that computes $\langle \prod_{i=1}^s \delta(\gamma(u_i)) \prod_{j=1}^s \delta(\beta(w_j)) \rangle$. From this, we deduce that the normalized correlation function of a product of delta function operators is simply the inverse of (10.12):

$$(10.28) \quad \left\langle \prod_{i=1}^s \delta(\gamma(u_i)) \prod_{j=1}^s \delta(\beta(w_j)) \right\rangle_N = \frac{1}{\sum_{\pi} (-1)^{\pi} \prod_{i=1}^s S(u_i, w_{\pi(i)})}.$$

Just as for $s = 1$, the poles of this correlation function arise from zero-modes of the extended system in which it can be computed as a Gaussian integral. One will never encounter such poles in superstring perturbation theory with a correct gauge-fixing.

Eqn. (10.28) has a number of interesting consequences. First of all, this formula is antisymmetric in the u_i and also antisymmetric in the w_j , so we see that the local operators $\delta(\gamma)$ and also $\delta(\beta)$ are fermionic.

We have already analyzed the $\delta(\gamma) \cdot \delta(\beta)$ operator product in eqn. (10.25). Now we can similarly study the $\delta(\gamma) \cdot \delta(\gamma)$ and $\delta(\beta) \cdot \delta(\beta)$ operator products. For fermions, we have $\gamma^*(u_1) \gamma^*(u_2) \sim -(u_1 - u_2) \gamma^* \partial \gamma^*(u_2)$. (This statement is shorthand for saying that the product $\gamma^*(u_1) \gamma^*(u_2)$ is for $u_1 - u_2 \rightarrow 0$ the product of $u_1 - u_2$ with a dimension 2 primary that is conveniently denoted $\gamma^* \partial \gamma^*$. That the leading operator that appears in the product $\gamma^*(u_1) \gamma^*(u_2)$ is of dimension 2 can be deduced from (10.12).) Equivalently, for fermions $\delta(\gamma^*(u_1)) \cdot \delta(\gamma^*(u_2)) \sim -(u_1 - u_2) \delta(\gamma^*) \delta(\partial \gamma^*)$. Since the bosonic correlation function is the inverse of the fermionic one, it follows that the product $\delta(\gamma(u_1)) \cdot \delta(\gamma(u_2))$ for $u_1 \rightarrow u_2$ is the product of $1/(u_1 - u_2)$ times an operator that is conveniently denoted $-\delta(\gamma) \delta(\partial \gamma)$:

$$(10.29) \quad \delta(\gamma(u_1)) \delta(\gamma(u_2)) \sim -\frac{1}{u_1 - u_2} \delta(\gamma) \delta(\partial \gamma)(u_2).$$

Since $\delta(\gamma)$ has dimension $-1/2$, the factor of $1/(u_1 - u_2)$ implies that the operator $\delta(\gamma) \delta(\partial \gamma)$ has dimension -2 , the opposite of the dimension of the corresponding fermionic operator. Note that $\delta(\partial \gamma)$ is again fermionic, and in particular $\delta(\gamma) \delta(\partial \gamma) = -\delta(\partial \gamma) \delta(\gamma)$.

More generally, the leading singularity as t operators $\delta(\gamma)$ approach each other is a primary field of dimension $-t^2/2$ that can conveniently be denoted

$$(10.30) \quad \Theta_{-t} = \delta(\gamma)\delta(\partial\gamma) \dots \delta(\partial^{t-1}\gamma).$$

The justification for the notation is that near a point at which Θ_{-t} is inserted, the elementary field γ has a zero of order t . This can be established using formulas of section 10.2.4. Similarly, β has a pole of order t at an insertion point of this operator. In conventional language, the fact that γ has a zero of order t at a point at which Θ_{-t} is inserted while β has a pole of order t at such a point means that this operator is associated to the ground state in picture number $-t$. The ground state with positive picture number is similarly constructed from delta functions of β .

Eqns. (10.29) and (10.30) have obvious analogs with β replacing γ . The interested reader can deduce additional OPE relations from eqn. (10.28), such as

$$(10.31) \quad \delta(\gamma(u_1))\delta(\beta)\delta(\partial\beta)(u_2) \sim (u_1 - u_2)^2\delta(\beta(u_2)).$$

10.2.4. Mixed correlation functions So far we have understood how to compute correlation functions of elementary fields β and γ , and also of delta function insertions $\delta(\beta)$ and $\delta(\gamma)$. For superstring perturbation theory, in general one also needs mixed correlation functions such as $\langle\delta(\gamma(u))\delta(\beta(w))\gamma(u')\beta(w')\rangle$. These can be computed by slightly extending what we have described so far.

First of all, for fermions, since there is no difference between a delta function and an elementary field, this correlation function is just the correlation function (10.12) of a product of elementary fields. However, it will be useful to write (10.12) in terms of unnormalized path integrals, and also to break the symmetry between the operator insertions by writing two operators as delta functions and the other two as elementary fields:

$$(10.32) \quad \begin{aligned} &\langle\delta(\gamma^*(u))\delta(\beta^*(w))\gamma^*(u')\beta^*(w')\rangle \\ &= \det M (S(u, w)S(u', w') - S(u, w')S(u', w)). \end{aligned}$$

We will now proceed as in eqn. (10.15), introducing new variables τ^* and σ^* and using the exponential representation of the delta functions. As before, we introduce extended variables $\widehat{\gamma}^* = (\gamma^*(z), \tau^*)$ and $\widehat{\beta}^* = (\beta^*(z), \sigma^*)$, and define the extended action $\widehat{\mathcal{I}}_{\widehat{\beta}^*\widehat{\gamma}^*}$. Then we can write the correlation function

that we want as a two-point function in the extended Gaussian theory with this extended action:

$$(10.33) \quad \langle \delta(\gamma^*(u)) \delta(\beta^*(w)) \gamma^*(u') \beta^*(w') \rangle = \int \mathcal{D}\widehat{\beta}^* \mathcal{D}\widehat{\gamma}^* \exp(-\widehat{I}_{\widehat{\beta}^*, \widehat{\gamma}^*}) \gamma^*(u') \beta^*(w').$$

This is simply a Gaussian integral with insertion of elementary fields $\gamma^*(u')$ and $\beta^*(w')$. So

$$(10.34) \quad \langle \delta(\gamma^*(u)) \delta(\beta^*(w)) \gamma^*(u') \beta^*(w') \rangle = (\det \widehat{M}) \widehat{S}(u', w'),$$

where as before \widehat{M} is the extended kinetic operator, and $\widehat{S}(u', w')$ is a matrix element of the inverse operator to \widehat{M} , which we will call the extended propagator.

Using the formula (10.20) for $\det \widehat{M}$, and eqn. (10.32) for the correlation function in (10.34), we deduce a formula for the relevant matrix elements of the extended propagator:

$$(10.35) \quad \widehat{S}(u', w') = \frac{1}{S(u, w)} (S(u, w)S(u', w') - S(u, w')S(u', w)).$$

Now it is straightforward to understand the analog of this for bosons. To compute a mixed correlation function $\langle \delta(\gamma(u)) \delta(\beta(w)) \gamma(u') \beta(w') \rangle$, we introduce the new variables σ and τ and use the integral representation of the delta functions. In this way, we arrive at (10.33) without the $*$'s. Performing the Gaussian integral gives the same result as before, except that in the case of a bosonic Gaussian integral, the determinant appears in the denominator:

$$(10.36) \quad \langle \delta(\gamma(u)) \delta(\beta(w)) \gamma(u') \beta(w') \rangle = \frac{1}{\det \widehat{M}} \widehat{S}(u', w').$$

Upon using (10.20) and (10.35) and also multiplying by $\det M$ to pass to normalized correlation functions, we get finally

$$(10.37) \quad \begin{aligned} & \langle \delta(\gamma(u)) \delta(\beta(w)) \gamma(u') \beta(w') \rangle_N \\ &= \frac{1}{S(u, w)^2} (S(u, w)S(u', w') - S(u, w')S(u', w)). \end{aligned}$$

Given this, we can work out the OPE of an elementary field times a delta function operator. For example, eqn. (10.37) has a simple zero for $u' \rightarrow u$.

This means that the leading contribution to the product of $\gamma(u')$ and $\delta(\gamma(u))$ is an operator of dimension 1 that is naturally understood as $\partial\gamma \cdot \delta(\gamma)$:

$$(10.38) \quad \gamma(u')\delta(\gamma(u)) \sim (u' - u) \partial\gamma \cdot \delta(\gamma)(u).$$

Similarly, eqn. (10.37) has a simple pole for $w' \rightarrow u$. This means that the leading contribution to the product $\beta(w') \cdot \delta(\gamma(u))$ involves an operator of dimension -1 . It is natural to interpret this operator as $\delta'(\gamma(u))$:

$$(10.39) \quad \beta(w')\delta(\gamma(u)) \sim \frac{1}{w' - u} \delta'(\gamma(u)).$$

The justification for this notation comes from the OPE of γ and $\delta'(\gamma)$, which one can compute by letting u' approach u in (10.37) after having let w' approach u to define the operator that we are calling $\delta'(\gamma)$. After we extract the pole of (10.39) as $w' \rightarrow u$, there is no further singularity as $u' \rightarrow u$, but there is a minus sign because of the sign of the second term in (10.37). Apart from this minus sign, the result of taking $u' \rightarrow u$ after $w' \rightarrow u$ gives back the two-point function $\langle \delta(\gamma(u)) \delta(\beta(w)) \rangle_N$ with the operators $\gamma(u')$ and $\beta(w')$ simply omitted, so⁸⁶

$$(10.40) \quad \gamma(u')\delta'(\gamma(u)) \sim -\delta(\gamma(u)).$$

The name $\delta'(\gamma)$ for the operator in question is an intuitive way to express this fact. Eqns. (10.38) and (10.39), taken together, mean in the terminology of [1] that the operator $\delta(\gamma)$ is the vertex operator of the $\beta\gamma$ vacuum of picture number -1 . Similarly, we have

$$(10.41) \quad \begin{aligned} \gamma(u')\delta(\beta(w)) &\sim \frac{1}{u' - w} \delta'(\beta(w)) \\ \beta(w')\delta(\beta(w)) &\sim (w' - w) \partial\beta \cdot \delta(\beta)(w). \end{aligned}$$

These formulas mean that, in conventional language, $\delta(\beta)$ is the vertex operator of the $\beta\gamma$ ground state at picture number $+1$. Similarly, the operator in eqn. (10.30) represents the $\beta\gamma$ vacuum with picture number $-t$, and its cousin with γ replaced by β represents the vacuum with picture number $+t$.

⁸⁶To fully verify that the operator on the right hand side of this OPE is indeed $-\delta(\gamma(u))$ rather than a new operator that we have not seen before, we really should extend this analysis to consider a limit $u' \rightarrow u$ after $w' \rightarrow u$ in an arbitrary correlation function containing multiple insertions of $\delta(\gamma)$, $\delta(\beta)$, γ , and β , not just the example considered in the text. This is fairly straightforward given what we have explained, and is left to the reader.

To determine the ghost number of the operators $\delta(\gamma)$ and $\delta(\beta)$, we can proceed as follows. The ghost number current of the $\beta\gamma$ system is defined as the normal-ordered expression $J_{\beta\gamma}(u') = - : \beta\gamma : (u')$. Taking the limit as $w' \rightarrow u'$ in (10.37) after subtracting out the pole term to define the normal-ordered operator J , we get the three-point function $\langle \delta(\gamma(u))\delta(\beta(w))J_{\beta\gamma}(u') \rangle$. This three-point function has simple poles as $u' \rightarrow u$ or $u' \rightarrow w$. The poles come from the term $-S(u, w')S(u', w)$ in the numerator of (10.37). The residues of these poles determine the ghost numbers of the operators $\delta(\gamma)$ and $\delta(\beta)$. A quick and largely error-proof way to proceed is to observe that we could compute the ghost numbers of the elementary fields γ and β in the same way starting with the four-point function

$$(10.42) \quad \langle \gamma(u) \beta(w) \gamma(u') \beta(w') \rangle = S(u, w)S(u', w') + S(u, w')S(u', w).$$

Again from the behavior for $w' \rightarrow u'$, we extract the three-point function $\langle \gamma(u) \beta(w) J_{\beta\gamma}(u') \rangle$, and then from the residues of the poles at $u' = u$ or $u' = w$, one finds the ghost numbers of the elementary fields γ and β . The only material difference in the two computations is that the crucial term $S(u, w')S(u', w)$ appears with opposite sign in (10.42) relative to (10.37). As a result, the operators $\delta(\gamma)$ and $\delta(\beta)$ have opposite ghost numbers from γ and β ; thus $\delta(\gamma)$ and $\delta(\beta)$ have ghost numbers -1 and $+1$, respectively.

Once this is known, the OPE's given above suffice to determine the ghost numbers of the other operators that we have encountered. For example, (10.39) implies that $\delta'(\gamma)$ has ghost number -2 and (10.29) implies that $\delta(\gamma)\delta(\partial\gamma)$ also has ghost number -2 .

Now that we have encountered operators such as $\delta'(\gamma)$, it is instructive to analyze the operator product in (10.29) in higher orders. To shorten the formulas, we replace u_1 and u_2 by u and 0 , and also write γ' and γ'' for $\partial\gamma$ and $\partial^2\gamma$. For $u \rightarrow 0$, we have $\gamma(u) = \gamma(0) + u\gamma'(0) + \frac{u^2}{2}\gamma''(0) + \dots$, so

$$(10.43) \quad \delta(\gamma(u))\delta(\gamma(0)) = \delta \left(\gamma(0) + u\gamma'(0) + \frac{u^2}{2}\gamma''(0) + \dots \right) \delta(\gamma(0)).$$

Since $\delta(a + b)\delta(a) = \delta(b)\delta(a)$ and $\delta(\lambda b) = \delta(b)/\lambda$, this is

$$(10.44) \quad \begin{aligned} \delta \left(\gamma(0) + u\gamma'(0) + \frac{u^2}{2}\gamma''(0) + \dots \right) \delta(\gamma(0)) \\ = \delta \left(u\gamma'(0) + \frac{u^2}{2}\gamma''(0) + \dots \right) \delta(\gamma(0)) \end{aligned}$$

$$= \frac{1}{u} \delta \left(\gamma'(0) + \frac{u}{2} \gamma''(0) + \dots \right) \delta(\gamma(0)).$$

Now we simply expand $\delta(\gamma'(0) + \frac{u}{2} \gamma''(0) + \dots) = \delta(\gamma'(0)) + \frac{u}{2} \gamma''(0) \delta'(\gamma'(0)) + \dots$ to get

$$(10.45) \quad \delta(\gamma(u)) \delta(\gamma(0)) \sim \frac{1}{u} \delta(\gamma') \delta(\gamma)(0) + \frac{1}{2} \gamma'' \delta'(\gamma') \delta(\gamma)(0) + \dots,$$

where the expansion can be straightforwardly carried out to any desired order. All these manipulations are valid in the context of an algebraic treatment of Gaussian integrals.

10.2.5. Extended delta functions In superstring perturbation theory, one encounters more general delta functions that are not necessarily local operators. Let f and g be $(0, 1)$ -forms on Σ with values in $K^{1/2}$ and define

$$(10.46) \quad \gamma_f = \int_{\Sigma} f(z) \gamma(z), \quad \beta_g = \int_{\Sigma} g(z) \beta(z).$$

We would like to calculate corresponding correlation functions such as $\langle \delta(\gamma_f) \delta(\beta_g) \rangle_N$.

For fermions, this would be completely straightforward. With γ_f^* and β_g^* defined by the obvious analogs of eqn. (10.46), we have

$$(10.47) \quad \langle \delta(\gamma_f^*) \delta(\beta_g^*) \rangle_N = \langle \gamma_f^* \beta_g^* \rangle_N = \int_{\Sigma \times \Sigma} f(z) S(z, z') g(z').$$

Repeating the derivation in section 10.2.2, we find that the normalized bosonic correlation function is the inverse of this:

$$(10.48) \quad \langle \delta(\gamma_f) \delta(\beta_g) \rangle_N = \frac{1}{\int_{\Sigma \times \Sigma} f(z) S(z, z') g(z')}.$$

In superstring perturbation theory, one meets operators $\delta(\beta_g)$ for arbitrary g , though usually one only encounters $\delta(\gamma_f)$ for the case that f is a delta function. All our statements about correlation functions with multiple delta function insertions or with insertions of elementary fields as well as delta function operators have fairly immediate analogs in the presence of nonlocal operators such as $\delta(\beta_g)$ and/or $\delta(\gamma_f)$.

If we specialize eqn. (10.48) to the case that f and g are delta functions, we recover eqn. (10.24). More generally, if f and g are derivatives of delta

functions, we get formulas such as

$$(10.49) \quad \langle \delta(\gamma(u)) \delta(\partial\beta(w)) \rangle_N = \frac{1}{\partial_w S(u, w)}.$$

Here $\delta(\partial\beta)$ is a local operator, though not a primary field. (In fact, a Virasoro module containing a state corresponding to $\delta(\partial\beta)$ is not a highest weight module and does not contain a primary state. So $\delta(\partial\beta)$ is also not a descendant. Still, it is possible to make sense of its correlation functions. By contrast, the operator $\delta(\beta)\delta(\partial\beta)$ is primary.)

Hopefully the $\beta\gamma$ system with spin 1/2 does not retain much mystery. The generalization that we need for superstring perturbation theory is only a short step, to which we turn next.

10.3. The $\beta\gamma$ system with zero-modes

Now we generalize the $\beta\gamma$ system to the case that γ is a section of a holomorphic line bundle \mathcal{L} and β is a section of $K \otimes \mathcal{L}^{-1}$. The main novelty is that depending on the choice of \mathcal{L} , β or γ may have zero-modes. According to the Riemann-Roch theorem, the number of γ zero-modes minus the number of β zero-modes is $n_\gamma - n_\beta = 1 - g + \text{deg } \mathcal{L}$, where $\text{deg } \mathcal{L}$ is the degree of \mathcal{L} . For a generic choice of the moduli of Σ or \mathcal{L} , the total number of zero-modes is the minimum required by the Riemann-Roch theorem, which means that generically either $n_\gamma = 0$ or $n_\beta = 0$. Correlation functions have poles at values of the moduli at which the number of zero-modes exceeds the minimum required by the Riemann-Roch theorem, so it is natural to begin with the generic case. Ordinarily,⁸⁷ in superstring perturbation theory, \mathcal{L} is such that $n_\beta \neq 0$, $n_\gamma = 0$. In the absence of Ramond punctures, one has $\mathcal{L} = K^{-1/2}$, with $\text{deg } \mathcal{L} = 1 - g$, $n_\gamma = 0$, $n_\beta = 2g - 2$. With Ramond punctures, $\text{deg } \mathcal{L}$ becomes more negative, n_β becomes larger, and n_γ remains zero.

For these reasons, in developing the theory, we will assume that $n_\gamma = 0$, $n_\beta > 0$. Since γ and β zero-modes are respectively holomorphic sections of \mathcal{L} and of $K \otimes \mathcal{L}^{-1}$, our hypothesis means that $H^0(\Sigma, \mathcal{L}) = 0$, $H^0(\Sigma, K \otimes$

⁸⁷There are a few exceptional cases. For $g = 0$ and no more than 2 Ramond punctures, one has $n_\gamma \neq 0$, $n_\beta = 0$. The $\beta\gamma$ system can be described exactly as we do below, with the roles of β and γ reversed. We actually do this for $n_R = 2$ in section 10.3.4 below. For $g = 1$ and no Ramond punctures, in an odd spin structure, $n_\beta = n_\gamma = 1$. What we will say can be adapted to this case, starting with the fact that for the $\beta^*\gamma^*$ system, the vacuum amplitude vanishes in this example because of fermion zero modes, but $\langle \gamma^*(u)\beta^*(w) \rangle \neq 0$.

$\mathcal{L}^{-1}) \neq 0$. According to Serre duality, $H^0(\Sigma, \mathcal{L})$ is dual to $H^1(\Sigma, K \otimes \mathcal{L}^{-1})$, and therefore our hypothesis implies that

$$(10.50) \quad H^1(\Sigma, K \otimes \mathcal{L}^{-1}) = 0.$$

10.3.1. The minimal delta function insertion Setting $n_\beta = t$, let us suppose that β has zero-modes y_1, \dots, y_t , obeying

$$(10.51) \quad M^t y_i(z) = 0, \quad i = 1, \dots, t,$$

with⁸⁸ $M^t = -\partial_z/2\pi$. This being so, the vacuum path integral of the $\beta\gamma$ system is not well-defined. To make sense of it, we need delta function operators that will remove the zero-modes from the path integral.

As usual, it is easier to begin with fermions. So we consider a $\beta^*\gamma^*$ system with γ^* and β^* being fermionic variables that are sections respectively of \mathcal{L} and of $K \otimes \mathcal{L}^{-1}$. The vacuum amplitude $\langle 1 \rangle$ vanishes because of the zero modes $y_i(z)$. A minimal non-zero correlation function is $\langle \beta^*(w_1) \dots \beta^*(w_t) \rangle$, with enough β^* insertions to absorb the zero-modes.

The correlation function $\langle \beta^*(w_1) \dots \beta^*(w_t) \rangle$ will be annihilated by ∂_w in each of its arguments. This ensures that it is a linear combination of expressions $\prod_{i=1}^t y_{j_i}(w_i)$ for some sequence j_1, \dots, j_t . But fermi statistics imply that $\langle \beta^*(w_1) \dots \beta^*(w_t) \rangle$ must be antisymmetric in w_1, \dots, w_t . This implies that it must be proportional to $\det N_{(t)}$, where $N_{(t)}$ is the $t \times t$ matrix whose ij matrix element is $N_{(t)ij} = y_i(w_j)$. The constant of proportionality is by definition $\det' M$, the determinant of M in the space of $\beta\gamma$ fields with the zero-modes divided out. So for fermions

$$(10.52) \quad \langle \beta^*(w_1) \dots \beta^*(w_t) \rangle = \det' M \det N.$$

As usual, we can write this more suggestively as

$$(10.53) \quad \langle \delta(\beta^*(w_1)) \dots \delta(\beta^*(w_t)) \rangle = \det' M \det N.$$

The reader will probably now not be surprised that the result for bosons is simply the inverse of this:

$$(10.54) \quad \langle \delta(\beta(w_1)) \dots \delta(\beta(w_t)) \rangle = \frac{1}{\det' M} \frac{1}{\det N}.$$

⁸⁸The transpose of the operator $M = \partial_z/2\pi$ acting on γ is the operator $M^t = -\partial_z/2\pi$ acting on β .

To deduce this, we simply use the fact that the left-hand side of (10.53) can be written as a Gaussian integral. As in eqn. (10.27), we add variables τ_i^* , $i = 1, \dots, t$, of ghost number 1, and write

$$(10.55) \quad \delta(\beta^*(w_1)) \dots \delta(\beta^*(w_t)) = \int d\tau_1^* \dots d\tau_t^* \exp\left(-\sum_i \beta(w_i)\tau_i^*\right).$$

Letting $\widehat{\gamma}^* = (\gamma^*(z), \tau_1^*, \dots, \tau_t^*)$, we define the extended action

$$(10.56) \quad \widehat{I}_{\beta^*\widehat{\gamma}^*} = I_{\beta^*\gamma^*} + \sum_{i=1}^t \beta^*(w_i)\tau_i^*,$$

and then

$$(10.57) \quad \langle \delta(\beta^*(w_1)) \dots \delta(\beta^*(w_t)) \rangle = \int \mathcal{D}\beta^* \mathcal{D}\widehat{\gamma}^* \exp(-\widehat{I}_{\beta^*\widehat{\gamma}^*}).$$

Now that we have expressed the correlation function of interest as a Gaussian integral, the generalization to bosons is straightforward. We simply drop the *'s everywhere, introducing bosonic variables τ_1, \dots, τ_t and the extended action

$$(10.58) \quad \widehat{I}_{\beta\widehat{\gamma}} = I_{\beta\gamma} + \sum_{i=1}^t \beta(w_i)\tau_i,$$

where $\widehat{\gamma} = (\gamma(z), \tau_1, \dots, \tau_t)$. Then

$$(10.59) \quad \langle \delta(\beta(w_1)) \dots \delta(\beta(w_t)) \rangle = \int \mathcal{D}\beta \mathcal{D}\widehat{\gamma} \exp(-\widehat{I}_{\beta\widehat{\gamma}}).$$

The result of the bosonic Gaussian integral is inverse to the fermionic one, so we arrive at (10.54).

This formula shows that $\langle \delta(\beta(w_1)) \dots \delta(\beta(w_t)) \rangle$ has a pole precisely when the matrix N has a non-zero kernel. This occurs precisely if a non-zero linear combination of the zero-modes y_1, \dots, y_t vanishes at all the points w_1, \dots, w_t , so that the insertion of the delta functions $\delta(\beta(w_1)) \dots \delta(\beta(w_t))$ does not remove all of the zero-modes from the functional integral.

10.3.2. More general correlation functions To understand more general correlation functions, we again begin with fermions.

Consider a correlation function with $t+1$ insertions $\beta^*(w_i)$, $i = 1, \dots, t+1$ and a single insertion $\gamma^*(u)$. We would like to compute the correlation function

$$(10.60) \quad \langle \gamma^*(u) \beta^*(w_1) \dots \beta^*(w_{t+1}) \rangle.$$

As a function of any one variable w_i , this correlation function is a holomorphic section of $K \otimes \mathcal{L}^{-1}$ except for a simple pole at $w_i = u$. Differently put, the correlation function in its dependence on w_i is a holomorphic section of $K \otimes \mathcal{L}^{-1} \otimes \mathcal{O}(u)$.

The space $H^0(\Sigma, K \otimes \mathcal{L}^{-1} \otimes \mathcal{O}(u))$ has dimension $t + 1$. This follows⁸⁹ from the Riemann-Roch theorem plus the assumption (10.50). Of the $t + 1$ linearly independent holomorphic sections of $K \otimes \mathcal{L}^{-1} \otimes \mathcal{O}(u)$, t are the modes y_1, \dots, y_t that actually come from holomorphic sections of $K \otimes \mathcal{L}^{-1}$ with no pole at u . To get a basis of $H^0(\Sigma, K \otimes \mathcal{L}^{-1} \otimes \mathcal{O}(u))$, we need one more section of $K \otimes \mathcal{L}^{-1}$ that actually does have a pole at u . We can denote that last section as $f_u(w)$. It is convenient to ask that in its dependence on u , $f_u(w)$ should be a section of \mathcal{L} ; we denote it as $S(u, w)$, a section of $\mathcal{L} \boxtimes (K \otimes \mathcal{L}^{-1}) \rightarrow \Sigma \times \Sigma$ (that is, $S(u, w)$ is a section of \mathcal{L} in the first variable and a section of $K \otimes \mathcal{L}^{-1}$ in the second) with a pole on the diagonal. We can constrain $S(u, w)$ by requiring that the residue of the pole at $w = u$ is 1 (the fact that this condition makes sense is the reason to define $S(u, w)$ as a section of \mathcal{L} in its dependence on the first variable); once we do this, $S(u, w)$ is uniquely determined modulo the possibility of adding a linear combination of the zero-modes y_i with u -dependent coefficients

$$(10.61) \quad S(u, w) \rightarrow S(u, w) + \sum_i h_i(u) y_i(w).$$

$S(u, w)$ is the closest analog of the propagator, given the presence of the zero-modes y_i .

Now we can analyze the correlation function (10.60) by the same reasoning that led to (10.52). As a function of any of the w_j , this correlation function is a linear combination of the $t + 1$ modes $y_i(w_j)$, $i = 1, \dots, t$, and $S(u, w_i)$. Fermi statistics implies that as a function of the $t + 1$ variables w_1, \dots, w_{t+1} , the correlation function must be, up to a multiplicative constant, the ‘‘Slater

⁸⁹The Riemann-Roch theorem says that $\dim H^0(\Sigma, K \otimes \mathcal{L}^{-1} \otimes \mathcal{O}(u)) - \dim H^1(\Sigma, K \otimes \mathcal{L}^{-1} \otimes \mathcal{O}(u)) = t + 1$. But $H^1(\Sigma, K \otimes \mathcal{L}^{-1} \otimes \mathcal{O}(u))$ vanishes by virtue of (10.50) and the long exact cohomology sequence derived from the short exact sequence of sheaves $0 \rightarrow K \otimes \mathcal{L}^{-1} \rightarrow K \otimes \mathcal{L}^{-1} \otimes \mathcal{O}(u) \rightarrow \mathcal{L}^{-1}|_u \rightarrow 0$.

determinant” constructed from those $t + 1$ modes. This Slater determinant is by definition the determinant of the $t + 1 \times t + 1$ matrix

$$(10.62) \quad N_{(t+1)} = \begin{pmatrix} y_1(w_1) & y_1(w_2) & \dots & y_1(w_{t+1}) \\ y_2(w_1) & y_2(w_2) & \dots & y_2(w_{t+1}) \\ & & \ddots & \\ y_t(w_1) & y_t(w_2) & \dots & y_t(w_{t+1}) \\ S(u, w_1) & S(u, w_2) & \dots & S(u, w_{t+1}) \end{pmatrix}.$$

Note that this determinant is not affected by the nonuniqueness (10.61) of $S(u, w)$ (and therefore we do not need to ask whether there is a global choice of $S(u, w)$).

The constant multiplying $\det N_{(t+1)}$ in the correlation function is precisely $\det' M$, the same factor that appeared in (10.52). To show this, we simply take the limit that one of the w_i approaches u , and use the condition that the residue of $S(u, w)$ at $u = w$ is 1. So finally, we get a formula for the correlation function (10.60):

$$(10.63) \quad \langle \gamma^*(u) \beta^*(w_1) \dots \beta^*(w_{t+1}) \rangle = \det' M \det N_{(t+1)}.$$

There are two ways to extrapolate from this formula to a corresponding formula for bosons. In one approach, we replace all operators β^* or γ^* with $\delta(\beta^*)$ or $\delta(\gamma^*)$:

$$(10.64) \quad \langle \delta(\gamma^*(u)) \delta(\beta^*(w_1)) \dots \delta(\beta^*(w_{t+1})) \rangle = \det' M \det N_{(t+1)}.$$

In the now familiar fashion, by using an exponential representation of the delta functions, one can express the left hand side as a fermionic Gaussian integral. The analogous correlation function for bosons can be expressed as the corresponding bosonic Gaussian integral, so for the $\beta\gamma$ system, we have

$$(10.65) \quad \langle \delta(\gamma(u)) \delta(\beta(w_1)) \dots \delta(\beta(w_{t+1})) \rangle = \frac{1}{\det' M} \frac{1}{\det N_{(t+1)}}.$$

The second thing we can do is to replace only the fields $\beta^*(w_1), \dots, \beta^*(w_t)$ in (10.63) with delta functions, and leave $\gamma^*(u)$ and $\beta^*(w_{t+1})$ as elementary fields. In this case, it is convenient to consider a ratio of (10.63) and the corresponding correlator (10.53) with $\gamma^*(u)$ and $\beta^*(w_{t+1})$ omitted:

$$(10.66) \quad \frac{\langle \gamma^*(u) \delta(\beta^*(w_1)) \dots \delta(\beta^*(w_t)) \beta^*(w_{t+1}) \rangle}{\langle \delta(\beta^*(w_1)) \dots \delta(\beta^*(w_t)) \rangle} = \frac{\det N_{(t+1)}}{\det N_{(t)}}.$$

We can regard the left hand side as the propagator in the fermionic Gaussian theory of eqn. (10.56). As a propagator, this ratio is unchanged in replacing fermions by bosons. So for the $\beta\gamma$ system

$$(10.67) \quad \frac{\langle \gamma(u) \delta(\beta(w_1)) \dots \delta(\beta(w_t)) \beta(w_{t+1}) \rangle}{\langle \delta(\beta(w_1)) \dots \delta(\beta(w_t)) \rangle} = \frac{\det N_{(t+1)}}{\det N_{(t)}}.$$

Using (10.54) for the denominator, we find

$$(10.68) \quad \langle \gamma(u) \delta(\beta(w_1)) \dots \delta(\beta(w_t)) \beta(w_{t+1}) \rangle = \frac{\det N_{(t+1)}}{\det' M (\det N_{(t)})^2}.$$

But actually, for some purposes, the normalized two-point function (10.67) of the operators $\gamma(u)$ and $\beta(w_{t+1})$ in the presence of the given delta function insertions is more useful. Once one has this formula, the correlation function of an arbitrary collection of elementary fields in the presence of the delta function insertions follows via the general free-field formula of eqn. (10.13).

It is straightforward to extend all this to a correlation function with additional delta function insertions. Given any s points $u_1, \dots, u_s \in \Sigma$, and $t + s$ points w_1, \dots, w_{t+s} , we define a $(t + s) \times (t + s)$ Slater determinant $\det N_{(t+s)}$, where

$$(10.69) \quad N_{(t+s)} = \begin{pmatrix} y_1(w_1) & y_1(w_2) & \dots & y_1(w_{t+s}) \\ y_2(w_1) & y_2(w_2) & \dots & y_2(w_{t+s}) \\ & & \ddots & \\ y_t(w_1) & y_t(w_2) & \dots & y_t(w_{t+s}) \\ S(u_1, w_1) & S(u_1, w_2) & \dots & S(u_1, w_{t+s}) \\ & & \ddots & \\ S(u_s, w_1) & S(u_s, w_2) & \dots & S(u_s, w_{t+s}) \end{pmatrix}.$$

Then

$$(10.70) \quad \langle \gamma^*(u_1) \dots \gamma^*(u_s) \beta^*(u_1) \dots \beta^*(u_{t+s}) \rangle = \det' M \det N_{(t+s)}.$$

For bosons, this extrapolates to a generalization of eqn. (10.65) with $s > 1$:

$$(10.71) \quad \langle \delta(\gamma(u_1)) \dots \delta(\gamma^*(u_s)) \delta(\beta^*(u_1)) \dots \delta(\beta^*(u_{t+s})) \rangle = \frac{1}{\det' M} \frac{1}{\det N_{(t+s)}}.$$

After replacing t and s by $t + 1$ and $s + 1$, it also extrapolates to the following generalization of eqn. (10.68) for the two-point function of elementary fields

β and γ in the presence of any number of delta function insertions:

$$(10.72) \quad \begin{aligned} & \langle \delta(\gamma(u_1))\delta(\gamma(u_2)) \dots \delta(\gamma(u_s))\gamma(u_{s+1}) \delta(\beta(w_1)) \dots \delta(\beta(w_{t+s})) \beta(w_{t+s+1}) \rangle \\ &= \frac{\det N_{(t+s+1)}}{\det' M (\det N_{(t+s)})^2}. \end{aligned}$$

The normalized correlation function of β and γ in the presence of the delta function insertions is

$$(10.73) \quad \begin{aligned} & \frac{\langle \delta(\gamma(u_1))\delta(\gamma(u_2)) \dots \delta(\gamma(u_s))\gamma(u_{s+1}) \delta(\beta(w_1)) \dots \delta(\beta(w_{t+s})) \beta(w_{t+s+1}) \rangle}{\langle \delta(\gamma(u_1))\delta(\gamma(u_2)) \dots \delta(\gamma(u_s)) \delta(\beta(w_1)) \dots \delta(\beta(w_{t+s})) \rangle} \\ &= \frac{\det N_{(t+s+1)}}{\det N_{(t+s)}}. \end{aligned}$$

This formula again has an immediate extension to the case of a larger number of elementary field insertions in the presence of the same delta functions. One just uses eqn. (10.13) to express the $2n$ -point functions of elementary free fields in terms of the two-point functions.

From these general formulas for correlation functions, one can evaluate the operator products of the general $\beta\gamma$ system, for an arbitrary collection of elementary fields and delta function operators. As in the spin 1/2 case, a variety of interesting additional operators appear in this operator product expansion. The operators built from products of elementary fields and delta functions are all of the local operators of the $\beta\gamma$ system that one needs in superstring perturbation theory. (They correspond to the operators of the “small Hilbert space” in the language of [1].) It is convenient in superstring perturbation theory to also include nonlocal delta function operators described in section 10.2.5; as we have seen, the extension of the theory to include such operators is straightforward.

10.3.3. Spin fields? At this point, the reader may well ask what has happened to the usual spin fields [1] of superstring perturbation theory. The answer to this question is that in fact, the spin fields are hidden in what we have already explained. (See section 5.3.2 for related remarks.)

In superstring perturbation theory, in the absence of Ramond insertions, the line bundle \mathcal{L} is $K^{-1/2}$, or more precisely it is endowed with an isomor-

phism $\mathcal{L}^2 \cong K^{-1}$. In the presence of Ramond punctures supported at points⁹⁰ $p_1, \dots, p_{n_R} \in \Sigma$, the isomorphism becomes

$$(10.74) \quad \mathcal{L}^2 \cong K^{-1} \otimes_{i=1}^{n_R} \mathcal{O}(-p_i).$$

In particular, with arbitrary Ramond insertions, the $\beta\gamma$ system is the system that we have already described, with a particular choice of \mathcal{L} . There is no need to discuss separately a case in which spin fields are included.

Still, the reader may feel that something is missing from what we have said so far. In fact, what is missing is not a discussion of spin fields, but a discussion of reparametrization invariance.

10.3.4. Reparametrization invariance For general \mathcal{L} , the question of reparametrization invariance does not arise. Given an abstract holomorphic line bundle \mathcal{L} over a Riemann surface Σ , there is no natural way to lift local reparametrizations of Σ to symmetries of \mathcal{L} . For this question to make sense, \mathcal{L} must be defined in terms of the geometry of Σ , such that we know how reparametrizations of Σ should act on \mathcal{L} .

For example, this is the case if $\mathcal{L} = K^s$ for some s ; reparametrizations of Σ lift in a natural way to an action on K and also on K^s .

The space of local operators of the $\beta\gamma$ system is independent of \mathcal{L} , in the following sense. A local operator is a locally-defined notion, and since any \mathcal{L} is locally trivial, the choice of \mathcal{L} cannot affect the space of local operators. Likewise, operator product relations can be computed locally and thus are independent of \mathcal{L} .

What depends on \mathcal{L} is the way that reparametrizations act on the local operators. Let us consider, for example, the vertex operator (10.30) that represents the ground state of the $\beta\gamma$ system at picture number $-t$:

$$(10.75) \quad \Theta_{-t} = \delta(\gamma)\delta(\partial\gamma) \dots \delta(\partial^{t-1}\gamma).$$

It can be defined for any \mathcal{L} . The question of its conformal dimension makes sense if \mathcal{L} is of the form $\mathcal{L} = K^s$ for any s , but the answer certainly depends on s . Indeed, γ has dimension s . Hence, reasoning classically, $\delta(\gamma)$ has dimension $-s$, and $\delta(\partial^k\gamma)$ has dimension $-s-k$, so Θ_{-t} has dimension $-ts - t(t-1)/2$. These formulas are also valid quantum mechanically, for the following reasons. The fact that $\delta(\gamma)$ has the opposite conformal dimension from γ follows from

⁹⁰We consider a split super Riemann surfaces so that the positions of the Ramond divisors can be described by points in an ordinary Riemann surface Σ . This assumption does not affect the local structure, which is of interest here.

the usual mapping from the $\beta^*\gamma^*$ system to the $\beta\gamma$ system. Given this, the conformal dimension of Θ_{-t} can be deduced from operator product relations such as eqn. (10.29), leading to the classical result $-ts - t(t-1)/2$. Ultimately, the reason that classical formulas for the dimensions are valid is that we are looking at operators made only from γ and there is no singularity in the $\gamma \cdot \gamma$ operator product. In superstring perturbation theory in the absence of Ramond punctures, this discussion applies with $s = -1/2$ and hence the dimension of Θ_{-t} is $-t(t-2)/2$.

Now let us discuss what happens in the presence of a Ramond puncture. In the presence of a Ramond puncture at a point p , \mathcal{L} is endowed with an isomorphism

$$(10.76) \quad \mathcal{L}^2 \cong K^{-1} \otimes \mathcal{O}(-p).$$

(We do not specify the positions of other Ramond punctures, since we will be looking at the local behavior near p .) The line bundle $\mathcal{O}(-p)$ is canonically trivial away from p ; it is trivialized by the section 1 (understood as a section of $\mathcal{O}(-p)$ with a simple pole at p). So away from p , \mathcal{L} is presented with an isomorphism to $K^{-1/2}$, and therefore it makes sense to ask how reparametrizations of the punctured z -plane act on the operators of the $\beta\gamma$ system. The generators of such reparametrizations comprise the Virasoro algebra, so in particular the Virasoro algebra acts on the operators that can be inserted at p . This action, however, is different from what it would be in the absence of the Ramond puncture at p .

To spell this out a little more fully, we will compare two descriptions of the $\beta\gamma$ system in the presence of a Ramond puncture – the same two descriptions that were considered in section 5.3. Pick a local holomorphic coordinate z that vanishes at the point p . Then the line bundle $K^{-1} \otimes \mathcal{O}(-p) \cong T \otimes \mathcal{O}(-p)$ is trivialized locally by the section $z^{-1}\partial_z$. \mathcal{L} can be trivialized locally by a section s whose square maps to $z^{-1}\partial_z$ under the isomorphism (10.76). This condition determines s only up to sign. Make a choice of sign and denote s as $(z^{-1}\partial_z)^{1/2}$. In the notation of section 5.3, we write the ghost field γ near $z = 0$ as

$$(10.77) \quad \hat{\gamma}(z)(z^{-1}\partial_z)^{1/2},$$

where the function $\hat{\gamma}(z)$ is single-valued and regular near $z = 0$. In this description, there is no spin operator at $z = 0$. The more conventional approach is to take advantage of the fact that $\mathcal{O}(-p)$ is canonically trivial when the point $z = 0$ is omitted, so on the complement of that point, the line bundle

$K^{-1} \otimes \mathcal{O}(-p)$ is isomorphic to $K^{-1} \cong T$, and is trivialized by the section ∂_z . So away from $z = 0$, the isomorphism (10.76) means that \mathcal{L} can be trivialized locally by a section s^\diamond whose square maps to ∂_z . Again, s^\diamond is locally determined up to sign, but now there is no consistent way to define the sign; s^\diamond has a monodromy around $z = 0$, since

$$(10.78) \quad s = \frac{1}{z^{1/2}} s^\diamond,$$

and s has no monodromy. Let us write $(\partial_z)^{1/2}$ for s^\diamond . The conventional description of the $\beta\gamma$ system is to write γ near $z = 0$ as

$$(10.79) \quad \gamma^\diamond(z)(\partial_z)^{1/2}.$$

Clearly, the function $\gamma^\diamond(z)$ has a monodromy around $z = 0$. $\gamma^\diamond(z)$ is usually denoted simply as $\gamma(z)$ (and we did so in section 5.3).

The advantage of the description by $\gamma^\diamond(z)$ (and its conjugate $\beta^\diamond(z)$) is that the action of reparametrizations is clear. Away from $z = 0$, the $\beta^\diamond\gamma^\diamond$ system is a standard $\beta\gamma$ system of spins $(3/2, -1/2)$; the local operators and operator product relations, the stress tensor, and the transformation of local operators under reparametrizations are all the same as if there were no Ramond puncture at $z = 0$. Of course, the local operators at $z = 0$ look exotic from this point of view; they are spin fields. However, the $\beta^\diamond\gamma^\diamond$ description gives a very effective way to determine how those spin fields transform under reparametrizations; for one approach relying on bosonization, see [1], and for another approach, relying on a knowledge of how the stress tensor is defined away from $z = 0$, see section 5.3.1 above. In the $\widehat{\beta}\widehat{\gamma}$ description, there is nothing unusual about the operators that are inserted at $z = 0$, but it is less obvious how to compute their behavior under reparametrization.

10.3.5. Another interpretation of the delta function insertions We will conclude by explaining another interpretation of the formula (10.73) for the two-point function of elementary fields γ and β in the presence of arbitrary delta function insertions. We simplify notation by writing u and w instead of u_{s+1} and w_{t+s+1} , and we write the left hand side of eqn. (10.73) as $\langle \gamma(u)\beta(w) \rangle_{N,\delta}$, where the notation is meant to indicate a normalized two-point function in the presence of a collection of delta function insertions:

$$(10.80) \quad \langle \gamma(u)\beta(w) \rangle_{N,\delta} = \frac{\det N_{(t+s+1)}}{\det N_{(t+s)}}.$$

We want to consider this as a function of u and w with the locations of the delta function insertions held fixed. This means that the denominator is a constant, and all zeroes and poles come from the numerator $\det N_{(t+s+1)}$.

This Slater determinant has a simple pole at $u = w$, with a coefficient such that the residue of $\langle \gamma(u)\beta(w) \rangle_{N,\delta}$ at $u = w$ is 1. It also has simple poles at $u = w_i, i = 1, \dots, t + s$, and simple zeroes at $u = u_i, i = 1, \dots, s$, with no other zeroes or poles. These conditions define the normalized two-point function $\langle \gamma(u)\beta(w) \rangle_{N,\tilde{\mathcal{L}}}$, with no delta functions at all but with the line bundle \mathcal{L} that enters the definition of the $\beta\gamma$ system replaced by another line bundle

$$(10.81) \quad \tilde{\mathcal{L}} = \mathcal{L} \otimes_{i=1}^s \mathcal{O}(-u_i) \otimes_{j=1}^{t+s} \mathcal{O}(w_j).$$

Thus, the $\beta\gamma$ system defined with a line bundle \mathcal{L} and an arbitrary collection of delta function insertions is simply equivalent to a $\beta\gamma$ system with a different line bundle $\tilde{\mathcal{L}}$ and no delta function insertions at all.

Hopefully the bosonic delta function operators of superstring perturbation theory do not retain much mystery.

Appendix A. Pullback of differential forms

Suppose that a group G acts on a manifold Y with quotient $M = Y/G$. Let $y_i, i = 1, \dots, n = \dim Y$ be local coordinates on Y . A differential form on Y is a function $F(y_1 \dots y_n | dy_1 \dots dy_n)$.

F is said to be a pullback from M if it can be written as $F(x_1 \dots x_m | dx_1 \dots dx_m)$, where $x_i, i = 1, \dots, m = \dim M$ are local coordinates on M . We can pick the local coordinates of Y to be the x_i and also local coordinates $f_i, i = 1, \dots, n - m$ on the fibers of the fibration $Y \rightarrow M$. For F to be a pullback from M means that it is independent of the f_i and of the df_i .

This is equivalent to the following two conditions:

- (1) F must be G -invariant.
- (2) F must be annihilated by contraction with any of the vector fields that generate the action of G .

Concretely, G is generated by vector fields of the general form

$$(A.1) \quad V_a = \sum_{i=1}^{n-m} v_{a,i} \frac{\partial}{\partial f_i}$$

and the corresponding contraction operators are

$$(A.2) \quad \mathbf{i}_{V_a} = \sum_{i=1}^{n-m} v_{a,i} \frac{\partial}{\partial \mathbf{d}f_i}.$$

The condition that $\mathbf{i}_{V_a} F = 0$, $a = 1, \dots, n - m$ means precisely that F is independent of the $\mathbf{d}f_i$, and then the condition that F is G -invariant means that it is also independent of the f_i . (The matrix $v_{a,i}$ is invertible if G acts freely on M , but in any event the vector fields V_a span the tangent space to the fibers of the projection $Y \rightarrow M$, so the condition that $\mathbf{i}_{V_a} F = 0$ for all a means that F is independent of all $\mathbf{d}f_i$.)

This argument is equally applicable if n and/or m is infinite. In the application in section 2.1, we have $Y = \mathcal{J}$, the space of all complex structures on a given surface Σ of genus g ; $M = \mathcal{M}_g$ the moduli space of Riemann surfaces; and $G = \mathcal{D}$, the group of orientation-preserving diffeomorphisms of Σ . So n is infinite but m is finite. The argument also generalizes without difficulty to supermanifolds and supergroups, a fact we exploit in section 3. In the context of supermanifolds, the reasoning applies to forms of any picture number.

Appendix B. Bosonic string gauge parameters

B.1. Examples

In bosonic string theory, in compactification to \mathbb{R}^d with $d \geq 2$, every physical state of non-zero momentum is associated to a primary field V of dimension $(1, 1)$ constructed from matter fields only.⁹¹ However, in general this representation is not unique; V can be shifted by adding to it a null vector.

A primary field V is said to be a null vector if it is a Virasoro descendant, meaning that it is of the form

$$(B.1) \quad V = \sum_{n=1}^{\infty} L_{-n} U_n,$$

with Virasoro generators L_{-n} and some states U_n . (In practice, if V is a state of definite momentum, the sum over n is a finite sum since there is a lower bound on the possible dimension of U_n .) If V is of this form, then according to the no-ghost theorem, the corresponding BRST-invariant vertex operator

⁹¹This statement and some others below are part of the BRST version of the no-ghost theorem; see [33–36], or, for example, section 4.4 of [32]. For the original no-ghost theorem, see [38, 37].

$\mathcal{V} = cV$ is a BRST commutator, $\mathcal{V} = \{Q_B, \mathcal{W}\}$ for some \mathcal{W} , and should decouple from the S -matrix.⁹²

As we saw in section 2.4.3, to analyze this decoupling, it helps to know that we can assume that

$$(B.2) \quad b_n \mathcal{W} = 0, \quad n \geq 0,$$

or equivalently that \mathcal{W} is constructed using only c and not its derivatives. Before presenting any general theory, we will describe the first few examples.

The first case of a null vector is

$$(B.3) \quad V = L_{-1} \Phi_0,$$

where Φ_0 is a matter primary of dimension 0. In this case, $\mathcal{V} = cV$ obeys $\mathcal{V} = Q_B \mathcal{W}$, with

$$(B.4) \quad \mathcal{W} = \Phi_0.$$

So \mathcal{W} does not involve c at all. The next case is

$$(B.5) \quad V = \left(L_{-2} + \frac{3}{2} L_{-1}^2 \right) \Phi_{-1},$$

where Φ_{-1} is a matter primary of dimension -1 . In this case $\mathcal{V} = cV$ is of the form $Q_B \mathcal{W}$ with

$$(B.6) \quad \mathcal{W} = bc\Phi_{-1} + \frac{3}{2} L_{-1} \Phi_{-1}.$$

So \mathcal{W} depends on c , though not on its derivatives.

B.2. General proof

The BRST version of the no-ghost theorem says that if V is a Virasoro primary of dimension 1 constructed from matter fields that is also null, meaning that it is a Virasoro descendant in the sense of eqn. (B.1), then $\mathcal{V} = cV$ is of the form $Q_B \mathcal{W}$ for some \mathcal{W} . We want to show that in general, not just in the examples described above, we can pick \mathcal{W} so that $b_n \mathcal{W} = 0$, $n \geq 0$.

⁹²For brevity, in this appendix, we consider either open strings or a chiral sector of closed strings, so we omit the antiholomorphic ghosts \tilde{c} . Also, motivated by the operator-state correspondence of conformal field theory, we write just $Q_B \mathcal{W}$ instead of $\{Q_B, \mathcal{W}\}$.

There is actually a surprising shortcut to this conclusion, provided by the appendix to [36]. If V is a dimension 1 primary that can be expanded as in eqn. (B.1), then actually this expansion can be drastically shortened and put in the very special form

$$(B.7) \quad V = L_{-1}\Phi_0 + \left(L_{-2} + \frac{3}{2}L_{-1}^2\right)\Phi_{-1},$$

where Φ_0 and Φ_{-1} are primaries of the indicated dimension. In other words, a dimension 1 primary V that is null is always a linear combination of null vectors of the sort that we analyzed in appendix B.1. So we can use equations (B.4) and (B.6) to explicitly write $\mathcal{V} = cV$ in the desired form $\mathcal{V} = \{Q, \mathcal{W}\}$ where \mathcal{W} is the primary

$$(B.8) \quad \mathcal{W} = \Phi_0 + \left(bc + \frac{3}{2}L_{-1}\right)\Phi_{-1}.$$

We have actually learned more than we needed. We aimed to prove that \mathcal{W} can be chosen not to depend on the derivatives of c ; we have learned that it can be chosen to depend on derivatives of neither b nor c . The significance of this is unclear.

The considerations in [36] are closely related to the no-ghost theorem, and here we will give an alternative argument that is closely related to the BRST version of the no-ghost theorem and especially to the proofs given in [33] and in section 4.4 of [32]. This argument will have two steps: (i) a general argument to show that we can assume that \mathcal{W} does not involve $\partial^n c$, $n \geq 2$; (ii) a more special argument to show that we can assume that \mathcal{W} also does not involve ∂c .

Step (ii) goes as follows. Associated to every matter primary V of dimension 1, there are a dual pair of Q_B -invariant vertex operators, namely $\mathcal{V} = cV$ and $\mathcal{V}^* = c\partial cV$. The no-ghost theorem shows that \mathcal{V} is trivial in Q_B -cohomology if and only if \mathcal{V}^* is (and if and only if V is a descendant). Now suppose that $\mathcal{V}^* = Q_B\mathcal{W}^*$ for some \mathcal{W}^* . After averaging over the compact group generated by L_0 , we can assume that $L_0\mathcal{W}^* = 0$; the whole analysis that follows will be made in the subspace with $L_0 = 0$. Our reasoning in step (i) will apply equally to \mathcal{V}^* and \mathcal{V} and show that, if $\mathcal{V}^* = Q_B\mathcal{W}^*$ for some \mathcal{W}^* , we can assume that $b_n\mathcal{W}^* = 0$, $n > 0$. Then setting $\mathcal{W} = -b_0\mathcal{W}^*$, we see that $b_n\mathcal{W} = 0$, $n \geq 0$. Moreover $Q_B\mathcal{W} = -Q_Bb_0\mathcal{W}^* = b_0Q_B\mathcal{W}^* = b_0\mathcal{V}^* = \mathcal{V}$. So we have found an operator \mathcal{W} with the desired properties $Q_B\mathcal{W} = \mathcal{V}$ and $b_n\mathcal{W} = 0$, $n \geq 0$.

As for step (i), we define a grading on the space of operators by assigning degree 0 to matter fields as well as b , c , ∂c , and derivatives of b , but degree 1 to

$\partial^n c$, $n \geq 2$. We call this grading the c'' -degree. So for example, the operator $bc\partial X\partial^2 c\partial^5 c$ has c'' -degree 2, with a contribution of 1 from $\partial^2 c$ and from $\partial^5 c$ and no contribution from b , c , or ∂X . (The name c'' -degree is motivated by the fact that we are counting the number of times that the second or higher derivative of c appears in an operator.) The c'' -degree of any operator is non-negative.

We can write the BRST operator Q_B as

$$(B.9) \quad Q_B = Q_{>} + Q_0,$$

where $Q_{>}$ increases the c'' -degree by 1, and Q_0 leaves it unchanged. The equation $Q_B^2 = 0$ implies that $Q_{>}^2 = 0$, so we can define the cohomology of $Q_{>}$.

Moreover, since $Q_{>}$ increases the c'' -degree by a definite amount (namely 1), we can define the cohomology of $Q_{>}$ for states of any given c'' -degree. We will show below that this cohomology vanishes except for c'' -degree 0. Let us first explain how this implies what we want.

We assume that there exists some \mathcal{W} such that

$$(B.10) \quad Q_B \mathcal{W} = \mathcal{V},$$

where \mathcal{V} has c'' -degree 0. (So \mathcal{V} can be either cV or $c\partial cV$, where V is constructed from matter fields only.) We assume that \mathcal{V} has energy-momentum k , and we can require that \mathcal{W} has the same energy-momentum. Suppose that the expansion of \mathcal{W} in operators of definite c'' -degree is $\mathcal{W} = \sum_{r=0}^t \mathcal{W}_r$ and to begin with suppose that t , which is the maximal c'' -degree of any term in \mathcal{W} , is greater than 0. (t is finite since the c'' -degree of states of fixed energy-momentum and $L_0 = 0$ is bounded.) Since the right hand side of (B.10) has c'' -degree 0, (B.10) implies that $Q_{>} \mathcal{W}_t = 0$. If it is true that the cohomology of $Q_{>}$ vanishes for c'' -degree greater than 0, it follows that $\mathcal{W}_t = Q_{>} \mathcal{U}_{t-1}$ for some \mathcal{U}_{t-1} of c'' -degree $t - 1$. This being so, we can replace \mathcal{W} by $\mathcal{W} - Q_B \mathcal{U}_{t-1}$ without disturbing eqn. (B.10); so we reduce to the case that the maximum c'' -degree of any term in \mathcal{W} is $t - 1$. Continuing in this way, we reduce to the case that \mathcal{W} has maximum c'' -degree 0. At this point, we have $b_n \mathcal{W} = 0$, $n > 0$ (since the operators b_n with $n > 0$ lower the c'' -degree), and we have completed step (i).

It remains to show that the cohomology of $Q_{>}$ vanishes as claimed for positive c'' -degree. For this we will imitate the proof of the no-ghost theorem, as presented in section 4.4 of [32], whose notation and reasoning we follow as closely as possible.

The argument assumes that the matter sector of the theory has at least two free fields X^0 and X^1 , with Lorentz metric $-(dX^0)^2 + (dX^1)^2$. We use a lightcone basis $X^\pm = (X^0 \pm X^1)/\sqrt{2}$. The corresponding oscillators α_m^\pm , $m \in \mathbb{Z}$, obey $[\alpha_m^+, \alpha_n^-] = -m\delta_{m+n}$, with other commutators vanishing. We introduce the operator

$$(B.11) \quad N_*^{\text{lc}} = \sum_{m \geq 1} \frac{1}{m} \alpha_{-m}^+ \alpha_m^-,$$

which counts minus the number of + excitations and has eigenvalues $0, -1, -2, \dots$ (Our N_*^{lc} is the relevant half of the definition in eqn. (4.4.8) of [32].) $Q_{>}$ has a decomposition

$$(B.12) \quad Q_{>} = Q_{>,1} + Q_{>,0} + Q_{>,-1},$$

where $[N_*^{\text{lc}}, Q_{>,j}] = jQ_{>,j}$, so that $Q_{>,j}$ shifts N_*^{lc} by j units. The fact that $Q_{>}^2 = 0$ implies that $Q_{>,1}^2 = 0$, so we can compute its cohomology.

Assuming that the energy-momentum k carried by \mathcal{V} is non-zero, we can pick our coordinates so that $k^+ \neq 0$. (Because of this step, the analysis given here fails for $k = 0$. This case requires special treatment and has unusual properties; see section 7.2.2.) Explicitly

$$(B.13) \quad Q_{>,1} = -(2\alpha')^{1/2} k^+ \sum_{m \geq 1} \alpha_m^- c_{-m}.$$

(From Q_1 as defined in eqn. (4.4.13) of [32], we have omitted the terms that reduce the c'' -degree.) A convenient way⁹³ to compute the cohomology of $Q_{>,1}$ is to define

$$(B.14) \quad R = \frac{1}{(2\alpha')^{1/2} k^+} \sum_{m \geq 1} \alpha_{-m}^+ b_m,$$

which reduces N_*^{lc} by 1. We also define

$$(B.15) \quad S = \{Q_{>,1}, R\} = \sum_{m=1}^{\infty} (m c_{-m} b_m - \alpha_{-m}^+ \alpha_m^-),$$

⁹³Since $Q_{>,1}$ is bilinear in oscillators, and commutes with L_0 , one can also simply compute its cohomology via a mode-by-mode analysis. We will proceed that way in the superstring case, though it would also be possible to imitate the argument involving R and S .

which commutes with N_*^{lc} .

The operator S is positive semi-definite. Its kernel consists of states annihilated by b_m and also by α_m^- , with $m > 0$. In particular, states annihilated by S have c'' -degree 0.

Since S commutes with $Q_{>,1}$, the cohomology of $Q_{>,1}$ can be decomposed in the eigenspaces of S . $Q_{>,1}$ annihilates the kernel of S , as is clear from the description of that kernel in the last paragraph. On the other hand, the cohomology of $Q_{>,1}$ vanishes for $S \neq 0$. Indeed, if $S\Psi_0 = s\Psi_0$ with $s \neq 0$, and $Q_{>,1}\Psi_0 = 0$, then $\Psi_0 = s^{-1}\{Q_{>,1}, R\}\Psi_0 = Q_{>,1}(s^{-1}R\Psi_0)$, so Ψ_0 vanishes in the cohomology of $Q_{>,1}$.

These statements imply that the cohomology of $Q_{>,1}$ is supported at c'' -degree 0. Now let us examine the cohomology of $Q_{>}$.

We will give an argument similar to one above. Suppose that \mathcal{Y} is an element of the cohomology of $Q_{>}$ at c'' -degree $q > 0$, and let $\mathcal{Y} = \sum_{r=-t}^{-t'} \mathcal{Y}_r$ be the expansion of \mathcal{Y} in states \mathcal{Y}_r of $N_*^{\text{lc}} = r$. This expansion is a finite sum since N_*^{lc} is bounded above and below in the space of states of fixed energy-momentum with $L_0 = 0$; in fact, $-t' \leq 0$ since N_*^{lc} is bounded above by 0. If $Q_{>}\mathcal{Y} = 0$, then $Q_{>,1}\mathcal{Y}_{-t'} = 0$. Since the cohomology of $Q_{>,1}$ vanishes at positive c'' -degree, there is a state $\mathcal{U}_{-t'-1}$ of $N_*^{\text{lc}} = -t' - 1$ (and with c'' -degree 1 less than that of \mathcal{Y}) such that $Q_{>,1}\mathcal{U}_{-t'-1} = \mathcal{Y}_{-t'}$. Without changing the $Q_{>}$ cohomology class of \mathcal{Y} , we can replace \mathcal{Y} by $\mathcal{Y} - Q_{>}\mathcal{U}_{-t'-1}$, whose expansion in eigenstates of N_*^{lc} now runs over eigenvalues that are bounded above by $-t' - 1$. After repeating this process finitely many times, we eventually reach the lower bound on N_*^{lc} in the chosen sector and reduce to the case $\mathcal{Y} = 0$. So the $Q_{>}$ cohomology vanishes for positive c'' -degree.

Appendix C. Superstring gauge parameters

C.1. Examples

The goal here is to generalize the results of appendix B to superstring theory. We begin by describing the first few examples of null vectors and gauge parameters in superstring theory.

In the Neveu-Schwarz sector, a physical state that should decouple from the S -matrix is associated to a primary field V of dimension 1/2 constructed from the matter system that is also null. In other words, V is a superconformal descendant, that is it can be written

$$(C.1) \quad V = \sum_{n>0} L_{-n}^X W_n + \sum_{r>0} G_{-r}^X \Lambda_r,$$

where L_n^X, G_r^X are the superconformal generators of the matter system and $W_n, \Lambda_r, n, r > 0$, are some states of the matter system. The first such null vector arises at the massless level and takes the form

$$(C.2) \quad V = G_{-1/2}^X \Phi_0,$$

where Φ_0 is a matter primary of dimension 0. The corresponding superconformal vertex operator $\mathcal{V} = c\delta(\gamma)V$ can be written $\mathcal{V} = \{Q_B, \mathcal{W}\}$, with

$$(C.3) \quad \mathcal{W} = c\delta'(\gamma)\Phi_0.$$

In terms of modes, if $|-1\rangle$ represents the $\beta\gamma$ ground state of picture number -1 , obeying

$$(C.4) \quad \beta_r |-1\rangle = \gamma_r |-1\rangle = 0, \quad r > 0,$$

then

$$(C.5) \quad \mathcal{W} = -c_1\beta_{-1/2} |-1\rangle \otimes \Phi_0.$$

\mathcal{W} is annihilated by b_n and $\beta_r, n, r \geq 0$.

The first example of gauge-invariance for massive Neveu-Schwarz states is associated to a level 1 null vector

$$(C.6) \quad V = \left(G_{-3/2}^X + 2G_{-1/2}^X L_{-1}^X\right) \Phi_{-1}$$

of the matter system. Here Φ_{-1} is a matter primary of dimension -1 . The corresponding superconformal vertex operator $\mathcal{V} = c\delta(\gamma)V$ is $\mathcal{V} = \{Q_B, \mathcal{W}\}$, with

$$(C.7) \quad \mathcal{W} = \left(\delta(\gamma)G_{-1/2}^X - c\beta\delta(\gamma) + c\delta'(\gamma)L_{-1}^X\right) \Phi_{-1}.$$

In modes

$$(C.8) \quad \mathcal{W} = \left(G_{-1/2}^X - c_1\beta_{-3/2} - c_1\beta_{-1/2}L_{-1}^X\right) |-1\rangle \otimes \Phi_{-1}.$$

\mathcal{W} is annihilated by the modes $b_n, \beta_r, n, r \geq 0$.

In the Ramond sector, a physical state that should decouple from the S -matrix is associated to a primary field V of dimension $5/8$ constructed from the matter system only that is a descendant in the sense that it can be written as in (C.1), again with $n, r > 0$. (Now r takes integer values. We do

not allow a term $G_0^X \Lambda_0$ in the sum in (C.1), or else we would be claiming that the usual massless fermions should decouple.) There are no massless string states of this form. The first non-trivial example arises at the first massive level of the Ramond sector and is

$$(C.9) \quad V = \left(L_{-1}^X - \frac{1}{2} G_{-1}^X G_0^X \right) \Phi,$$

where Φ is a matter primary of dimension $-3/8$. The corresponding vertex operator is $\mathcal{V} = c\Theta_{-1/2}V$, where $\Theta_{-1/2}$ is the $-1/2$ picture spin field of the $\beta\gamma$ system, obeying

$$(C.10) \quad \begin{aligned} \beta_n \Theta_{-1/2} &= 0, \quad n \geq 0, \\ \gamma_n \Theta_{-1/2} &= 0, \quad n > 0. \end{aligned}$$

We have $\mathcal{V} = \{Q_B, \mathcal{W}\}$, with

$$(C.11) \quad \mathcal{W} = \left(1 - G_0^{\text{gh}} G_0^X \right) \Theta_{-1/2} \Phi,$$

where G_0^{gh} is the G_0 operator of the ghost system. In terms of ghost oscillator modes, this can be written

$$(C.12) \quad \mathcal{W} = \left(1 + \frac{1}{2} c_1 \beta_{-1} G_0^X \right) \Theta_{-1/2} \Phi.$$

In particular \mathcal{W} depends non-trivially on c , as is typical of gauge parameters for massive string states, but is annihilated by b_n and β_r , $n, r \geq 0$.

C.2. General proof

Now we want to show that in general, gauge parameters in superstring theory can be chosen to be annihilated by antighost modes b_n and β_r , $n, r \geq 0$, so that in analyzing gauge-invariance, a superconformal formalism is possible.

Let us consider first the NS sector. Let $\mathcal{V} = c\delta(\gamma)V$ be a superconformal vertex operator, where V is a superconformal primary of the matter system of dimension $1/2$. The no-ghost theorem says that $\mathcal{V} = \{Q_B, \mathcal{W}\}$ for some \mathcal{W} if and only if V is a null vector in the sense that it can be written as in eqn. (C.1). To the matter primary V we can also associate the Q_B -invariant operator $\mathcal{V}^* = c\partial c\delta(\gamma)V$. The no-ghost theorem says further that \mathcal{V}^* is BRST-trivial, $\mathcal{V}^* = \{Q_B, \mathcal{W}^*\}$ for some \mathcal{W}^* , if and only if V is a null vector. In these statements, we can assume that \mathcal{W} and \mathcal{W}^* are annihilated by L_0 .

We aim to show that if $\mathcal{V} = \{Q_B, \mathcal{W}\}$ for some \mathcal{W} , then \mathcal{W} can be chosen to be independent of $\partial^n c$ and $\partial^m \gamma$, with $n, m > 0$. Equivalently, we want to show that one can choose \mathcal{W} to be annihilated by antighost modes b_n and β_r , $n, r \geq 0$.

It is possible to generalize either of the two arguments described in section B.2, but here we will present a generalization only of the second one.⁹⁴ As before, the argument consists of (i) a general proof to show that we can assume that \mathcal{W} and \mathcal{W}^* do not involve $\partial^n c$, $n \geq 2$ or $\partial^m \gamma$, $m \geq 1$; and (ii) a special argument to show that we can further assume that \mathcal{W} does not involve ∂c .

Step (ii) is exactly as it was in section B.2: once we find \mathcal{W}^* obeying $\mathcal{V}^* = \{Q_B, \mathcal{W}^*\}$, annihilated by L_0 , and independent of $\partial^n c$, $n \geq 2$ and $\partial^m \gamma$, $m \geq 1$, we simply set $\mathcal{W} = b_0 \mathcal{W}^*$. Then \mathcal{W} obeys all the desired properties.

For step (i), we imitate the proof in section B.2. We define what we will call the $c''\gamma'$ degree of an operator to be the number of times that $\partial^n c$, $n \geq 2$ or $\partial^m \gamma$, $m \geq 1$ appear in an operator. For example, $c\partial c\partial^2 c\delta'(\gamma)\partial\gamma$ has $c''\gamma'$ degree 2. We write the BRST operator as

$$(C.13) \quad Q_B = Q_{>} + Q_0,$$

where $Q_{>}$ increases the $c''\gamma'$ degree by 1, and Q_0 leaves it unchanged. The condition $Q_B^2 = 0$ implies that $Q_{>}^2 = 0$, so we can define the cohomology of $Q_{>}$. Just as in section B.2, it suffices to show that the cohomology of the operator $Q_{>}$ is supported at $c''\gamma'$ degree 0.

To prove this, just as in the bosonic case, we introduce lightcone coordinates $X^\pm = (X^0 \pm X^1)/\sqrt{2}$, and an operator N_*^{lc} equal to minus the number of + excitations. Now, however, we must consider fermionic as well as bosonic + excitations. We expand the worldsheet fermions as $\psi^I(z) = \sum_{r \in \mathbb{Z} + 1/2} \psi_r^I / z^{r+1/2}$, $I = 0, \dots, 9$ with $\{\psi_r^I, \psi_s^J\} = \eta^{IJ} \delta_{rs}$, and define light cone operators $\psi_r^\pm = (\psi_r^0 \pm \psi_r^1)/\sqrt{2}$, obeying $\{\psi_r^+, \psi_s^-\} = \delta_{rs}$. Then we define

$$(C.14) \quad N_*^{\text{lc}} = \sum_{m \geq 1} \frac{1}{m} \alpha_{-m}^+ \alpha_m^- - \sum_{r \geq 1/2} \psi_{-r}^+ \psi_r^-.$$

As before, we expand $Q_{>}$ as a sum of terms $Q_{>,k}$ that shift N_*^{lc} by k :

$$(C.15) \quad Q_{>} = Q_{>,1} + Q_{>,0} + Q_{>,-1}.$$

⁹⁴For a generalization of the first one, see [91].

Exactly as in section B.2, it suffices to show that the cohomology of $Q_{>,1}$ is supported at $c''\dot{\gamma}$ -degree 0. Again, the explicit form of $Q_{>,1}$ makes this an easy result:

$$(C.16) \quad Q_{>,1} = -(2\alpha')^{1/2}k^+ \sum_{m \geq 1} \alpha_m^- c_{-m} + (2\alpha')^{1/2}k^+ \sum_{r \geq 1/2} \gamma_{-r} \psi_r^-.$$

To make a similar analysis for the Ramond sector, we mainly need to change our terminology slightly. A basis of states of the $\beta\gamma$ system at picture number $-1/2$ is given by

$$(C.17) \quad \prod_{s \geq 1} \beta_{-s}^{n_s} \prod_{r \geq 0} \gamma_{-r}^{m_r} \Theta_{-1/2},$$

where all but finitely many n_s and m_r are zero. We say that such a state has $\dot{\gamma}$ degree $\sum_{r \geq 0} m_r$. After including the bc ghosts and the matter fields, we define the $c''\dot{\gamma}$ -degree of a state to be the sum of the c'' -degree and the $\dot{\gamma}$ -degree of that state. (The $c''\dot{\gamma}$ -degree is bounded in states of picture number $-1/2$ and fixed L_0 and ghost number.) Consider a pair of Q_B -invariant states $\mathcal{V} = c\Theta_{-1/2}\Phi$, $\mathcal{V}^* = c\partial c\Theta_{-1/2}$, where Φ is a Ramond-sector matter primary of dimension $5/8$. According to the no-ghost theorem, \mathcal{V} and \mathcal{V}^* are Q_B -trivial if and only if Φ is a null vector. In this case, we want to prove that we can write $\mathcal{V} = \{Q_B, \mathcal{W}\}$, with \mathcal{W} not depending on derivatives of c and constructed only from states of $m_r = 0$ in (C.17). As in the other examples, it suffices to find \mathcal{W}^* of $c''\dot{\gamma}$ -degree 0 with $\mathcal{V}^* = \{Q, \mathcal{W}^*\}$; then we set $\mathcal{W} = b_0\mathcal{W}^*$.

To find a suitable \mathcal{W}^* , we make the same expansion as in (C.13), but now using the $c''\dot{\gamma}$ -degree, and again it will suffice to show that the cohomology of $Q_{>}$ is supported at $c''\dot{\gamma}$ -degree 0. The worldsheet fermion fields still have an expansion $\psi^I(z) = \sum_{r \in \mathbb{Z}} \psi_r^I / z^{r+1/2}$, but now with integer r . In defining N_*^{lc} , we now need to be careful with the treatment of modes with $r = 0$. We set

$$(C.18) \quad N_*^{\text{lc}} = \sum_{m \geq 1} \frac{1}{m} \alpha_{-m}^+ \alpha_m^- - \sum_{r \geq 0} \psi_{-r}^+ \psi_r^-.$$

Again we expand $Q_{>}$ as in (C.15) in terms of operators $Q_{>,k}$ that shift N_*^{lc} by k , and it suffices to show that the cohomology of $Q_{>,1}$ is supported at $c''\dot{\gamma}$ degree 0. This follows directly from the explicit form:

$$(C.19) \quad Q_{>,1} = -(2\alpha')^{1/2}k^+ \sum_{m \geq 1} \alpha_m^- c_{-m} + (2\alpha')^{1/2}k^+ \sum_{r \geq 0} \gamma_{-r} \psi_r^-.$$

Appendix D. An example of fermion integration

Here we explore in more detail the example briefly cited in section 6.6.2. This will also enable us to show some of the subtleties of fermionic integration in a concrete example. (The same example is explored in much more detail in [28].)

The example, which was studied in [39, 42, 43], involves a one-loop heterotic string amplitude with two NS vertex operators, with an even spin structure. We will not describe here the full string theory context for this computation. We will simply describe what is involved in this example in constructing the right integration cycle and integrating over it.

First we describe the string worldsheet. From a holomorphic point of view, a heterotic string worldsheet Σ is a super Riemann surface. A genus one super Riemann surface with an even spin structure can be described by superconformal coordinates $z|\theta$ with the equivalence relations

$$(D.1) \quad \begin{aligned} z &\cong z + 1 \\ \theta &\cong -\theta \end{aligned}$$

and

$$(D.2) \quad \begin{aligned} z &\cong z + \tau \\ \theta &\cong \theta. \end{aligned}$$

From an antiholomorphic point of view, Σ is an ordinary genus 1 Riemann surface, described by a complex coordinate \tilde{z} with the equivalence relations

$$(D.3) \quad \tilde{z} \cong \tilde{z} + 1 \cong \tilde{z} + \tilde{\tau}.$$

We will just set $\tilde{\tau} = \bar{\tau}$, though the general formalism would let us relax this slightly. τ is an even modulus of Σ , and in superstring perturbation theory, one integrates over it. However, the interesting subtleties do not involve the integral over τ , and for our purposes we will just set τ to a constant.

In general, the two NS vertex operators can be inserted at arbitrary points $z|\theta = u_1|\zeta_1$ and $z|\theta = u_2|\zeta_2$. The interesting subtlety will arise when the two vertex operators approach each other, or in other words when $u_1 \rightarrow u_2$. As depicted in fig. 41, this is actually a special case of a separating degeneration of a Riemann surface. The natural gluing parameter at this degeneration is

$$(D.4) \quad q_{\text{NS}} = u_1 - u_2 - \zeta_1 \zeta_2,$$

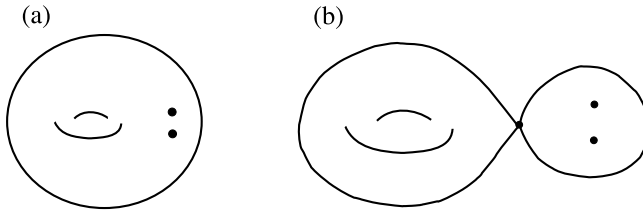


Figure 41: (a) A Riemann surface Σ of genus 1 with two marked points. (b) As the two marked points approach each other, Σ splits into two components Σ_ℓ and Σ_r . The external vertex operators are on Σ_r , which has genus 0.

as already explained in section 6.6.2 (and in section 6.3.2 of [16]). We should stress that what is important in (D.4) is that the function whose vanishing defines the separating degeneration is $u_1 - u_2 - \zeta_1\zeta_2$, rather than, say $u_1 - u_2$ or $u_1 - u_2 - 2\zeta_1\zeta_2$. The precise coefficient multiplying $u_1 - u_2 - \zeta_1\zeta_2$ is not relevant and has been chosen for convenience in (D.4). (The normalization used in [16] actually differs from this by a factor of 2.)

A genus 1 surface has a translation symmetry $z \rightarrow z + \text{constant}$. Using this, we can set, say, $u_2 = 0$. We write simply u instead of u_1 . With this choice, the insertion points of the two NS vertex operators are $\tilde{z};z|\theta = \tilde{u};u|\zeta_1$ and $0;0|\zeta_2$. The equivalence relations (D.1), (D.2), and (D.3) become

$$\begin{aligned}
 u &\cong u + 1 \\
 \zeta_1 &\cong -\zeta_1 \\
 \zeta_2 &\cong \zeta_2 \\
 \tilde{u} &\cong \tilde{u} + 1
 \end{aligned}
 \tag{D.5}$$

and

$$\begin{aligned}
 u &\cong u + \tau \\
 \zeta_1 &\cong \zeta_1 \\
 \zeta_2 &\cong \zeta_2 \\
 \tilde{u} &\cong \tilde{u} + \bar{\tau}
 \end{aligned}
 \tag{D.6}$$

The gluing parameters are

$$\tilde{q} = \tilde{u}
 \tag{D.7}$$

and

$$(D.8) \quad q_{\text{NS}} = u - \zeta_1 \zeta_2.$$

The moduli space \mathcal{M}_R is parameterized by $u|\zeta_1, \zeta_2$, and the moduli space \mathcal{M}_L is parameterized by \tilde{u} , in each case with equivalence relations stated in (D.5) and (D.6). Now we want to define an integration cycle $\Gamma \subset \mathcal{M}_L \times \mathcal{M}_R$. We will aim for the simplest choice of Γ , meaning that its reduced space Γ_{red} will be the “diagonal,” by which we mean the subspace of $\mathcal{M}_{L,\text{red}} \times \mathcal{M}_{R,\text{red}}$ defined by $\tilde{u} = u$. Once we specify Γ_{red} , the most general possible choice of Γ is made by generalizing the equation $\tilde{u} = u$ to include the odd variables. Since in this example, we only have two odd variables ζ_1 and ζ_2 , the most general choice of Γ is defined by

$$(D.9) \quad \tilde{u} = u + \zeta_1 \zeta_2 h(\tilde{u}, u)$$

for some function $h(u, \tilde{u})$. For Γ to be invariant under the equivalences (D.5) and (D.6), we require

$$(D.10) \quad h(\tilde{u} + 1, u + 1) = -h(\tilde{u}, u), \quad h(\tilde{u} + \bar{\tau}, u + \tau) = h(\tilde{u}, u).$$

These conditions would allow us to set $h = 0$, but they would not allow us, for example, to set $h = -1$. However, we need one more condition for h . At the degeneration $u = \tilde{u} = 0$, Γ is supposed to be defined by $\tilde{q} = q_{\text{NS}}$, and this implies that we want

$$(D.11) \quad h(0, 0) = -1.$$

We cannot set h to a constant, because no constant is compatible with both (D.11) and (D.10). Functions h obeying the conditions do exist, but there is no canonical choice. And therefore, there is no canonical choice for the integration cycle Γ of superstring perturbation theory in this situation. However, any two choices give homologous integration cycles. Indeed, if h_1 and h_2 are any two functions that obey the conditions, we can interpolate between them via

$$(D.12) \quad h_\lambda = \lambda h_1 + (1 - \lambda) h_2, \quad 0 \leq \lambda \leq 1,$$

and this gives an explicit homology between the two choices of Γ .

Now let us discuss what sort of measure we want to integrate over Γ . Perhaps the most obvious idea is to consider a section of the Berezinian of

$\mathcal{M}_L \times \mathcal{M}_R$ of the simple form $[d\tilde{u};du|d\zeta_1 d\zeta_2]$. This measure behaves well at infinity, since it is invariant under the change of variables $u \rightarrow u - \zeta_1\zeta_2$, where according to (D.8), $u - \zeta_1\zeta_2$ is a good coordinate at infinity. However, $[d\tilde{u};du|d\zeta_1 d\zeta_2]$ is odd under (D.5) and so cannot be the form we want to integrate. Instead we consider an integration form

$$(D.13) \quad \Omega = [d\tilde{u};du|d\zeta_1 d\zeta_2] P(\tilde{u}),$$

where

$$(D.14) \quad P(\tilde{u} + 1) = -P(\tilde{u}), \quad P(\tilde{u} + \bar{\tau}) = P(\tilde{u}),$$

so that Ω respects the necessary equivalences and is well-defined and holomorphic near $\Delta^* \subset \mathcal{M}_L \times \mathcal{M}_R$. We further require that $P(\tilde{u})$ is a meromorphic function of \tilde{u} whose only singularity is a simple pole with residue 1 at $\tilde{u} = 0$. These conditions determine $P(\tilde{u})$:

$$(D.15) \quad P(\tilde{u}) = \sum_{n,m \in \mathbb{Z}} \frac{(-1)^n}{\tilde{u} + n + m\bar{\tau}}.$$

We now want to evaluate the integral

$$(D.16) \quad I = \int_{\Gamma} \Omega.$$

How such an integral arises in superstring perturbation theory is explained in [39, 42]. The singularity of Ω at $\tilde{u} = u = 0$ is $d\tilde{u} du/\tilde{u}$, milder than the $d\tilde{u} du/\tilde{u}u$ singularity that leads to subtleties with massless tadpoles, but singular enough to lead to the interesting phenomenon that we are about to describe.

To evaluate the integral, we simply use the condition (D.9) that defines the integration cycle to solve for u . Since (D.9) says that $u = \bar{\tilde{u}} \pmod{\zeta_1\zeta_2}$, and since $(\zeta_1\zeta_2)^2 = 0$, we can solve (D.9) by

$$(D.17) \quad u = \bar{\tilde{u}} - \zeta_1\zeta_2 h(\tilde{u}, \bar{\tilde{u}}).$$

This enables us to express the measure $[d\tilde{u};du|d\zeta_1 d\zeta_2]$ in terms of \tilde{u} , $\bar{\tilde{u}}$, ζ_1 , and ζ_2 only:

$$(D.18) \quad [d\tilde{u};du|d\zeta_1 d\zeta_2] = \left(1 - \zeta_1\zeta_2 \frac{\partial h}{\partial \bar{\tilde{u}}}\right) [d\tilde{u} d\bar{\tilde{u}}|d\zeta_1 d\zeta_2].$$

So our integral is

$$(D.19) \quad I = \int_{\Gamma} [d\tilde{u} d\bar{\tilde{u}} |d\zeta_1 d\zeta_2] \left(1 - \zeta_1 \zeta_2 \frac{\partial h}{\partial \bar{\tilde{u}}}\right) P(\tilde{u}).$$

Integrating first over the fermions, via the Berezin integral $\int [d\zeta_1 d\zeta_2] 1 = 0$, $\int [d\zeta_1 d\zeta_2] \zeta_1 \zeta_2 = 1$, we reduce to an ordinary integral over the torus Σ_{red} :

$$(D.20) \quad I = - \int_{\Sigma_{\text{red}}} d\tilde{u} \wedge d\bar{\tilde{u}} \frac{\partial h}{\partial \bar{\tilde{u}}} P(\tilde{u}).$$

Integrating by parts and using the fact that $\partial_{\bar{\tilde{u}}} P(\tilde{u}) = 2\pi \delta^2(\tilde{u})$, because of the pole of $P(\tilde{u})$ at $\tilde{u} = 0$, and the fact that $h(0, 0) = -1$, we get

$$(D.21) \quad I = 4\pi i.$$

As expected, the integral did not depend on the choice of h , as long as it behaves correctly at infinity.

D.1. What we have learned

One lesson to learn from this example is that the evaluation of a superstring scattering amplitude depends crucially on knowing how the integration cycle Γ is supposed to behave near infinity. Another lesson is that although the integral has a well-defined value, there is no natural answer to the question of where on moduli space the answer came from; this depends on the unnatural choice of Γ .

It is true that after integration by parts, the answer in this example seemed to come from a delta function at $u = 0$. However, this appears to be special to this low genus example, somewhat analogous to the fact that in genus 1, the vanishing of the dilaton tadpole can be established by summing over spin structures.

The basic reason that in this example it is possible to isolate a delta function contribution at infinity is that the moduli spaces \mathcal{M}_L and \mathcal{M}_R are holomorphically split (this notion was defined in section 3.3.2). \mathcal{M}_R is holomorphically split by the map that takes $z|\zeta_1\zeta_2$ to z , and \mathcal{M}_L is trivially split since it is purely bosonic. When \mathcal{M}_L and \mathcal{M}_R are holomorphically split, we have a natural holomorphic map $\pi : \mathcal{M}_L \times \mathcal{M}_R \rightarrow \mathcal{M}_{L,\text{red}} \times \mathcal{M}_{R,\text{red}}$, and then there is a fairly natural choice of integration cycle, namely $\Gamma_0 = \pi^{-1}(\Delta)$, where $\Delta \subset \mathcal{M}_{L,\text{red}} \times \mathcal{M}_{R,\text{red}}$ is the “diagonal.” The only thing that may be wrong with Γ_0 is that it may have the wrong behavior at infinity. In our

example, this is the case; Γ_0 corresponds to taking h to be identically zero, while the desired behavior at infinity is $h(0,0) = -1$. Still, we can pick Γ to coincide with Γ_0 except in a small neighborhood of infinity, so one can write the superstring amplitude as an integral over Γ_0 plus a correction at infinity. Our example has the further property that the bulk contribution – the integral over Γ_0 – vanishes. In a more generic situation, there is no holomorphic splitting [45], and one should expect that there is no natural choice of integration cycle even away from divisors at infinity, and no natural way to write a superstring amplitude as the sum of a “bulk” contribution and a contribution at infinity.

D.2. The moduli space as an orbifold

Now let us look more closely at the moduli space $\mathfrak{M}_{1,2,0,+}$ that parametrizes a super Riemann surface Σ with two NS punctures. As above, it can be parametrized by $u|\zeta_1, \zeta_2$, with the equivalences in (D.5) and (D.6).

However, there is one more symmetry that we should take into account. The super Riemann surface Σ that we started with in eqns. (D.1) and (D.2) has the additional symmetry

$$(D.22) \quad \begin{aligned} z &\rightarrow -z \\ \theta &\rightarrow \pm\sqrt{-1}\theta. \end{aligned}$$

The transformation of θ by $\pm\sqrt{-1}\theta$ ensures that $\varpi = dz - \theta d\theta$, which defines the superconformal structure, maps to a multiple of itself. With some choice of the sign, let us call this automorphism κ . There is no natural choice of sign and indeed we must allow both signs. κ^2 is the universal symmetry $z|\theta \rightarrow z|-\theta$ of any split super Riemann surface, and κ^3 is obtained from κ by reversing the sign.

Accordingly, we should impose on the variables $u|\zeta_1, \zeta_2$ the equivalence relation

$$(D.23) \quad \begin{aligned} u &\rightarrow -u \\ \zeta_i &\rightarrow \pm\sqrt{-1}\zeta_i, \quad i = 1, 2. \end{aligned}$$

generated by κ . κ^2 is the universal symmetry

$$(D.24) \quad \begin{aligned} u &\rightarrow u \\ \zeta_i &\rightarrow -\zeta_i, \quad i = 1, 2 \end{aligned}$$

that reverses the sign of all odd moduli, and again κ^3 is obtained from κ by reversing the sign.

So $\mathfrak{M}_{1,2,0,+}$ must be understood as an orbifold (or stack). There is a \mathbb{Z}_2 automorphism group, generated by κ^2 , whenever $\zeta_1 = \zeta_2 = 0$. This automorphism group is enhanced to \mathbb{Z}_4 , generated by κ , at $u = \zeta_1 = \zeta_2 = 0$. In fact, taking account the equivalences (D.5) and (D.6), the automorphism group is enhanced to \mathbb{Z}_4 at the four points $u = 0, 1/2, \tau/2$, and $(1 + \tau)/2$, with $\zeta_1 = \zeta_2 = 0$.

κ maps $q_{\text{NS}} = u - \zeta_1\zeta_2$ to $-q_{\text{NS}}$, consistent with the claim that the gluing parameters are well-defined up to multiplication by an invertible function.

We did not take the automorphism κ into account in our discussion of integration. However, the form (D.13) that we integrated is κ -invariant, with κ acting on \tilde{u} by $\tilde{u} \rightarrow -\tilde{u}$.

Acknowledgments

Research partly supported by NSF grant PHY-0969448. I thank N. Berkovits, L. Brink, P. Goddard, J. Maldacena, V. Pestun, J. Polchinski, M. Rangamani, A. Sen, N. Seiberg, H. Verlinde, and B. Zwiebach and especially K. Becker, P. Deligne, G. Moore, D. Robbins, B. Safdi, and C. Thorn for helpful comments and discussions. M. Turansick assisted with the figures.

References

- [1] D. FRIEDAN, E. MARTINEC, and S. SHENKER, “Covariant Quantization Of Superstrings,” *Phys. Lett.* **B160** (1985) 55, “Conformal Invariance, Supersymmetry, and String Theory,” *Nucl. Phys.* **B271** (1986) 93. [MR0845945](#).
- [2] V. KNIZHNIK, “Covariant Fermion Vertex In Superstrings,” *Phys. Lett.* **B160** (1985) 403–407. [MR0818682](#).
- [3] A. CAPPELLI, E. CASTELLANI, F. COLOMO, and P. DIVECCHIA, eds., *The Birth Of String Theory* (Cambridge University Press, 2012).
- [4] G. MOORE, P. NELSON, and J. POLCHINSKI, “Strings And Supermoduli,” *Phys. Lett.* **B169** (1986) 47–53. [MR0834339](#).
- [5] L. ALVAREZ-GAUMÉ, C. GOMEZ, P. C. NELSON, G. SIERRA, and C. VAFA, “Fermionic Strings In The Operator Formalism,” *Nucl. Phys.* **B311** (1988) 333. [MR0980864](#).

- [6] A. A. ROSLY, A. S. SCHWARZ, and A. A. VORONOV, “Superconformal Geometry And String Theory,” *Commun. Math. Phys.* **120** (1989) 437–450.
- [7] E. VERLINDE and H. VERLINDE, “Multiloop Calculations In Covariant Superstring Theory,” *Phys. Lett.* **B192** (1987) 95.
- [8] E. D’HOKER and D. H. PHONG, “The Geometry Of String Perturbation Theory,” *Rev. Mod. Phys.* **60** (1988) 917–1065. [MR0969998](#).
- [9] A. BELOPOLSKY, “De Rham Cohomology Of The Supermanifolds And Superstring BRST Cohomology,” *Phys. Lett.* **B403** (1997), [hep-th/9609220](#).
- [10] A. BELOPOLSKY, “New Geometrical Approach To Superstrings,” [hep-th/9703183](#).
- [11] A. BELOPOLSKY, “Picture Changing Operators In Supergeometry And Superstring Theory,” [hep-th/9706033](#).
- [12] M. B. GREEN and J. H. SCHWARZ, “Supersymmetric Dual String Theory (III). Loops and Renormalization,” *Nucl. Phys.* **B198** (1982) 441–460.
- [13] L. BRINK, M. B. GREEN, and J. H. SCHWARZ, “ $\mathcal{N} = 4$ Yang-Mills And $\mathcal{N} = 8$ Supergravity As Limits Of String Theories,” *Nucl. Phys.* **B198** (1982) 474–492.
- [14] E. D’HOKER and D. H. PHONG, “Lectures On Two-Loop Superstrings,” *Adv. Lect. Math.* **1** 85–123, [hep-th/0211111](#), and references therein.
- [15] E. WITTEN, “Notes On Supermanifolds And Integration,” *Pure Appl. Math. Q.* **15** (2019) 3–56, [arXiv:1209.2199](#).
- [16] E. WITTEN, “Notes On Super Riemann Surfaces And Their Moduli,” *Pure Appl. Math. Q.* **15** (2019) 57–211, [arXiv:1209.2459](#).
- [17] J. POLCHINSKI, “Factorization Of Bosonic String Amplitudes,” *Nucl. Phys.* **B307** (1988) 61.
- [18] P. C. NELSON, “Covariant Insertions Of General Vertex Operators,” *Phys. Rev. Lett.* **62** (1989) 993–996. [MR0981301](#).
- [19] R. PIUS, A. RUDRA, and A. SEN, “Mass Renormalization in String Theory: General States,” *JHEP* **1407** (2014) 062, [arXiv:1401.7014](#).
- [20] A. SEN, “Off-shell Amplitudes In Superstring Theory,” *Fortsch. Phys.* **63** (2015) 149–188, [arXiv:1408.0571](#).

- [21] A. SEN, “One Loop Mass Renormalization Of Unstable Particles In Superstring Theory,” JHEP **1611** (2016) 050, [arXiv:1607.06500](#).
- [22] A. SEN, “Unitarity of Superstring Field Theory,” JHEP **1612** (2016) 115, [arXiv:1607.06244](#).
- [23] A. SEN, “Wilsonian Effective Action Of Superstring Theory,” JHEP **1701** (2017) 108, [arXiv:1609.00459](#).
- [24] A. SEN, “Background Independence Of Closed Superstring Field Theory,” [arXiv:1711.08468](#).
- [25] N. BERKOVITS, “Super-Poincaré Covariant Quantization Of The Superstring,” JHEP **0004** (2000) 018.
- [26] J. SHAPIRO, “Loop Graph In The Dual-Tube Model,” Phys. Rev. **D5** (1972) 1945–1948.
- [27] C. LOVELACE, “Pomeron Form Factors And Dual Regge Cuts,” Phys. Lett. **B34** (1971) 500–506.
- [28] E. WITTEN, “More On Superstring Perturbation Theory: An Overview of Superstring Perturbation Theory Via Super Riemann Surfaces,” Pure Appl. Math. Q. **15** (2019) 517–607, [arXiv:1304.2832](#).
- [29] O. ALVAREZ, “Theory Of Strings With Boundaries: Fluctuations, Topology, and Quantum Geometry,” Nucl. Phys. **B216** (1983) 125–184.
- [30] S. GIDDINGS and E. MARTINEC, “Conformal Geometry And String Field Theory,” Nucl. Phys. **B278** (1986) 91–120.
- [31] L. ALVAREZ-GAUMÉ, C. GOMEZ, G. W. MOORE, and C. VAFA, “Strings In The Operator Formalism,” Nucl. Phys. **B303** (1988) 455.
- [32] J. POLCHINSKI, *String Theory* (Cambridge University Press, 1998).
- [33] M. KATO and K. OGAWA, “Covariant Quantization Of String Based On BRS Invariance,” Nucl. Phys. **B212** (1983) 443.
- [34] I. B. FRENKEL, H. GARLAND, and G. ZUCKERMAN, “Semi-Infinite Cohomology And String Theory,” Proc. Natl. Acad. Sci. USA **83** (1986) 8442–8446. [MR0865483](#).
- [35] M. D. FREEMAN and D. OLIVE, “BRS Cohomology In String Theory And The No Ghost Theorem,” Phys. Lett. **B175** (1986) 151. [MR0851340](#).
- [36] C. B. THORN, “A Detailed Study Of The Physical State Conditions In Covariantly Quantized String Theories,” Nucl. Phys. **B286** (1987) 61.

- [37] P. GODDARD and C. THORN, “Compatibility Of The Dual Pomeron With Unitarity And The Absence Of Ghosts In The Dual Resonance Model,” *Phys. Lett.* **B40** (1972) 235–238.
- [38] R. C. BROWER, “Spectrum Generating Algebra And No Ghost Theorem For The Dual Model,” *Phys. Rev.* **D6** (1972) 1655–1662.
- [39] M. DINE, I. ICHINOSE, and N. SEIBERG, “ F Terms And D Terms In String Theory,” *Nucl. Phys.* **B293** (1987) 253.
- [40] P. DELIGNE, letter to Yu. I. Manin (September 25, 1987).
- [41] P. DELIGNE, lectures at the workshop on Supermoduli (Simons Center for Geometry and Physics, May, 2015), video available at http://scgp.stonybrook.edu/video_portal/video.php?id=954.
- [42] J. J. ATICK, L. J. DIXON, and A. SEN, “String Calculation Of Fayet-Iliopoulos D -Terms In Arbitrary Supersymmetric Compactifications,” *Nucl. Phys.* **B292** (1987) 109–149. [MR0900776](#).
- [43] M. B. GREEN and N. SEIBERG, “Contact Interactions In Superstring Theory,” *Nucl. Phys.* **B299** (1988) 559.
- [44] M. DINE, N. SEIBERG, and E. WITTEN, “Fayet-Iliopoulos Terms In String Theory,” *Nucl. Phys.* **B278** (1986) 769.
- [45] R. DONAGI and E. WITTEN, “Supermoduli Space Is Not Projected,” *Proc. Symp. Pure Math.* **90** (2015) 19–72, [arXiv:1304.7798](#).
- [46] E. VERLINDE and H. VERLINDE, “Lectures On String Perturbation Theory,” in *Superstrings ’89* (proceedings of the ICTP Spring School), ed. M. B. GREEN et al. (World-Scientific, 1990). [MR1091556](#).
- [47] O. LECHTENFELD, “Superconformal Ghost Correlations On Riemann Surfaces,” *Phys. Lett.* **B232** (1989) 193.
- [48] J. J. ATICK, J. M. RABIN, and A. SEN, “An Ambiguity In Fermionic String Perturbation Theory,” *Nucl. Phys.* **B299** (1988) 279.
- [49] A. SEN and E. WITTEN, “Filling The Gaps With PCO’s,” *JHEP* **1509** (2015) 004, [arXiv:1504.00609](#).
- [50] P. DELIGNE and J. W. MORGAN, “Notes On Supersymmetry (following Joseph Bernstein),” in P. DELIGNE et al., eds., *Quantum Fields And Strings: A Course For Mathematicians, Vol. 1* (American Mathematical Society, 1999).
- [51] P. RAMOND, “Dual Theory For Free Fermions,” *Phys. Rev.* **D3** (1971) 2415–2418.

- [52] F. GLIOZZI, J. SCHERK, and D. I. OLIVE, “Supergravity And The Dual Spinor Model,” *Phys. Lett.* **B65** (1976) 282.
- [53] L. DIXON, D. FRIEDAN, E. MARTINEC, and S. SHENKER, “The Conformal Field Theory Of Orbifolds,” *Nucl. Phys.* **B282** (1987) 13–73.
- [54] E. WITTEN, “Noncommutative Geometry And String Field Theory,” *Nucl. Phys.* **B268** (1986) 253.
- [55] A. BERERA, “Unitary String Amplitudes,” *Nucl. Phys.* **B411** (1994) 157–180.
- [56] E. WITTEN, “The Feynman $i\epsilon$ In String Theory,” *JHEP* **1504** (2015) 055, [arXiv:1307.5124](#).
- [57] L. ALVAREZ-GAUMÉ and E. WITTEN, “Gravitational Anomalies,” *Nucl. Phys.* **B234** (1984) 269.
- [58] S. GIDDINGS and S. WOLPERT, “A Triangulation Of Moduli Space From Light Cone String Theory,” *Commun. Math. Phys.* **109** (1987) 177–190.
- [59] S. GIDDINGS, “Conformal Techniques In String Theory And String Field Theory,” *Phys. Reports* **170** (1988) 167–212.
- [60] S. MANDELSTAM, “The n Loop String Amplitude: Explicit Formulas, Finiteness, And Absence Of Ambiguities,” *Phys. Lett.* **B277** (1992) 82–88.
- [61] K. AOKI, E. D’HOKER, and D. PHONG, “Unitarity Of Closed Superstring Perturbation Theory,” *Nucl. Phys.* **B342** (1990) 149–230.
- [62] E. D’HOKER and D. H. PHONG, “The Box Graph In Superstring Theory,” *Nucl. Phys.* **B440** (1995) 24–94. [MR1336085](#).
- [63] R. BRITTO, F. CACHAZO, and B. FENG, “Generalized Unitarity and One-Loop Amplitudes in N=4 Super-Yang-Mills,” *Nucl. Phys.* **B725** (2005) 275–305, [hep-th/0412103](#). [MR2164293](#).
- [64] J. J. ATICK, G. MOORE, and A. SEN, “Catoptric Tadpoles,” *Nucl. Phys.* **B307** (1988) 221–273. [MR0960840](#).
- [65] S. WEINBERG, “Radiative Corrections In String Theory,” in *Proc. Oregon Meeting*, ed. R. C. HWA (World Scientific, 1986), p. 850. [MR0856955](#).
- [66] N. SEIBERG, “Anomalous Dimensions And Mass Renormalization In String Theory,” *Phys. Lett.* **B187** (1987) 56.

- [67] A. SEN, “Mass Renormalization And BRST Anomaly In String Theories,” Nucl. Phys. **B304** (1988) 403–416.
- [68] M. DINE and N. SEIBERG, “Is The Superstring Weakly Coupled?” Phys. Lett. **B162** (1985) 299.
- [69] M. DINE and N. SEIBERG, “Nonrenormalization Theorems In Superstring Theory,” Phys. Rev. Lett. **57** (1986) 2625–2628.
- [70] L. J. ROMANS, “Massive $\mathcal{N} = 2a$ Supergravity In Ten-Dimensions,” Phys. Lett. **B169** (1986) 374.
- [71] A. BELOPOLSKY and B. ZWIEBACH, “Who Changes The String Coupling?” Nucl. Phys. **B472** (1996) 109–138.
- [72] M. KALB and P. RAMOND, “Classical Direct Interstring Action,” Phys. Rev. **D9** (1974) 2273–2284.
- [73] H. LA and P. NELSON, “Unambiguous Fermionic-String Amplitudes,” Phys. Rev. Lett. **63** (1989) 24.
- [74] S. WEINBERG, “Photons And Gravitons In S -Matrix Theory: Derivation Of Charge Conservation And Equality of Gravitational And Inertial Mass,” Phys. Rev. **135** (1964) B1049–B1056.
- [75] M. T. GRISARU and H. N. PENDLETON, “Soft Spin 3/2 Fermions Require Gravity And Supersymmetry,” Phys. Lett. **B67** (1977) 323–326.
- [76] T. YONEYA, “Quantum Gravity And The Zero Slope Limit Of The Generalized Virasoro Model,” Lett. Nuovo Cimento **8** (1973) 951–955.
- [77] J. SCHERK and J. H. SCHWARZ, “Dual Models For Nonhadrons,” Nucl. Phys. **B81** (1974) 118–144.
- [78] S. WEINBERG, “Photons And Gravitons In Perturbation Theory: Derivation Of Maxwell’s and Einstein’s Equations,” Phys. Rev. **138** (1965) B988–B1002.
- [79] S. FERRARA and C. KOUNNAS, “Extended Supersymmetry In Four-Dimensional Type II Strings,” Nucl. Phys. **B328** (1989) 406–438.
- [80] J. DAI, R. G. LEIGH, and J. POLCHINSKI, “New Connections Among String Theories,” Mod. Phys. Lett. **A4** (1989) 2073–2083.
- [81] P. HORAVA, “Strings On World Sheet Orbifolds,” Nucl. Phys. **B327** (1989) 461.

- [82] M. B. GREEN and J. H. SCHWARZ, “Anomaly Cancellation In Supersymmetric D=10 Gauge Theory And Superstring Theory,” Phys. Lett. **B149** (1984) 117–122.
- [83] M. B. GREEN and J. H. SCHWARZ, “Infinity Cancellations In SO(32) Superstring Theory,” Phys. Lett. **B151** (1985) 21–25.
- [84] Y. CAI and J. POLCHINSKI, “Consistency Of Open Superstring Theories,” Nucl. Phys. **B296** (1988) 91.
- [85] H. W. DIEHL and S. DIETRICH, “Field Theoretical Approach To Multicritical Behavior Near Free Surfaces,” Z. Phys. **B42** (1981) 65.
- [86] J. POLCHINSKI, “Dirichlet Branes And Ramond-Ramond Charges,” Phys. Rev. Lett. **75** (1995) 4724–4727. [MR1366179](#).
- [87] J. H. SCHWARZ and E. WITTEN, “Anomaly Analysis of Brane-Antibrane Systems,” JHEP **0103** (2001) 032.
- [88] C. G. CALLAN, JR., C. LOVELACE, C. R. NAPPI, and S. A. YOST, “Adding Holes And Crosscaps To The Superstring,” Nucl. Phys. **B293** (1987) 83.
- [89] S. YOST, “Bosonized Superstring Boundary States And Partition Functions,” Nucl. Phys. **B321** (1989) 629.
- [90] N. BERKOVITS and B. ZWIEBACH, “On The Picture Dependence Of Ramond-Ramond Cohomology,” Nucl. Phys. **B523** (1998) 311–343.
- [91] C. THORN, “Null Physical States In String Models,” JHEP **02** (2013) 030, [arXiv:1212.4429](#).

Edward Witten
School of Natural Sciences
Institute for Advanced Study
1 Einstein Drive
Princeton, NJ 08540
USA
E-mail: witten@ias.edu