4
Strings on circles and T-duality

In this chapter we shall study the spectrum of strings propagating in a spacetime that has a compact direction. The theory has all of the properties we might expect from the knowledge that at low energy we are placing gravity and field theory on a compact space. Indeed, as the compact direction becomes small, the parts of the spectrum resulting from momentum in that direction become heavy, and hence less important, but there is much more. The spectrum has additional sectors coming from the fact that closed strings can wind around the compact direction, contributing states whose mass is proportional to the radius. Thus, they become light as the circle shrinks. This will lead us to T-duality, relating a string propagating on a large circle to a string propagating on a small circle[14]. This is just the first of the remarkable symmetries relating two string theories in different situations that we shall encounter here. It is a crucial consequence of the fact that strings are extended objects. Studying its consequences for open strings will lead us to D-branes, since T-duality will relate the Neumann boundary conditions we have already encountered to Dirichlet ones[9, 11], corresponding to open strings ending on special hypersurfaces in spacetime.

4.1 Fields and strings on a circle

Let us remind ourselves of what happens in field theory, for the case of placing gravity on a spacetime with a compact direction. This will help us appreciate the extra features encountered in the case of strings, and will also prepare for remarks to be made in a variety of cases much later. We start with the idea of Kaluza, later refined by Klein.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

4.1.1 The Kaluza–Klein reduction

Imagine that we are in five dimensions, with metric components G_{MN}, $M, N = 0, \ldots, 4$, and that the spacetime is actually of topology $\mathbb{R}^4 \times S^1$, and so has one compact direction. So we will have the usual four dimensional coordinates on $\mathbb{R}^4$, $(x^\mu, \mu = 0, \ldots, 3)$ and a periodic coordinate, $x^4 = x^4 + 2\pi R$, where R is the radius of the circle.

Now as we have seen before, the five dimensional coordinate transformation $x^M \to x'^M = x^M + \epsilon^M(x)$ is an invariance of our five dimensional theory, under which

$$G_{MN} \to G'_{MN} = G_{MN} - \partial_M \epsilon_N - \partial_N \epsilon_M. \tag{4.1}$$

The metric has the natural decomposition into $G^{(5)}_{\mu\nu}$, $G^{(5)}_{44}$, and $G^{(5)}_{\mu 4}$, where the superscript is necessary to distinguish similar-looking quantities in four dimensions, as we shall see.

Let us consider the class of transformations $\epsilon_4(x^\mu)$, $\epsilon_\mu = 0$, which corresponds to an x^μ-dependent isometry (rotation) of the circle. Then $G^{(5)}_{\mu\nu}$ and $G^{(5)}_{44}$ are invariant, and

$$G^{(5)}_{\mu 4} \to G'^{(5)}_{\mu 4} = G^{(5)}_{\mu 4} - \partial_\mu \epsilon_4(x). \tag{4.2}$$

However, from the four dimensional point of view, $G^{(5)}_{44}$ is a scalar, $G^{(5)}_{\mu\nu}$ is proportional to the metric, and $G^{(5)}_{\mu 4}$ is a vector, proportional to what we will call A_μ, and so equation (4.2) is simply a $U(1)$ gauge transformation: $A_\mu \to A_\mu - \partial_\mu \Lambda(x)$. So the $U(1)$ of electromagnetism can be thought of as resulting from compactifying gravity, the gauge field being an internal component of the metric. The idea of using this, as a first attempt at unifying gravity with electromagnetism, was that R is small enough that the world would be effectively four dimensional on larger scales, so an observer would have to work hard to see it. On distance scales much longer than that set by R, physical quantities in the theory would be effectively x^4-independent.

Let us be a bit more precise. Explicitly, we can write the most general metric consistent with the translation invariance in x^4 as

$$ds^2 = G^{(5)}_{MN} dx^M dx^N = G^{(4)}_{\mu\nu} dx^\mu dx^\nu + G_{44} \left(dx^4 + A_\mu dx^\mu\right)^2, \tag{4.3}$$

and we write $G_{44} = e^{2\phi}$. The five dimensional Ricci scalar decomposes as

$$R^{(5)} = R^{(4)} - 2e^{-\phi} \nabla^2 e^\phi - \tfrac{1}{4} e^{2\phi} F_{\mu\nu} F^{\mu\nu}, \tag{4.4}$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Notice for future reference that the lower dimensional metric components in the $0, 1, 2, 3$ directions are a modification of the higher dimensional metric components:

$$G^{(4)}_{\mu\nu} = G^{(5)}_{\mu\nu} - e^{2\phi} A_\mu A_\nu,$$

which is an important observation for later. So, suppressing the x^4 dependence of the fields, we get

$$\begin{aligned} S &= \frac{1}{16\pi G^N_{(5)}} \int (-G_{(5)})^{1/2} R^{(5)} d^5x \\ &= \frac{1}{16\pi G^N_{(4)}} \int (-G_{(4)})^{1/2} \left(R^{(4)} - \tfrac{3}{2}\partial_\mu\phi\partial^\mu\phi - \tfrac{1}{4} e^{3\phi} F_{\mu\nu}F^{\mu\nu} \right) d^4x, \end{aligned}$$

where we have defined $\tilde{G}^{(4)}_{\mu\nu} = e^{\phi} G^{(4)}_{\mu\nu}$ and used equation (2.110). Now we have a relation between the five dimensional and four dimensional Newton constants:

$$\frac{2\pi R}{G^N_{(5)}} = \frac{1}{G^N_{(4)}}, \tag{4.5}$$

and the gauge coupling is set by ϕ and Newton's constant.

Let us be more careful about following how the x^4-independence of the theory arises. Since momentum in x^4 is quantised as $p_4 = n/R$, any scalar (or component of a field) in $D = 5$ (which obeys $\partial^M \partial_M \phi = 0$) can be expanded:

$$\phi(x^\mu) = \sum_{n\in\mathbb{Z}} \phi_n(x^\mu) e^{inx^4/R}, \tag{4.6}$$

giving

$$\partial^\mu\partial_\mu\phi_n - \frac{n^2}{R^2}\phi = 0, \tag{4.7}$$

and so we see that the ϕ_n appear in four dimensions as a family of scalars of mass $m = n/R$, and $U(1)$ charge n. We get a tower of states which becomes extremely heavy for very small R, and are therefore hard to excite. We shall see this sort of spectrum arise in the closed string theory as well (since it contains gravity at low energy), but accompanied by new features.

4.1.2 Closed strings on a circle

The mode expansion (2.84) for the closed string theory can be written as:

$$X^\mu(z,\bar{z}) = \frac{x^\mu}{2} + \frac{\tilde{x}^\mu}{2} - i\sqrt{\frac{\alpha'}{2}}(\alpha_0^\mu + \tilde{\alpha}_0^\mu)\tau + \sqrt{\frac{\alpha'}{2}}(\alpha_0^\mu - \tilde{\alpha}_0^\mu)\sigma + \text{oscillators}. \tag{4.8}$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

We have already identified the spacetime momentum of the string:

$$p^\mu = \frac{1}{\sqrt{2\alpha'}}(\alpha_0^\mu + \tilde{\alpha}_0^\mu). \tag{4.9}$$

If we run around the string, i.e. take $\sigma \to \sigma + 2\pi$, the oscillator terms are periodic and we have

$$X^\mu(z, \bar{z}) \to X^\mu(z, \bar{z}) + 2\pi\sqrt{\frac{\alpha'}{2}}(\alpha_0^\mu - \tilde{\alpha}_0^\mu). \tag{4.10}$$

So far, we have studied the situation of non-compact spatial directions for which the embedding function $X^\mu(z, \bar{z})$ is single-valued, and therefore the above change must be zero, giving

$$\alpha_0^\mu = \tilde{\alpha}_0^\mu = \sqrt{\frac{\alpha'}{2}}p^\mu. \tag{4.11}$$

Indeed, momentum p^μ takes a continuum of values reflecting the fact that the direction X^μ is non-compact.

Let us consider the case that we have a compact direction, say X^{25}, of radius R. Our direction X^{25} therefore has period $2\pi R$. The momentum p^{25} now takes the discrete values n/R, for $n \in \mathbb{Z}$. Now, under $\sigma \sim \sigma + 2\pi$, $X^{25}(z, \bar{z})$ is not single valued, and can change by $2\pi w R$, for $w \in \mathbb{Z}$. Solving the two resulting equations gives:

$$\begin{aligned} \alpha_0^{25} + \tilde{\alpha}_0^{25} &= \frac{2n}{R}\sqrt{\frac{\alpha'}{2}} \\ \alpha_0^{25} - \tilde{\alpha}_0^{25} &= \sqrt{\frac{2}{\alpha'}}wR \end{aligned} \tag{4.12}$$

and so we have:

$$\begin{aligned} \alpha_0^{25} &= \left(\frac{n}{R} + \frac{wR}{\alpha'}\right)\sqrt{\frac{\alpha'}{2}} \equiv P_{\mathrm{L}}\sqrt{\frac{\alpha'}{2}} \\ \tilde{\alpha}_0^{25} &= \left(\frac{n}{R} - \frac{wR}{\alpha'}\right)\sqrt{\frac{\alpha'}{2}} \equiv P_{\mathrm{R}}\sqrt{\frac{\alpha'}{2}}. \end{aligned} \tag{4.13}$$

We can use this to compute the formula for the mass spectrum in the remaining uncompactified 24+1 dimensions, using the fact that $M^2 = -p_\mu p^\mu$, where now $\mu = 0, \ldots, 24$.

$$\begin{aligned} M^2 = -p^\mu p_\mu &= \frac{2}{\alpha'}(\alpha_0^{25})^2 + \frac{4}{\alpha'}(N-1) \\ &= \frac{2}{\alpha'}(\tilde{\alpha}_0^{25})^2 + \frac{4}{\alpha'}(\bar{N}-1), \end{aligned} \tag{4.14}$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

where $N, \bar{N}$ denote the total levels on the left- and right-moving sides, as before. These equations follow from the left and right $L_0, \bar{L}_0$ constraints. Recall that the sum and difference of these give the Hamiltonian and the level-matching formulae. Here, they are modified, and a quick computation gives:

$$\begin{aligned} M^2 &= \frac{n^2}{R^2} + \frac{w^2R^2}{\alpha'^2} + \frac{2}{\alpha'}\left(N + \tilde{N} - 2\right) \\ nw &+ N - \tilde{N} = 0. \end{aligned} \tag{4.15}$$

The key features here are that there are terms in addition to the usual oscillator contributions. In the mass formula, there is a term giving the familiar contribution of the Kaluza–Klein tower of momentum states for the string (see section 4.1.1), and a new term from the tower of winding states. This latter term is a very stringy phenomenon. Notice that the level matching term now also allows a mismatch between the number of left and right oscillators excited, in the presence of discrete winding and momenta.

In fact, notice that we can get our usual massless Kaluza–Klein states* by taking

$$n = w = 0; \qquad N = \bar{N} = 1, \tag{4.16}$$

exciting an oscillator in the compact direction. There are two ways of doing this, either on the left or the right, and so there are two $U(1)$s following from the fact that there is an internal component of the metric and also of the antisymmetric tensor field. We can choose to identify the two gauge fields of this $U(1) \times U(1)$ as follows:

$$A_{\mu(\mathrm{R})} \equiv \frac{1}{2}(G - B)_{\mu,25}; \quad A_{\mu(\mathrm{L})} \equiv \frac{1}{2}(G + B)_{\mu,25}.$$

We have written these states out explicitly, together with the corresponding spacetime fields, and the vertex operators (at zero momentum), below.

field	state	operator
$G_{\mu\nu}$	$(\alpha^{\mu}_{-1}\tilde{\alpha}^{\nu}_{-1} + \alpha^{\nu}_{-1}\tilde{\alpha}^{\mu}_{-1})\vert 0;k\rangle$	$\partial X^{\mu}\partial X^{\nu} + \partial X^{\mu}\partial X^{\nu}$
$B_{\mu\nu}$	$(\alpha^{\mu}_{-1}\tilde{\alpha}^{\nu}_{-1} - \alpha^{\nu}_{-1}\tilde{\alpha}^{\mu}_{-1})\vert 0;k\rangle$	$\partial X^{\mu}\bar{\partial} X^{\nu} - \partial X^{\mu}\bar{\partial} X^{\nu}$
$A_{\mu(R)}$	$\alpha^{\mu}_{-1}\tilde{\alpha}^{25}_{-1}\vert 0;k\rangle$	$\partial X^{\mu}\bar{\partial} X^{25}$
$A_{\mu(L)}$	$\tilde{\alpha}^{\mu}_{-1}\alpha^{25}_{-1}\vert 0;k\rangle$	$\partial X^{25}\bar{\partial} X^{\mu}$
$\phi \equiv \frac{1}{2}\log G_{25,25}$	$\alpha^{25}_{-1}\tilde{\alpha}^{25}_{-1}\vert 0;k\rangle$	$\partial X^{25}\bar{\partial} X^{25}$

* We shall sometimes refer to Kaluza–Klein states as 'momentum' states, to distinguish them from 'winding' states, in what follows.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

So we have these 25-dimensional massless states which are basically the components of the graviton and antisymmetric tensor fields in 26 dimensions, now relabelled. (There is also of course the dilaton Φ, which we have not listed.) There is a pair of gauge fields giving a $U(1)_{\rm L}\times U(1)_{\rm R}$ gauge symmetry, and in addition a massless scalar field ϕ. Actually, ϕ is a massless scalar which can have any background vacuum expectation value (vev), which in fact sets the radius of the circle. This is because the square root of the metric component $G_{25,25}$ is indeed the measure of the radius of the X^{25} direction.

4.2 T-duality for closed strings

Let us now study the generic behaviour of the spectrum (4.15) for different values of R. For larger and larger R, momentum states become lighter, and therefore it is less costly to excite them in the spectrum. At the same time, winding states become heavier, and are more costly. For smaller and smaller R, the reverse is true, and it is gets cheaper to excite winding states while it is momentum states which become more costly.

We can take this further: as $R\to\infty$, all of the winding states, i.e. states with $w\neq 0$, become infinitely massive, while the $w=0$ states with all values of n go over to a continuum. This fits with what we expect intuitively, and we recover the fully uncompactified result.

Consider instead the case $R\to 0$, where all of the momentum states, i.e. states with $n\neq 0$, become infinitely massive. If we were studying field theory we would stop here, as this would be all that would happen – the surviving fields would simply be independent of the compact coordinate, and so we have performed a dimension reduction. In closed string theory things are quite different: the pure winding states (i.e. $n=0$, $w\neq 0$, states) form a continuum as $R\to 0$, following from our observation that it is very cheap to wind around the small circle. *Therefore, in the $R\to 0$ limit, an effective uncompactified dimension actually reappears!*

Notice that the formula (4.15) for the spectrum is invariant under the exchange

$$n\leftrightarrow w \qquad \text{and} \qquad R\leftrightarrow R'\equiv \alpha'/R. \tag{4.17}$$

The string theory compactified on a circle of radius R' (with momenta and windings exchanged) is the 'T-dual' theory[14], and the process of going from one theory to the other will be referred to as 'T-dualising'.

The exchange takes (see (equation 4.13))

$$\alpha_0^{25}\to\alpha_0^{25}, \quad \tilde\alpha_0^{25}\to-\tilde\alpha_0^{25}. \tag{4.18}$$

The dual theories are identical in the fully interacting case as well (after a shift of the coupling to be discussed shortly)[15]. Simply rewrite the radius

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

R theory by performing the exchange

$$X^{25}(z,\bar{z}) = X^{25}(z) + X^{25}(\bar{z}) \longrightarrow X'^{25}(z,\bar{z}) = X^{25}(z) - X^{25}(\bar{z}). \quad (4.19)$$

The energy-momentum tensor and other basic properties of the conformal field theory are invariant under this rewriting, and so are therefore all of the correlation functions representing scattering amplitudes, etc. The only change, as follows from equation (4.18), is that the zero mode spectrum in the new variable is that of the α'/R theory.

So these theories are physically identical. T-duality, relating the R and α'/R theories, is an exact symmetry of perturbative closed string theory. Shortly, we shall see that it is non-perturbatively exact as well.

N.B. The transformation (4.19) can be regarded as a spacetime parity transformation acting only on the right-moving (in the world sheet sense) degrees of freedom. We shall put this picture to good use in what is to come.

4.3 A special radius: enhanced gauge symmetry

Given the relation we deduced between the spectra of strings on radii R and α'/R, it is clear that there ought to be something interesting about the theory at the radius $R = \sqrt{\alpha'}$. The theory should be self-dual, and this radius is the 'self-dual radius'. There is something else special about this theory besides just self-duality.

At this radius we have, using (4.13),

$$\alpha_0^{25} = \frac{(n+w)}{\sqrt{2}}; \qquad \tilde{\alpha}_0^{25} = \frac{(n-w)}{\sqrt{2}}, \quad (4.20)$$

and so from the left and right we have:

$$\begin{aligned} M^2 = -p^\mu p_\mu &= \frac{2}{\alpha'}(n+w)^2 + \frac{4}{\alpha'}(N-1) \\ &= \frac{2}{\alpha'}(n-w)^2 + \frac{4}{\alpha'}(\bar{N}-1). \end{aligned} \quad (4.21)$$

So if we look at the massless spectrum, we have the conditions:

$$(n+w)^2 + 4N = 4; \qquad (n-w)^2 + 4\bar{N} = 4. \quad (4.22)$$

As solutions, we have the cases $n = w = 0$ with $N = 1$ and $\bar{N} = 1$ from before. These are include the vectors of the $U(1) \times U(1)$ gauge symmetry of the compactified theory.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

Now, however, we see that we have more solutions. In particular:

$$n=-w=\pm 1, \quad N=1,\ \bar{N}=0; \qquad n=w=\pm 1, \quad N=0,\ \bar{N}=1. \tag{4.23}$$

The cases where the excited oscillators are in the non-compact direction yield two pairs of massless vector fields. In fact, the first pair go with the left $U(1)$ to make an $SU(2)$, while the second pair go with the right $U(1)$ to make another $SU(2)$. Indeed, they have the correct ± 1 charges under the Kaluza–Klein $U(1)$s in order to be the components of the W-bosons for the $SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}}$ 'enhanced gauge symmetries'. The term is appropriate since there is an extra gauge symmetry at this special radius, given that new massless vectors appear there.

When the oscillators are in the compact direction, we get two pairs of massless bosons. These go with the massless scalar ϕ to fill out the massless adjoint Higgs field for each $SU(2)$. These are the scalars whose vevs give the W-bosons their masses when we are away from the special radius.

In fact, this special property of the string theory is succinctly visible at all mass levels, by looking at the partition function (4.30). At the self-dual radius, it can be rewritten as a sum of squares of 'characters' of the $su(2)$ affine Lie algrebra:

$$Z(q, R=\sqrt{\alpha'}) = |\chi_1(q)|^2 + |\chi_2(q)|^2, \tag{4.24}$$

where

$$\chi_1(q) \equiv \eta^{-1}\sum_n q^{n^2}, \quad \chi_2(q) \equiv \eta^{-1}\sum_n q^{(n+1/2)^2}. \tag{4.25}$$

It is amusing to expand these out (after putting in the other factors of $(\eta\bar{\eta})^{-1}$ from the uncompactified directions) and find the massless states we discussed explicitly above.

It does not matter if an affine Lie algebra has not been encountered before by the reader. We can take this as an illustrative example, arising in a natural and instructive way. See insert 4.1 for further discussion[12]. In the language of two dimensional conformal field theory, there are additional left- and right-moving currents (i.e. fields with weights (1,0) and (0,1)) present. We can construct them as vertex operators by exponentiating some of the existing fields. The full set of vertex operators of the $SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}}$ spacetime gauge symmetry:

$$\begin{aligned} SU(2)_{\mathrm{L}}&: \ \bar{\partial}X^\mu \partial X^{25}(z), \quad \bar{\partial}X^\mu \exp(\pm 2iX^{25}(z)/\sqrt{\alpha'}) \\ SU(2)_{\mathrm{R}}&: \ \partial X^\mu \bar{\partial} X^{25}(z), \quad \partial X^\mu \exp(\pm 2iX^{25}(\bar{z})/\sqrt{\alpha'}), \end{aligned} \tag{4.26}$$

corresponding to the massless vectors we constructed by hand above.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

Insert 4.1. Affine Lie algebras

The key structure of an affine Lie algebra is just what we have seen arise naturally in this self-duality example. In addition to all of the nice structures that the conformal field theory has – most pertinently, the Virasoro algebra – there is a family of unit weight operators, often constructed as vertex operators as we saw in equation (4.26), which form the Lie algebra of some group G. They are unit weight as measured either from the left or the right, and so we can have such structures on either side. Let us focus on the left. Then, as $(1,0)$ operators, $J^a(z)$, (a is a label) we have:

$$[L_n, J^a_m] = mJ^a_{n+m}, \tag{4.27}$$

where

$$J^a_n = \frac{1}{2\pi i}\oint dz \quad z^{-n-1}J^a(z), \tag{4.28}$$

and

$$[J^a_n, J^b_m] = if^{ab}{}_c J^c_{n+m} + mkd^{ab}\delta_{n+m}, \tag{4.29}$$

where it should be noticed that the zero modes of these currents form a Lie algebra, with structure constants $f^{ab}{}_c$. The constants d^{ab} define the inner product between the generators $(t^a, t^b) = d^{ab}$. Since in bosonic string theory a mode with index -1 creates a state that is massless in spacetime, J^a_{-1} can be placed either on the left with $\tilde{\alpha}^\mu_{-1}$ on the right (or vice versa) to give a state $J^a_{-1}\tilde{\alpha}^\mu_{-1}|0\rangle$ which is a massless vector $A^{\mu a}$ in the adjoint of G, for which the low energy physics must be Yang–Mills theory.

The full algebra is called an 'affine Lie algebra', or a 'current algebra', and sometimes a 'Kac–Moody' algebra[275]. In a standard normalisation, k is an integer and is called the 'level' of the affinisation. In the case that we first see this sort of structure, the string at a self-dual radius, the level is 1. The currents in this case are:

$$J^3(z) = i\alpha'^{-1/2}\partial_z X^{25}(z),$$
$$J^1(z) = :\cos(2\alpha'^{-1/2}X^{25}(z)):, \quad J^2(z) =: \sin(2\alpha'^{-1/2}X^{25}(z)):$$

which satisfy the algebra in (4.29) with $f^{abc} = \epsilon^{abc}$, $k = 1$, and $d^{ab} = \frac{1}{2}\delta^{ab}$, as appropriate to the fundamental representation.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

The vertex operator for the change of radius, $\partial X^{25}\bar{\partial}X^{25}$, corresponding to the field ϕ, transforms as a $(\mathbf{3},\mathbf{3})$ under $SU(2)_{\rm L}\times SU(2)_{\rm R}$, and therefore a rotation by π in one of the $SU(2)$s transforms it into minus itself. The transformation $R\to\alpha'/R$ is therefore the $\mathbb{Z}_2$ Weyl subgroup of the $SU(2)\times SU(2)$. Since T-duality is part of the spacetime gauge theory, this is a clue that it is an exact symmetry of the closed string theory, if we assume that non-perturbative effects preserve the spacetime gauge symmetry. We shall see that this assumption seems to fit with non-perturbative discoveries to be described later.

4.4 The circle partition function

It is useful to consider the partition function of the theory on the circle. This is a computation as simple as the one we did for the uncompactified theory earlier, since we have done the hard work in working out L_0 and $\bar{L}_0$ for the circle compactification. Each non-compact direction will contribute a factor of $(\eta\bar{\eta})^{-1}$, as before, and the non-trivial part of the final τ-integrand, coming from the compact X^{25} direction is:

$$Z(q,R)=(\eta\bar{\eta})^{-1}\sum_{n,w}q^{\frac{\alpha'}{4}P_{\rm L}^2}\bar{q}^{\frac{\alpha'}{4}P_{\rm R}^2}, \tag{4.30}$$

where $P_{\rm L,R}$ are given in (4.13). Our partition function is manifestly T-dual, and is in fact also modular invariant. Under T, it picks us a phase $\exp(\pi i(P_{\rm L}^2-P_{\rm R}^2))$, which is again unity, as follows from the second line in (4.15): $P_{\rm L}^2-P_{\rm R}^2=2nw$. Under S, the role of the time and space translations as we move on the torus are exchanged, and this in fact exchanges the sums over momentum and winding. T-duality ensures that the S-transformation properties of the exponential parts involving $P_{\rm L,R}$ are correct, while the rest is S invariant as we have already discussed.

It is a useful exercise to expand this partition function out, after combining it with the factors from the other non-compact dimensions first, to see that at each level the mass (and level matching) formulae (4.15) which we derived explicitly is recovered.

In fact, the modular invariance of this circle partition function is part of a very important larger story. The left and right momenta $P_{\rm L,R}$ are components of a special two dimensional lattice, $\Gamma_{1,1}$. There are two basis vectors $k=(1/R,1/R)$ and $\hat{k}=(R,-R)$. We make the lattice with arbitrary integer combinations of these, $nk+w\hat{k}$, whose components are $(P_{\rm L},P_{\rm R})$. (cf. equation (4.13)). If we define the dot products between our basis vectors to be $k\cdot\hat{k}=2$ and $k\cdot k=0=\hat{k}\cdot\hat{k}$, our lattice then has a Lorentzian signature, and since $P_{\rm L}^2-P_{\rm R}^2=2nw\in 2\mathbb{Z}$, it is called

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

'even'. The 'dual' lattice $\Gamma^*_{1,1}$ is the set of all vectors whose dot product with $(P_{\rm L}, P_{\rm R})$ gives an integer. In fact, our lattice is self-dual, which is to say that $\Gamma_{1,1} = \Gamma^*_{1,1}$. It is the 'even' quality which guarantees invariance under T as we have seen, while it is the 'self-dual' feature which ensures invariance under S. In fact, S is just a change of basis in the lattice, and the self-duality feature translates into the fact that the Jacobian for this is unity.

4.5 Toriodal compactifications

It will be very useful later on for us to outline how things work more generally. The case of compactification on the circle encountered above can be easily generalised to compactification on the torus $T^d \simeq (S^1)^d$. Let us denote the compact dimensions by X^m, where $m, n = 1, \ldots, d$. Their periodicity is specified by

$$X^m \sim X^m + 2\pi R^{(m)}\mathsf{n}^m,$$

where the n^m are integers and $R^{(m)}$ is the radius of the mth circle. The metric on the torus, G_{mn}, can be diagonalised into standard unit Euclidean form by the veilbeins e^a_m where $a, b = 1, \ldots, d$:

$$G_{mn} = \delta_{ab} e^a_m e^b_n,$$

and it is convenient to use tangent space coordinates $X^a = X^m e^a_m$ so that the equivalence can be written:

$$X^a \sim X^a + 2\pi e^a_m \mathsf{n}^m.$$

We have defined for ourselves a lattice $\Lambda = \{e^a_m \mathsf{n}^m, \mathsf{n}^m \in \mathbb{Z}\}$. We now write our torus in terms of this as

$$T^d \equiv \frac{\mathbb{R}^d}{2\pi\Lambda}.$$

There are of course conjugate momenta to the X^a, which we denote as p^a. They are quantised, since moving from one lattice point to another, producing a change in the vector X by $\delta X \in 2\pi\Lambda$ are physically equivalent, and so single-valuedness of the wavefunction imposes $\exp(ip \cdot X) = \exp(ip \cdot [X + \delta X])$, i.e.

$$p \cdot \delta X \in 2\pi\mathbb{Z},$$

from which we see that clearly

$$p^n = G^{mn}\mathsf{n}_m,$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

where n_m are integers. In other words, the momenta live in the dual lattice, Λ^*, of Λ, defined by

$$\Lambda^* \equiv \{e^{*am}\mathsf{n}_m, \mathsf{n}_m \in \mathbb{Z}\},$$

where the inverse veilbiens $e^{*am}\mathsf{n}_m$ are defined in the usual way using the inverse metric:

$$e^{*am} \equiv e^a_m G^{mn}, \quad \text{or} \quad e^{*am}e^b{}_m = \delta^{ab}.$$

Of course we can have winding sectors as well, since as we go around the string via $\sigma \to \sigma + 2\pi$, we can change to a new point on the lattice characterised by a set of integers w^m, the winding number. Let us write out the string mode expansions. We have

$$X^a(\tau,\sigma) = X^a_{\rm L}(\tau-\sigma) + X^a_{\rm R}(\tau+\sigma), \qquad \text{where}$$

$$\begin{aligned} X^a_{\rm L} &= x^a_{\rm L} - i\sqrt{\frac{\alpha'}{2}}p^a_{\rm L}(\tau-\sigma) + \text{oscillators} \qquad x^a_{\rm L} = \frac{x^a}{2} - \theta^a \\ p^a_{\rm L} &= p^a + \frac{w^a R^{(a)}}{\alpha'} \equiv e^{*am}\mathsf{n}_m + \frac{1}{\alpha'}e^a_m w^m, \end{aligned} \tag{4.31}$$

for the left, while on the right we have

$$\begin{aligned} X^a_{\rm R} &= x^a_{\rm R} - i\sqrt{\frac{\alpha'}{2}}p^a_{\rm R}(\tau+\sigma) + \text{oscillators} \qquad x^a_{\rm R} = \frac{x^a}{2} + \theta^a \\ p^a_{\rm R} &= p^a - \frac{w^a R^{(a)}}{\alpha'} \equiv e^{*am}\mathsf{n}_m - \frac{1}{\alpha'}e^a_m w^m. \end{aligned} \tag{4.32}$$

The action of the manifest T-duality symmetry is simply to act with a right-handed parity, as before, swopping $p_{\rm L} \leftrightarrow p_{\rm L}$ and $p_{\rm R} \leftrightarrow -p_{\rm R}$, and hence momenta and winding and $X_{\rm L} \leftrightarrow X_{\rm L}$ and $X_{\rm R} \leftrightarrow -X_{\rm R}$.

To see more, let us enlarge our bases for the two separate lattices Λ, Λ^* into a singe one, via:

$$\hat{e}_m = \frac{1}{\alpha'}\begin{pmatrix} e^a_m \\ -e^a_m \end{pmatrix}, \quad \hat{e}^{*m} = \begin{pmatrix} e^{*am} \\ e^{*am} \end{pmatrix},$$

and now we can write

$$\hat{p} = \begin{pmatrix} p^a_{\rm L} \\ p^a_{\rm R} \end{pmatrix} = \hat{e}_m w^m + \hat{e}^{*m}\mathsf{n}_m,$$

which lives in a $(d+d)$-dimensional lattice which we will call $\Gamma_{d,d}$. We can choose the metric on this space to be of Lorentzian signature (d,d),

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

which is achieved by

$$G = \begin{pmatrix} \delta_{ab} & 0 \\ 0 & -\delta_{ab} \end{pmatrix},$$

and using this we see that

$$\begin{aligned} \hat{e}_m \cdot \hat{e}_n &= 0 = \hat{e}^{*m} \cdot \hat{e}^{*n} \\ \hat{e}_m \cdot \hat{e}^{*n} &= \frac{2}{\alpha'} \delta_n^m, \end{aligned} \tag{4.33}$$

which shows that the lattice is *self-dual*, since (up to a trivial overall scaling), the structure of the basis vectors of the dual is identical to that of the original: $\Gamma^*_{d,d} = \Gamma_{d,d}$. Furthermore, we see that the inner product between any two momenta is given by

$$(\hat{e}_m w^m + \hat{e}^{*m} \mathsf{n}_m) \cdot (\hat{e}_n w'^n + \hat{e}^{*n} \mathsf{n}'_n) = \frac{2}{\alpha'}(w^m \mathsf{n}'_m + \mathsf{n}_m w'^m). \tag{4.34}$$

In other words, the lattice is *even*, because the inner product gives even integer multiples of $2/\alpha'$.

It is these properties that guarantee that the string theory is modular invariant[173]. The partition function for this compactification is the obvious generalisation of the expression given in (4.30):

$$Z_{T^d} = (\eta\bar{\eta})^{-d} \sum_{\Gamma_{d,d}} q^{\frac{\alpha'}{4} p_{\mathrm{L}}^2} \bar{q}^{\frac{\alpha'}{4} p_{\mathrm{R}}^2}, \tag{4.35}$$

where the $p_{\mathrm{L,R}}$ are given in (4.32). Recall that the modular group is generated by $T : \tau \to \tau + 1$, and $S : \tau \to -1/\tau$. So T-invariance follows from the fact that its action produces a factor $\exp(i\pi\alpha'(p_{\mathrm{L}}^2 - p_{\mathrm{R}}^2)/2) = \exp(i\pi\alpha'(\hat{p}^2)/2)$ which is unity because the lattice is even, as shown in equation (4.34).

Invariance under S follows by rewriting the partition function $Z(-1/\tau)$ using the Poisson resummation formula given in insert 4.2, to get the result that

$$Z_\Gamma\left(-\frac{1}{\tau}\right) = \mathrm{vol}(\Gamma^*) Z_{\Gamma^*}(\tau).$$

The volume of the lattice's unit cell is unity, for a self-dual lattice, since $\mathrm{vol}(\Lambda)\mathrm{vol}(\Lambda^*) = 1$ for any lattice and its dual, and therefore S-invariance is demonstrated, and we can define a consistent string compactification.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

Insert 4.2. The Poisson resummation formula

A very useful trick is the following. Assume that we have a function $f(x)$ defined on $\mathbb{R}^n$. Then its Fourier transform is given as

$$f(x) = \int \frac{d^n k}{(2\pi)^n} e^{ik\cdot x} \hat{f}(k).$$

The formula we need is written in terms of this. If we sum over a lattice $\Lambda \subset \mathbb{R}^n$, then:

$$\sum_{n\in\Lambda} f(n) = \int \sum_{n\in\Lambda} \frac{d^n k}{(2\pi)^n} e^{ik\cdot m} \hat{f}(k) = \mathrm{vol}(\Lambda^*) \sum_{m\in\Lambda^*} \hat{f}(2\pi m).$$

We shall meet two very important examples of large even and self-dual lattices later in subsection 7.2. They are associated to the construction of the modular invariant partition functions of the ten dimensional $E_8 \times E_8$ and $SO(32)$ heterotic strings[20].

There is a large space of inequivalent lattices of the type under discussion, given by the shape of the torus (specified by background parameters in the metric G) and the fluxes of the B-field through it. We can work out this 'moduli space' of compactifications. It would naively seem to be simply $O(d,d)$, since this is the space of rotations naturally acting, taking such lattices into each other, i.e. starting with some reference lattice Γ_0, $\Gamma' = G\Gamma_0$ should be a different lattice. We must remember that the physics cares only about the values of p_L^2 and p_R^2, and so therefore we must count as equivalent any choices related by the $O(d)\times O(d)$ which acts independently on the left and right momenta: $G \sim G'G$, for $G' \in O(d)\times O(d)$. So at least *locally*, the space of lattices is isomorphic to

$$\mathcal{M} = \frac{O(d,d)}{O(d)\times O(d)}. \tag{4.36}$$

A quick count of the dimension of this space gives $2d(2d-1)/2 - 2 \times d(d-1)/2 = d^2$, which fits nicely, since this is the number of independent components contained in the metric G_{mn}, $(d(d+1)/2)$ and the antisymmetric tensor field B_{mn}, $(d(d-1)/2)$, for which we can switch on constant values (sourced by winding).

There are still a large number of discrete equivalences between the lattices, which follows from the fact that there is a discrete subgroup of

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

$O(d, d)$, called $O(d, d, \mathbb{Z})$, which maps our reference lattice Γ_0 into itself: $\Gamma_0 \sim G''\Gamma_0$. This is the set of discrete linear transformations generated by the subgroups of $SL(2d, \mathbb{Z})$ which preserves the inner product given in equations (4.33). This group includes the T-dualities on all of the d circles, linear redefinitions of the axes, and discrete shifts of the B-field. The full space of torus compactifications is often denoted:

$$\mathcal{M} = O(d, d, \mathbb{Z})\backslash O(d, d)/[O(d) \times O(d)], \tag{4.37}$$

where we divide by one action under left multiplication, and the other under right.

Now we see that there is a possibility of much more than just the $SU(2)_{\rm L} \times SU(2)_{\rm R}$ enhanced gauge symmetry which we got in the case of a single circle. We can have this large symmetry from any of the d circles, of course but there is more, since there are extra massless states that can be made by choices of momenta from more than one circle, corresponding to weight one vertex operators. This will allow us to make very large enhanced gauge groups, up to rank d, as we shall see later in section 7.2.

4.6 More on enhanced gauge symmetry

The reader is probably keen to see more of where some of the structures of sections 4.3, 4.4, and 4.5 come from, and so we will pause here to study a little about Lie groups and algebras.

4.6.1 Lie algebras and groups

Lie algebras are usually described in terms of a basis of *generators*, t^a, which have a specific antisymmetric product:

$$[t^a, t^b] = if^{ab}{}_c t^c, \tag{4.38}$$

where the $f^{ab}{}_c$ are often called the *structure constants*. This product must satisfy the Jacobi identity, which states that:

$$[t^a, [t^b, t^c]] + [t^b, [t^c, t^a]] + [t^c, [t^a, t^b]] = 0.$$

Once we have the algebra, we can form the group G by exponentiating the generators, to make a group element

$$g = e^{i\lambda_a t^a}.$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

N.B. One of the reasons why Lie groups are interesting is that *the group elements form a manifold*, and so there is a lot of familiar geometry to be found in their description. For example, one can think of the Lie algebra as the vector space that is simply the tangent space to the group manifold, G, and keep in mind a picture like that in figure 2.14. The natural way to make the Lie algebra from the group elements g is *via* the Maurer–Cartan forms, $g^{-1}dg$ which give a family of one-forms which are valued in the Lie algebra. We won't use this much, but the curious reader can look ahead to insert 7.4, where we make this explicit for $SU(2)$, which is the manifold S^3.

There is also an inner product between the generators, which is defined as $(t^a, t^b) = d^{ab}$, which is positive if the group is compact. We can lower and raise indices with this fellow, and having done this on the structure constants to get f^{abc}, there is an additional condition that they are totally antisymmetric in all of their indices. We shall restrict our attention mostly to the *simple* Lie algebras, for which a choice can be made to make d^{ab} proportional to δ^{ab}.

Most familiar is of course the representation of the algebra in (4.38) by matrices, for which we can use the notation t^a_R, where R stands for a representation, and the matrix elements are denoted $t^a_{R,ij}$. The antisymmetric product is then the familiar matrix commutator, and the inner product is matrix multiplication with the trace. Then we have $\mathrm{Tr}(t^a_R t^b_R) = T_R \delta^{ab}$, where T_R is a number which depends on the representation. Note that we can define the *Casimir invariant* of the representation R as $t^a_R t^b_R = Q_R \mathbf{1}$.

The Jacobi identity above translates into

$$f^{abd} f^{cde} + f^{bcd} f^{ade} + f^{cad} f^{bde} = 0.$$

A most convenient matrix representation of the algebra is given by

$$(t^a_A)_{bc} = -i f^a{}_{bc},$$

and for this we see that we get

$$[t^a_A, t^b_A] = i f^a{}_{bc} t^c_A,$$

and so we see that the structure constants themselves form a representation of the Lie algebra. This is the *adjoint representation*. Notice that the dimension of the representation is the number of generators of the group.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

It is useful to divide the generators t^a into two families. There is the maximal set of commuting generators, which are denoted H^i, where $i = 1, \ldots, r$ with r being the *rank* of the group, and there are the rest, denoted E^α of reasons to be given very shortly.

The set H^i, for which

$$[H^i, H^j] = 0,$$

is the *Cartan subalgebra*, and the H^i are often said to form the *maximal torus*, which we shall discuss more later. These elements are the generalisation of J_3 from the familiar case of $SU(2)$. For a representation of dimension d, we can think of the H^i as $d \times d$ matrices. We will pick a specific basis for these and keep in that basis to describe everything else. Being all mutually commutative, they may be simultaneously diagonalised, and there are d distinct eigenvalues for each H. Consider the nth entry along a diagonal. Each of the H^i supplies a component, w^i, of a vector w in a space $\mathbb{R}^r$. There are d such *weight vectors*.

Everything else can be given an assignment of 'charges' corresponding to the H-eigenvalues, via

$$[H^i, E^\alpha] = \alpha^i E^\alpha.$$

We can think of the α^i as components of an r-dimensional vector known as a root. It is a vector in the space $\mathbb{R}^r$ mentioned above. Every root is uniquely associated to a generator E^α. The remaining parts of the Lie algebra are:

$$[E^\alpha, E^\beta] = \begin{cases} \epsilon(\alpha, \beta) E^{\alpha,\beta} & \text{if } \alpha + \beta \text{ is a root,} \\ 2\alpha \cdot H/\alpha \cdot \alpha & \text{if } \alpha + \beta = 0, \\ 0 & \text{otherwise,} \end{cases}$$

where the dot product is defined with the relevant part of the inner product form, d_{ij}, and $\epsilon(\alpha, \beta)$ is ± 1. It is worth noting that the roots are the weights of the adjoint representation.

The E^α are the generalisations of the $J^\pm$ familiar from $SU(2)$, the raising and lowering operators. One can decompose weights into three classes, whether they are positive, negative, or zero. This is given by whether or not the first non-zero entry is positive, negative or zero (i.e. all components zero). There is a unique highest weight in any representation. Specialising to the weights of the adjoint representation, the roots, divides the E^α into raising operators, if α is positive, and lowering operators if α is negative. One can build the whole representation of the groups starting with the highest weight and acting with the lowering operators, while acting on a highest weight with a raising operator gives zero.

The *simple roots* are the positive roots that cannot be written as the sum of two positive roots, and they form a linearly independent set. The

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

number of them is equal to the rank of the group, r. Using these, it can be shown that the entire structure of the group may be reconstructed. A useful way of specifying the simple roots is to give their relative lengths and the angles between them, which turn out to be restricted to between 90° and 180°. The *Dynkin diagram* is a very useful way of giving that information in an easy to read form. Each simple root is a node in the diagram. There are links between nodes if the angle between them is not 90°. There is a single line if the angle is 120°, a double line if the angle is 135° and a triple line if it is 150°. To denote the odd root which is shorter than the rest, it is often a practice to make the note a different shade of colour in the diagram.

4.6.2 The classical Lie algebras

Let us list the *classical Lie algebras* of Cartan's classification.

- $SU(n)$ Denoted A_{n-1} in Cartan's classification. The generators are traceless $n\times n$ Hermitian matrices, and the group elements of $SU(n)$ are unit determinant unitary matrices.

- $SO(n)$ If $n = 2k+1$ this is denoted B_k, while if $n = 2k$ it is D_k. The generators are $n \times n$ antisymmetric Hermitian matrices, and the group elements of $SO(n)$ are real orthogonal matrices.

- $Sp(k) = USp(2k)$ This is denoted C_k in the classification. The generators are Hermitian $2k \times 2k$ matrices t satisfying

  $$MtM^{-1} = -t^T,$$

  where T denotes the transpose and

  $$M = i\begin{pmatrix} 0 & I_k \\ -I_k & 0 \end{pmatrix},$$

  where I_k is the $k\times k$ identity matrix. The groups is the set of unitary matrices u satisfying

  $$MuM^{-1} = u^{-T},$$

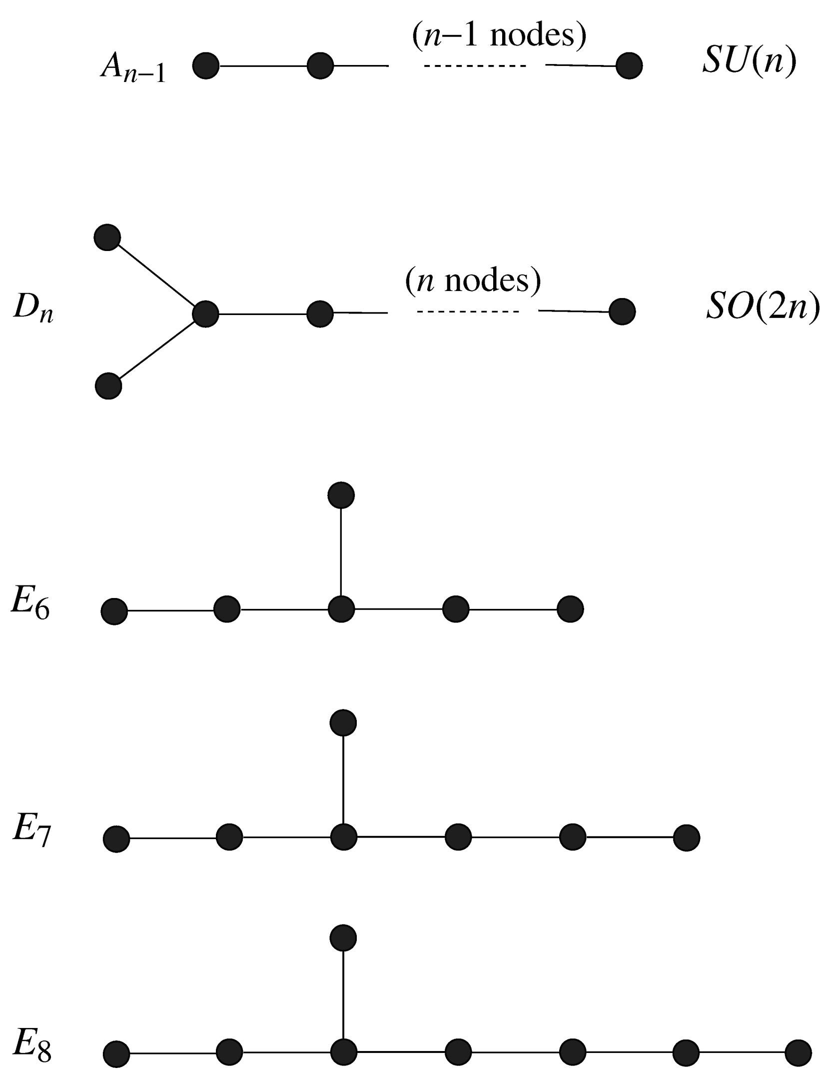
  where $-T$ denotes the inverse of the transpose.

We will often have cause to encounter some non-compact groups closely related to these. We obtain them by multiplying some generators by an i. In this way we will get the set of traceless imaginary matrices to make the group of real matrices of unit determinant, $SL(n)$ by continuing $SU(n)$. We have already encountered $O(n, m)$, which is a continuation of $O(n+m)$ made by such a continuation.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

Insert 4.3. The simply laced Lie algebras

It turns out that for the Lie algebras A_n, D_k, E_6, E_7 and E_8, all of the roots are the same length. These are called the *simply laced* algebras. It is very useful to know a bit about their structure, as manifest in the Dynkin diagrams given below.



https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

4.6.3 Physical realisations with vertex operators

Now we can return to some of the physical objects that we saw arising in the string theory and make contact with some of the structures we saw above. Recall that we represented the weights as vectors in $\mathbb{R}^r$, where r was the rank of the Lie algebra, arising as charges under the commuting generators or maximal torus given by the H^i. These vectors came with a specific set of entries, and we could build all representations out of them, by adding vectors. The set of points in $\mathbb{R}^r$ made in this way is the *Lie algebra lattice*, and it can be placed on a very physical footing in the context of toroidal compactification in the following way.

If we placed r directions X^i on a torus T^r, the weight (0,1) objects $H^i(z) = i\alpha'^{-1/2}\partial_z X^i$ parameterise the very object we have been working with: the maximal torus. The weight vectors that we had, with the additive structure allowing us to reach other points in the lattice, building up other representations, are simply the momenta, which are the zero modes of the $H^i(z)$, which are also additive.

In general, we can make states corresponding to the weight vector w^i with the vertex operator $\exp(2i\alpha'^{-1/2}w\cdot\phi)$. So now we see how to get a gauge symmetry, following the discussion in insert 4.1, we need to have vertex operators of weight (0,1) to go with the $H^i(z)$. These can be made with the vertex operators if the $w^2 = 2$. So we see that we need the simply laced algebras to do this. They are listed in insert 4.3, together with their Dynkin diagrams.

4.7 Another special radius: bosonisation

Before proceeding with the T-duality discussion, let us pause for a moment to remark upon something which will be useful later. In the case that $R = \sqrt{(\alpha'/2)}$, something remarkable happens. The partition function is:

$$Z\left(q, R = \sqrt{\frac{\alpha'}{2}}\right) = (\eta\bar{\eta})^{-1}\sum_{n,w} q^{\frac{1}{2}\left(n+\frac{w}{2}\right)^2}\bar{q}^{\frac{1}{2}\left(n-\frac{w}{2}\right)^2}. \tag{4.39}$$

Note that the allowed momenta at this radius are (cf. equation (4.13)):

$$\begin{aligned}\alpha_0^{25} &= P_{\rm L}\sqrt{\frac{\alpha'}{2}} = \left(n+\frac{w}{2}\right)\\ \tilde{\alpha}_0^{25} &= P_{\rm R}\sqrt{\frac{\alpha'}{2}} = \left(n-\frac{w}{2}\right),\end{aligned} \tag{4.40}$$

and so they span both integer and half-integer values. Now when $P_{\rm L}$ is an integer, then so is $P_{\rm R}$ and vice versa, and so we have two distinct sectors,

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

integer and half-integer. In fact, we can rewrite our partition function as a set of sums over these separate sectors:

$$Z_{R=\sqrt{\alpha'/2}}$$
$$= \frac{1}{2}\left\{\left|\frac{1}{\eta}\sum_n q^{\frac{1}{2}n^2}\right|^2 + \left|\frac{1}{\eta}\sum_n (-1)^n q^{\frac{1}{2}n^2}\right|^2 + \left|\frac{1}{\eta}\sum_n q^{\frac{1}{2}\left(n+\frac{1}{2}\right)^2}\right|^2\right\}. \tag{4.41}$$

The middle sum is rather like the first, except that there is a -1 whenever n is odd. Taking the two sums together, it is just like we have performed the sum (trace) over all the integer momenta, but placed a projection onto even momenta, using the projector

$$P = \frac{1}{2}(1+(-1)^n). \tag{4.42}$$

In fact, an investigation will reveal that the third term can be written with a partner just like it save for an insertion of $(-1)^n$ also, but that latter sum vanishes identically. This all has a specific meaning which we will uncover shortly.

Notice that the partition function can be written in yet another nice way, this time as

$$Z_{R=\sqrt{\alpha'/2}} = \frac{1}{2}\left(|f_4^2(q)|^2 + |f_3^2(q)|^2 + |f_2^2(q)|^2\right), \tag{4.43}$$

where, for here and for future use, let us define

$$\begin{aligned}
f_1(q) &\equiv = q^{\frac{1}{24}}\prod_{n=1}^{\infty}(1-q^n) \equiv \eta(\tau)\\
f_2(q) &\equiv = \sqrt{2}q^{\frac{1}{24}}\prod_{n=1}^{\infty}(1+q^n)\\
f_3(q) &\equiv = q^{-\frac{1}{48}}\prod_{n=1}^{\infty}(1+q^{n-\frac{1}{2}})\\
f_4(q) &\equiv = q^{-\frac{1}{48}}\prod_{n=1}^{\infty}(1-q^{n-\frac{1}{2}}),
\end{aligned} \tag{4.44}$$

and note that

$$f_2\left(-\frac{1}{\tau}\right) = f_4(\tau); \quad f_3\left(-\frac{1}{\tau}\right) = f_3(\tau); \tag{4.45}$$
$$f_3(\tau+1) = f_4(\tau); \quad f_2(\tau+1) = f_2(\tau). \tag{4.46}$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

While the rewriting as (4.43) might not look like much at first glance, this is in fact the partition function of a single Dirac fermion in two dimensions: $Z(R=\sqrt{\alpha'/2})=Z_{\rm Dirac}$. We have arrived at the result that a boson (at a special radius) is in fact equivalent to a fermion. This is called 'bosonisation' or 'fermionisation', depending upon one's perspective. How can this possibly be true?

The action for a Dirac fermion, $\Psi=(\Psi_{\rm L},\Psi_{\rm R})^T$ (which has two components in two dimensions) is, in conformal gauge:

$$S_{\rm Dirac}=\frac{i}{2\pi}\int d^2\sigma\ \bar{\Psi}\gamma^a\partial_a\Psi=\frac{i}{\pi}\int d^2\sigma\ \bar{\Psi}_{\rm L}\bar{\partial}\Psi_{\rm L}-\frac{i}{\pi}\int d^2\sigma\ \bar{\Psi}_{\rm R}\partial\Psi_{\rm R}, \tag{4.47}$$

where we have used

$$\gamma^0=i\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \gamma^1=i\begin{pmatrix}0&-1\\1&0\end{pmatrix}.$$

Now, as a fermion goes around the cylinder $\sigma\to\sigma+2\pi$, there are two types of boundary condition it can have. It can be periodic, and hence have integer moding, in which case it is said to be in the 'Ramond' (R) sector. It can instead be antiperiodic, have half-integer moding, and is said to be in the 'Neveu–Schwarz' (NS) sector.

In fact, these two sectors in this theory map to the two sectors of allowed momenta in the bosonic theory: integer momenta to NS and half-integer to R. The various parts of the partition function can be picked out and identified in fermionic language. For example, the contribution:

$$\left|f_3^2(q)\right|^2\equiv\left|q^{-\frac{1}{24}}\right|^2\left|\prod_{n=1}^{\infty}(1+q^{n-\frac{1}{2}})^2\right|^2,$$

looks very fermionic, (recall insert 3.4 (p. 92)) and is in fact the trace over the contributions from the NS sector fermions as they go around the torus. It is squared because there are two components to the fermion, Ψ and $\bar{\Psi}$. We have the squared modulus beyond that since we have the contribution from the left and the right.

The $f_4(q)$ contribution on the other hand, arises from the NS sector with a $(-)^F$ inserted, where F counts the number of fermions at each level. The $f_2(q)$ contribution comes from the R sector, and there is a vanishing contribution from the R sector with $(-1)^F$ inserted. We see that that the projector

$$P=\frac{1}{2}(1+(-1)^F) \tag{4.48}$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

is the fermionic version of the projector (4.42) we identified previously. Notice that there is an extra factor of two in front of the R sector contribution due to the definition of f_2. This is because the R ground state is in fact degenerate. The modes Ψ_0 and $\bar{\Psi}_0$ define two ground states which map into one another. Denote the vacuum by $|s\rangle$, where s can take the values $\pm\frac{1}{2}$. Then

$$\begin{aligned} \Psi_0|-\tfrac{1}{2}\rangle &= 0; \qquad & \bar{\Psi}_0|+\tfrac{1}{2}\rangle &= 0; \\ \bar{\Psi}_0|-\tfrac{1}{2}\rangle &= |+\tfrac{1}{2}\rangle; \qquad & \Psi_0|+\tfrac{1}{2}\rangle &= |-\tfrac{1}{2}\rangle, \end{aligned} \tag{4.49}$$

and Ψ_0 and $\bar{\Psi}_0$ therefore form a representation of the two dimensional Clifford algebra. We will see this in more generality later on. In D dimensions there are $D/2$ components, and the degeneracy is $2^{D/2}$.

As a final check, we can see that the zero point energies work out nicely too. The mnemonic (2.80) gives us the zero point energy for a fermion in the NS sector as $-1/48$, we multiply this by two since there are two components and we see that that we recover the weight of the ground state in the partition function. For the Ramond sector, the zero point energy of a single fermion is $1/24$. After multiplying by two, we see that this is again correctly obtained in our partition function, since $-1/24+1/8=1/12$. It is awfully nice that the function $f_2^2(q)$ has the extra factor of $2q^{1/8}$, just for this purpose.

This partition function is again modular invariant, as can be checked using elementary properties of the f-functions (4.46): f_2 transforms into f_4 under the S transformation, while under T, f_4 transforms into f_3.

At the level of vertex operators, the correspondence between the bosons and the fermions is given by:

$$\begin{aligned} \Psi_{\rm L}(z) &= e^{i\beta X_{\rm L}^{25}(z)}; \quad & \bar{\Psi}_{\rm L}(z) &= e^{-i\beta X_{\rm L}^{25}(z)}; \\ \Psi_{\rm R}(\bar{z}) &= e^{i\beta X_{\rm R}^{25}(\bar{z})}; \quad & \bar{\Psi}_{\rm R}(\bar{z}) &= e^{-i\beta X_{\rm R}^{25}(\bar{z})}, \end{aligned} \tag{4.50}$$

where $\beta=\sqrt{2/\alpha'}$. This makes sense, for the exponential factors define fields single-valued under $X^{25}\to X^{25}+2\pi R$, at our special radius $R=\sqrt{\alpha'/2}$. We also have

$$\Psi_{\rm L}(z)\bar{\Psi}_{\rm L}(z)=\partial_z X^{25}; \quad \Psi_{\rm R}(\bar{z})\bar{\Psi}_{\rm R}(\bar{z})=\partial_{\bar{z}} X^{25}, \tag{4.51}$$

which shows how to combine two $(0,1/2)$ fields to make a $(0,1)$ field, with a similar structure on the left. Notice also that the symmetry $X^{25}\to -X^{25}$ swaps $\Psi_{\rm L(R)}$ and $\bar{\Psi}_{\rm L(R)}$, a symmetry of interest in the next subsection. We will return to this bosonisation/fermionisation relation in later sections, where it will be useful to write vertex operators in various ways in the supersymmetric theories.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

4.8 String theory on an orbifold

There is a rather large class of string vacua, called 'orbifolds'[23], with many applications in string theory. We ought to study them, as many of the basic structures which will occur in their definition appear in more complicated examples later on.

The circle S^1, parametrised by X^{25}, has the obvious $\mathbb{Z}_2$ symmetry $R_{25} : X^{25} \to -X^{25}$. This symmetry extends to the full spectrum of states and operators in the complete theory of the string propagating on the circle. Some states are even under R_{25}, while others are odd. Just as we saw before in the case of Ω, it makes sense to ask whether we can define another theory from this one by truncating the theory to the sector which is even. This would define string theory propagating on the 'orbifold' space $S^1/\mathbb{Z}_2$.

In defining this geometry, note that it is actually a line segment, where the endpoints of the line are actually 'fixed points' of the $\mathbb{Z}_2$ action. The point $X^{25} = 0$ is clearly such a point and the other is $X^{25} = \pi R \sim -\pi R$, where R is the radius of the original S^1. A picture of the orbifold space is given in figure 4.1. In order to check whether string theory on this space is sensible, we ought to compute the partition function for it. We can work this out by simply inserting the projector

$$P = \frac{1}{2}(1 + R_{25}), \tag{4.52}$$

which will have the desired effect of projecting out the R_{25}-odd parts of the circle spectrum. So we expect to see two pieces to the partition function: a part that is $\frac{1}{2}$ times Z_{circle}, and another part which is Z_{circle} with R_{25} inserted. Noting that the action of R_{25} is

$$R_{25} : \begin{cases} \alpha_n^{25} \to -\alpha_n^{25} \\ \tilde{\alpha}_n^{25} \to -\tilde{\alpha}_n^{25} \end{cases}, \tag{4.53}$$

the partition function is:

$$Z_{\text{orbifold}} = \frac{1}{2}\left[Z(R,\tau) + 2\left(|f_2(q)|^{-2} + |f_3(q)|^{-2} + |f_4(q)|^{-2}\right)\right]. \tag{4.54}$$

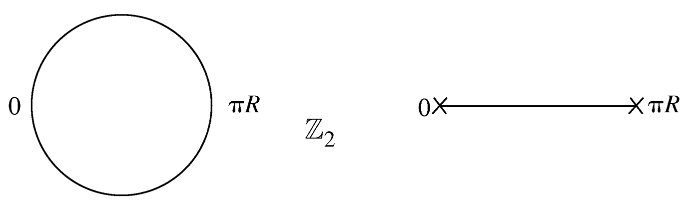


Fig. 4.1. A $\mathbb{Z}_2$ orbifold of a circle, giving a line segment with two fixed points.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

The f_2 part is what one gets if one works out the projected piece, but there are two extra terms. From where do they come? One way to see that those extra pieces must be there is to realise that the first two parts on their own cannot be modular invariant. The first part is of course already modular invariant on its own, while the second part transforms (4.46) into f_4 under the S transformation, so it has to be there too. Meanwhile, f_4 transforms into f_3 under the T-transformation, and so that must be there also, and so on.

While modular invariance is a requirement, as we saw, what is the physical meaning of these two extra partition functions? What sectors of the theory do they correspond to and how did we forget them?

The sectors we forgot are very stringy in origin, and arise in a similar fashion to the way we saw windings appear in earlier sections. There, the circle may be considered as a quotient of the real line $\mathbb{R}$ by a translation $X^{25} \to X^{25} + 2\pi R$. There, we saw that as we go around the string, $\sigma \to \sigma + 2\pi$, the embedding map $X^{25}(\sigma)$ is allowed to change by any amount of the lattice, $2\pi R w$. Here, the orbifold further imposes the equivalence $X^{25} \sim -X^{25}$, and therefore, as we go around the string, we ought to be allowed:

$$X^{25}(\sigma + 2\pi, \tau) = -X^{25}(\sigma, \tau) + 2\pi w R,$$

for which the solution to the Laplace equation is:

$$X^{25}(z, \bar{z}) = x^{25} + i\sqrt{\frac{\alpha'}{2}} \sum_{n=-\infty}^{\infty} \frac{1}{\left(n + \frac{1}{2}\right)} \left(\alpha^{25}_{n+\frac{1}{2}} z^{n+\frac{1}{2}} + \tilde{\alpha}^{25}_{n+\frac{1}{2}} \bar{z}^{n+\frac{1}{2}} \right), \tag{4.55}$$

with $x^{25} = 0$ or πR, no zero mode α_0^{25} (hence no momentum), and no winding: $w = 0$.

This is a configuration of the string allowed by our equations of motion and boundary conditions and therefore has to be included in the spectrum. We have two identical copies of these 'twisted sectors' corresponding to strings trapped at 0 and πR in spacetime. They are trapped, since x^{25} is fixed and there is no momentum.

Notice that in this sector, where the boson $X^{25}(w, \bar{w})$ is antiperiodic as one goes around the cylinder, there is a zero point energy of $1/16$ from the twisted sector: it is a weight $(1/16, 1/16)$ field, in terms of where it appears in the partition function.

Schematically therefore, the complete partition function ought to be

$$\begin{aligned} Z_{\text{orbifold}} &= \text{Tr}_{\text{untwisted}} \left(\frac{(1 + R_{25})}{2} q^{L_0 - \frac{1}{24}} \bar{q}^{\bar{L}_0 - \frac{1}{24}} \right) \\ &\quad + \text{Tr}_{\text{twisted}} \left(\frac{(1 + R_{25})}{2} q^{L_0 - \frac{1}{24}} \bar{q}^{\bar{L}_0 - \frac{1}{24}} \right) \end{aligned} \tag{4.56}$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

to ensure modular invariance, and indeed, this is precisely what we have in (4.54). The factor of two in front of the twisted sector contribution is because there are two identical twisted sectors, and we must sum over all sectors.

In fact, substituting in the expressions for the f-functions, one can discover the weight $(1/16, 1/16)$ twisted sector fields contributing to the vacuum of the twisted sector. This simply comes from the $q^{-1/48}$ factor in the definition of the $f_{3,4}$-functions. They appear inversely, and for example on the left, we have $1/48 = -c/24 + 1/16$, where $c = 1$.

Finally, notice that the contribution from the twisted sectors do not depend upon the radius R. This fits with the fact that the twisted sectors are trapped at the fixed points, and have no knowledge of the extent of the circle.

4.9 T-duality for open strings: D-branes

Let us now consider the $R \to 0$ limit of the open string spectrum. Open strings do not have a conserved winding around the periodic dimension and so they have no quantum number comparable to w, so something different must happen, as compared to the closed string case. In fact, it is more like field theory: when $R \to 0$ the states with non-zero internal momentum go to infinite mass, but there is no new continuum of states coming from winding. So we are left with a theory in one dimension fewer. A puzzle arises when one remembers that theories with open strings have closed strings as well, so that in the $R \to 0$ limit the closed strings live in D spacetime dimensions but the open strings only in $D - 1$.

This is perfectly fine, though, since the interior of the open string is indistinguishable from the closed string and so should still be vibrating in D dimensions. The distinguished part of the open string are the endpoints, and these are restricted to a $D - 1$ dimensional hyperplane.

This is worth seeing in more detail. Write the open string mode expansion as

$$\begin{aligned}
X^\mu(z, \bar{z}) &= X^\mu(z) + X^\mu(\bar{z}), \\
X^\mu(z) &= \frac{x^\mu}{2} + \frac{x'^\mu}{2} - i\alpha' p^\mu \ln z + i\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{n \neq 0} \frac{1}{n} \alpha_n^\mu z^{-n}, \\
X^\mu(\bar{z}) &= \frac{x^\mu}{2} - \frac{x'^\mu}{2} - i\alpha' p^\mu \ln \bar{z} + i\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{n \neq 0} \frac{1}{n} \alpha_n^\mu \bar{z}^{-n},
\end{aligned} \tag{4.57}$$

where x'^μ is an arbitrary number which cancels out when we make the usual open string coordinate. Imagine that we place X^{25} on a circle of

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

radius R. The T-dual coordinate is

$$\begin{aligned} X'^{25}(z,\bar{z}) &= X^{25}(z) - X^{25}(\bar{z}) \\ &= x'^{25} - i\alpha' p^{25} \ln\left(\frac{z}{\bar{z}}\right) + i(2\alpha')^{1/2} \sum_{n\neq 0} \frac{1}{n}\alpha_n^{25} e^{-in\tau} \sin n\sigma \\ &= x'^{25} + 2\alpha' p^{25}\sigma + i(2\alpha')^{1/2} \sum_{n\neq 0} \frac{1}{n}\alpha_n^{25} e^{-in\tau} \sin n\sigma \\ &= x'^{25} + 2\alpha' \frac{n}{R}\sigma + i(2\alpha')^{1/2} \sum_{n\neq 0} \frac{1}{n}\alpha_n^{25} e^{-in\tau} \sin n\sigma. \end{aligned} \tag{4.58}$$

Notice that there is no dependence on τ in the zero mode sector. This is where momentum usually comes from in the mode expansion, and so we have no momentum. In fact, since the oscillator terms vanish at the endpoints $\sigma = 0, \pi$, we see that *the endpoints do not move in the X'^{25} direction!* Instead of the usual Neumann boundary condition $\partial_n X \equiv \partial_\sigma X = 0$, we have $\partial_t X \equiv i\partial_\tau X = 0$. More precisely, we have the Dirichlet condition that the ends are at a fixed place:

$$X'^{25}(\pi) - X'^{25}(0) = \frac{2\pi\alpha' n}{R} = 2\pi n R'. \tag{4.59}$$

In other words, the values of the coordinate X'^{25} at the two ends are equal up to an integral multiple of the periodicity of the dual dimension, corresponding to a string that winds as in figure 4.2.

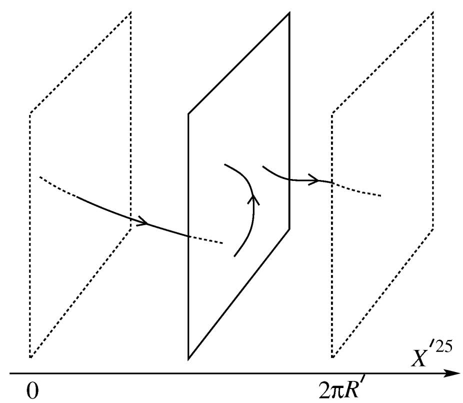


Fig. 4.2. Open strings with endpoints attached to a hyperplane. The dashed planes are periodically identified. The strings shown have winding numbers zero and one.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

This picture is consistent with the fact that under T-duality, the definition of the normal and tangential derivatives get exchanged:

$$
\begin{aligned}
\partial_n X^{25}(z,\bar{z}) &= \frac{\partial X^{25}(z)}{\partial z} + \frac{\partial X^{25}(\bar{z})}{\partial \bar{z}} = \partial_t X'^{25}(z,\bar{z}) \\
\partial_t X^{25}(z,\bar{z}) &= \frac{\partial X^{25}(z)}{\partial z} - \frac{\partial X^{25}(\bar{z})}{\partial \bar{z}} = \partial_n X'^{25}(z,\bar{z}).
\end{aligned}
\tag{4.60}
$$

Notice that this all pertains to just the direction which we T-dualised, X^{25}. So the ends are still free to move in the other 24 spatial dimensions, which constitutes a hyperplane called a 'D-brane'. There are 24 spatial directions, so we shall denote it a D24-brane.

4.9.1 Chan–Paton factors and Wilson lines

This picture becomes even more rich when we include Chan–Paton factors[25]. Consider the case of $U(N)$, the oriented open string. When we compactify the X^{25} direction, we can include a Wilson line

$$A_{25} = \text{diag}\{\theta_1, \theta_2, \ldots, \theta_N\}/2\pi R,$$

which generically breaks $U(N) \to U(1)^N$. (See insert 4.4 (p. 122) for a short discussion.) Locally this is pure gauge,

$$A_{25} = -i\Lambda^{-1}\partial_{25}\Lambda, \quad \Lambda = \text{diag}\{e^{iX^{25}\theta_1/2\pi R}, e^{iX^{25}\theta_2/2\pi R}, \ldots, e^{iX^{25}\theta_1/2\pi R}\}. \tag{4.61}$$

We can gauge A_{25} away, but since the gauge transformation is not periodic, the fields pick up a phase

$$\text{diag}\left\{e^{-i\theta_1}, e^{-i\theta_2}, \ldots, e^{-i\theta_N}\right\} \tag{4.62}$$

under $X^{25} \to X^{25} + 2\pi R$.

What is the effect in the dual theory? From the phase (4.62) the open string momenta are now fractional. As the momentum is dual to winding number, we conclude that the fields in the dual description have fractional winding number, i.e. their endpoints are no longer on the same hyperplane. Indeed, a string whose endpoints are in the state $|ij\rangle$ picks up a phase $e^{i(\theta_j-\theta_i)}$, so their momentum is $(2\pi n+\theta_j-\theta_i)/2\pi R$. Modifying the endpoint calculation (4.59) then gives

$$X'^{25}(\pi) - X'^{25}(0) = (2\pi n + \theta_j - \theta_i)R'. \tag{4.67}$$

In other words, up to an arbitrary additive constant, the endpoint in state i is at position

$$X'^{25} = \theta_i R' = 2\pi\alpha' A_{25,ii}. \tag{4.68}$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

Insert 4.4. Particles and Wilson lines

The following illustrates an interesting gauge configuration which arises when spacetime has the non-trivial topology of a circle (with coordinate X^{25}) of radius R. Consider the case of $U(1)$. Let us make the following choice of constant background gauge potential:

$$A_{25}(X^\mu) = -\frac{\theta}{2\pi R} = -i\Lambda^{-1}\frac{\partial\Lambda}{\partial X^{25}}, \tag{4.63}$$

where $\Lambda(X^{25}) = e^{-\frac{i\theta X^{25}}{2\pi R}}$. This is clearly pure gauge, but only locally. There still exists non-trivial physics. Form the gauge invariant quantity ('Wilson line'):

$$W_q = \exp\left(iq\oint dX^{25}A_{25}\right) = e^{-iq\theta}. \tag{4.64}$$

Where does this observable show up? Imagine a point particle of charge q under the $U(1)$. Its action can be written (see section 4.2) as:

$$S = \int d\tau\left\{\frac{1}{2}\dot{X}^\mu\dot{X}_\mu - iqA_\mu\dot{X}^\mu\right\} = \int d\tau\mathcal{L}. \tag{4.65}$$

The last term is just $-iq\int A = -iq\int A_\mu dx^\mu$, in the language of forms. This is the natural coupling of a world volume to an antisymmetric tensor, as we shall see.) Recall that in the path integral we are computing e^{-S}. So if the particle does a loop around X^{25} circle, it will pick up a phase factor of W_q. Notice: the conjugate momentum to X^μ is

$$\Pi^\mu = i\frac{\partial\mathcal{L}}{\partial\dot{X}^\mu} = i\dot{X}^\mu, \qquad \text{except for} \qquad \Pi^{25} = i\dot{X}^{25} - \frac{q\theta}{2\pi R} = \frac{n}{R},$$

where the last equality results from the fact that we are on a circle. Now we can of course gauge away A with the choice Λ^{-1}, but it will be the case that as we move around the circle, i.e. $X^{25} \to X^{25} + 2\pi R$, the particle (and all fields) of charge q will pick up a phase $e^{iq\theta}$. So the canonical momentum is shifted to:

$$p^{25} = \frac{n}{R} + \frac{q\theta}{2\pi R}. \tag{4.66}$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

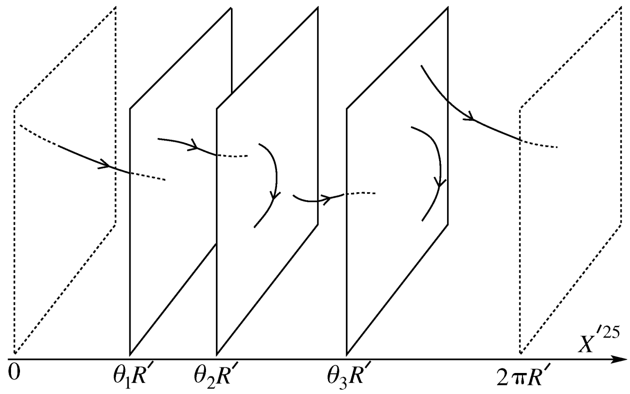


Fig. 4.3. Three D-branes at different positions, with various strings attached.

We have in general N hyperplanes at different positions as depicted in figure 4.3.

4.10 D-brane collective coordinates

Clearly, the whole picture goes through if several coordinates

$$X^m = \{X^{25}, X^{24}, \ldots, X^{p+1}\} \tag{4.69}$$

are periodic, and we rewrite the periodic dimensions in terms of the dual coordinates. The open string endpoints are then confined to N $(p+1)$-dimensional hyperplanes, the D$(p+1)$-branes. The Neumann conditions on the world-sheet, $\partial_n X^m(\sigma^1, \sigma^2) = 0$, have become Dirichlet conditions $\partial_t X'^m(\sigma^1, \sigma^2) = 0$ for the dual coordinates. In this terminology, the original 26 dimensional open string theory theory contains N D25-branes. A 25-brane fills space, so the string endpoint can be anywhere: it just corresponds to an ordinary Chan–Paton factor.

It is natural to expect that the hyperplane is dynamical rather than rigid[8]. For one thing, this theory still has gravity, and it is difficult to see how a perfectly rigid object could exist. Rather, we would expect that the hyperplanes can fluctuate in shape and position as dynamical objects. We can see this by looking at the massless spectrum of the theory, interpreted in the dual coordinates.

Taking for illustration the case where a single coordinate is dualised, consider the mass spectrum. The $D-1$ dimensional mass is

$$\begin{aligned} M^2 &= (p^{25})^2 + \frac{1}{\alpha'}(N-1) \\ &= \left(\frac{[2\pi n + (\theta_i - \theta_j)]R'}{2\pi\alpha'}\right)^2 + \frac{1}{\alpha'}(N-1). \end{aligned} \tag{4.70}$$

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

Note that $[2\pi n + (\theta_i - \theta_j)]R'$ is the minimum length of a string winding between hyperplanes i and j. Massless states arise generically only for non-winding (i.e. $n = 0$) open strings whose end points are on the same hyperplane, since the string tension contributes an energy to a stretched string. We have therefore the massless states (with their vertex operators):

$$\begin{aligned} &\alpha^{\mu}_{-1}|k;ii\rangle, \quad V = \partial_{\mathrm{t}} X^{\mu}, \\ &\alpha^{m}_{-1}|k;ii\rangle, \quad V = \partial_{\mathrm{t}} X^{25} = \partial_n X'^{25}. \end{aligned} \tag{4.71}$$

The first of these is a gauge field living on the D-brane, with $p+1$ components tangent to the hyperplane, $A^{\mu}(\xi^a)$, $\mu, a = 0, \ldots, p$. Here, $\xi^{\mu} = x^{\mu}$ are coordinates on the D-branes' world-volume. The second was the gauge field in the compact direction in the original theory. In the dual theory it becomes the transverse position of the D-brane (see equation (4.68)). From the point of view of the world-volume, it is a family of scalar fields, $\Phi^m(\xi^a)$, $(m = p+1, \ldots, D-1)$ living there.

We saw this in equation (4.68) for a Wilson line, which was a constant gauge potential. Now imagine that, as genuine scalar fields, the Φ^m vary as we move around on the world-volume of the D-brane. This therefore embeds the brane into a variable place in the transverse coordinates. This is simply describing a specific *shape* to the brane as it is embedded in spacetime. The $\Phi^m(\xi^a)$ are exactly analogous to the embedding coordinate map $X^{\mu}(\sigma, \tau)$ with which we described strings in the earlier sections.

The values of the gauge field backgrounds describe the shape of the branes as a soliton background, then. Meanwhile their quanta describe fluctuations of that background. This is the same phenomenon which we found for our description of spacetime in string theory. We started with strings in a flat background and discover that a massless closed string state corresponds to fluctuations of the geometry. Here we found first a flat hyperplane, and then discovered that a certain open string state corresponds to fluctuations of its shape. Remarkably, these open string states are simply gauge fields, and this is one of the reasons for the great success of D-branes. There are other branes in string theory (as we shall see) and they have other types of field theory describing their collective dynamics. D-branes are special, in that they have a beautiful description using gauge theory. Ultimately, we can use the long experience of working with gauge theories to teach us much about D-branes, and later, the geometry of D-branes and the string theories in which they live can teach us a lot about gauge theories. This is the basis of the dialogue between gauge theory and geometry which dominates the field at present.

It is interesting to look at the $U(N)$ symmetry breaking in the dual picture where the brane can move transverse to their world-volumes. When no D-branes coincide, there is just one massless vector each, or $U(1)^N$ in all, the generic unbroken group. If k D-branes coincide, there are new massless states because strings which are stretched between these branes can achieve vanishing length. Thus, there are k^2 vectors, forming the adjoint of a $U(k)$ gauge group[25, 26]. This coincident position corresponds to $\theta_1 = \theta_2 = \cdots = \theta_k$ for some subset of the original $\{\theta\}$, so in the original theory the Wilson line left a $U(k)$ subgroup unbroken. At the same time, there appears a set of k^2 massless scalars: the k positions are promoted to a matrix. This is not intuitive at first, but plays an important role in the dynamics of D-branes[26]. We will examine many consequences of this later in this book. Note that if all N branes are coincident, we recover the $U(N)$ gauge symmetry.

Although this picture seems quite odd, and will become more so in the unoriented theory, note that all we have done is to rewrite the original open string theory in terms of variables which are more natural in the limit $R \ll \sqrt{\alpha'}$. Various obscure features of the small-radius limit become clear in the T-dual picture.

Observe that, since T-duality interchanges Neumann and Dirichlet boundary conditions, a further T-duality in a direction tangent to a Dp-brane reduces it to a D$(p-1)$-brane, while a T-duality in a direction orthogonal turns it into a D$(p+1)$-brane.

4.11 T-duality for unoriented strings: orientifolds

The $R \to 0$ limit of an unoriented theory also leads to a new extended object. Recall that the effect of T-duality can also be understood as a one-sided parity transformation. For closed strings, the original coordinate is $X^m(z, \bar{z}) = X^m(z) + X^m(\bar{z})$. We have already discussed how to project string theory with these coordinates by Ω. The dual coordinate is $X'^m(z, \bar{z}) = X^m(z) - X^m(\bar{z})$. The action of world sheet parity reversal is to exchange $X^\mu(z)$ and $X^\mu(\bar{z})$. This gives for the dual coordinate:

$$X'^m(z, \bar{z}) \leftrightarrow -X'^m(\bar{z}, z). \tag{4.72}$$

This is the product of a world-sheet and a spacetime parity operation. In the unoriented theory, strings are invariant under the action of Ω, while in the dual coordinate the theory is invariant under the product of world-sheet parity and a spacetime parity. This generalisation of the usual unoriented theory is known as an 'orientifold', a term that mixes the term 'orbifold' with orientation reversal.

Imagine that we have separated the string wavefunction into its internal part and its dependence on the centre of mass, x^m. Furthermore, take the internal wavefunction to be an eigenstate of Ω. The projection then determines the string wavefunction at $-x^m$ to be the same as at x^m, up to a sign. The various components of the metric and antisymmetric tensor satisfy, for example,

$$\begin{aligned} G_{\mu\nu}(x^\mu, -x^m) &= G_{\mu\nu}(x^\mu, x^m), \quad B_{\mu\nu}(x^\mu, -x^m) = -B_{\mu\nu}(x^\mu, x^m), \\ G_{\mu n}(x^\mu, -x^m) &= -G_{\mu n}(x^\mu, x^m), \quad B_{\mu n}(x^\mu, -x^m) = B_{\mu n}(x^\mu, x^m), \\ G_{mn}(x^\mu, -x^m) &= G_{mn}(x^\mu, x^m), \quad B_{mn}(x^\mu, -x^m) = -B_{mn}(x^\mu, x^m). \end{aligned} \tag{4.73}$$

In other words, when we have k compact directions, the T-dual spacetime is the torus T^{25-k} moded by a $\mathbb{Z}_2$ reflection in the compact directions. So we are instructed to perform an orbifold construction, modified by the extra sign. In the case of a single periodic dimension, for example, the dual spacetime is the line segment $0 \le x^{25} \le \pi R'$. The reader should remind themselves of the orbifold construction in section 4.8. At the ends of the interval, there are fixed 'points', which are in fact spatially 24-dimensional planes. Looking at the projections (4.73) in this case, we see that on these fixed planes, the projection is just like we did for the Ω-projection of the 25+1 dimensional theory in section 2.6: the theory is unoriented there, and half the states are removed. These orientifold fixed planes are called 'O-planes' for short. For this case, we have two O24-planes. (For k directions we have 2^k $\mathrm{O}(25-k)$-planes arranged on the vertices of a hypercube.) In particular, we can usefully think of the original case of $k = 0$ as being on an O25-plane.

While the theory is unoriented on the O-plane, away from the orientifold fixed planes, the local physics is that of the *oriented* string theory. The projection relates the physics of a string at some point x^m to the string at the image point $-x^m$.

In string perturbation theory, orientifold planes are not dynamical. Unlike the case of D-branes, there are no string modes tied to the orientifold plane to represent fluctuations in its shape. Our heuristic argument in the previous subsection that gravitational fluctuations force a D-brane to move dynamically does not apply to the orientifold fixed plane. This is because the identifications (4.73) become *boundary conditions* at the fixed plane, such that the incident and reflected gravitational waves cancel. For the D-brane, the reflected wave is higher order in the string coupling.

The orientifold construction was discovered via T-duality[8] and independently from other approaches[27, 10]. One can of course consider more general orientifolds which are not simply T-duals of toroidal compactifications. The idea is simply to combine a group of discrete symmetries with Ω

such that the resulting group of operations (the 'orientifold group', G_Ω) is itself a symmetry of some string theory. One then has the right to ask what the nature of the projected theory obtained by dividing by G_Ω might be. This is a fruitful way of construction interesting and useful string vacua[28]. We shall have more to say about this later, since in superstring theory we shall find that O-planes, like D-branes , are sources of various closed string sector fields. Therefore there will be additional consistency conditions to be satisfied in constructing an orientifold, amounting to making sure that the field equations are satisfied.

So far our discussion of orientifolds was just for the closed string sector. Let us see how things are changed in the presence of open strings. In fact, the situation is similar. Again, let us focus for simplicity on a single compact dimension. Again there is one orientifold fixed plane at 0 and another at $\pi R'$. Introducing $SO(N)$ Chan–Paton factors, a Wilson line can be brought to the form

$$\mathrm{diag}\{\theta_1, -\theta_1, \theta_2, -\theta_2, \ldots, \theta_{N/2}, -\theta_{N/2}\}. \tag{4.74}$$

Thus in the dual picture there are $\frac{1}{2}N$ D-branes on the line segment $0 \le X'^{25} < \pi R'$, and $\frac{1}{2}N$ at their image points under the orientifold identification.

Strings can stretch between D-branes and their images, as shown in figure 4.4. The generic gauge group is $U(1)^{N/2}$, where all branes are separated. As in the oriented case, if m D-branes are coincident there is a $U(m)$ gauge group. However, now if the m D-branes in addition lie at one

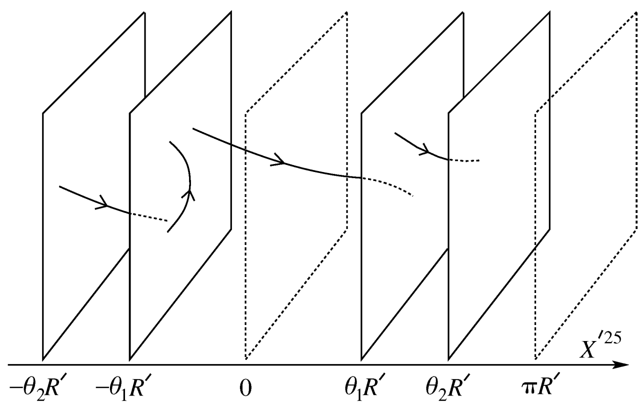


Fig. 4.4. Orientifold planes at 0 and $\pi R'$. There are D-branes at $\theta_1 R'$ and $\theta_2 R'$, and their images at $-\theta_1 R'$ and $-\theta_2 R'$. Ω acts on any string by a combination of a spacetime reflection through the planes and reversing the orientation arrow.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press

of the fixed planes, then strings stretching between one of these branes and one of the image branes also become massless and we have the right spectrum of additional states to fill out $SO(2m)$. The maximal $SO(N)$ is restored if all of the branes are coincident at a single orientifold plane. Note that this maximally symmetric case is asymmetric between the two fixed planes. Similar considerations apply to $USp(N)$. As we saw before, the difference between the appearance of the two groups is in a sign on the matrix M as it acts on the string wavefunction. Later, we shall see that this sign is correlated with the sign of the charge and tension of the orientifold plane.

We should emphasise that there are $\frac{1}{2}N$ dynamical D-branes but an N-valued Chan–Paton index. An interesting case is when $k+\frac{1}{2}$ D-branes lie on a fixed plane, which makes sense because the number $2k+1$ of indices is integer. A brane plus image can move away from the fixed plane, but the number of branes remaining is always half-integer. This anticipates a discussion which we shall have about fractional branes much later, in section 13.2, even outside the context of orientifolds.

https://doi.org/10.1017/9781009401371.005 Published online by Cambridge University Press