WIS-88/12/March-PH

MODULAR INVARIANCE, CHIRAL ANOMALIES AND CONTACT TERMS

David Kutasov

Department of Physics, Weizmann Institute of Science
Rehovot 76100, Israel

ABSTRACT

The chiral anomaly in heterotic strings with full and partial modular invariance in D=2n+2 dimensions is calculated.The boundary terms which were present in previous calculations are shown to be cancelled in the modular invariant case by contact terms, which can be obtained by an appropriate analytic continuation. The relation to the low energy field theory is explained. In theories with partial modular invariance, an expression for the anomaly is obtained and shown to be non zero in general.

§1 Introduction

Chiral anomalies in field theory are due to the impossibility of regularizing the divergent loop amplitudes in a way consistent with unitarity and with gauge invariance at the same time. The situation in open superstring theories[1] is similar: the amplitudes are divergent, and in the process of regularization a non zero contribution remains to give a finite answer (which is of course zero for SO(32)[2]).

In closed superstring theories[1,3] the situation is different. The relevant amplitudes are finite and no regularization is needed, which can explain the possibillity of an anomaly. It has been anticipated in [4] (see also [5]) that the anomaly should be given by a total derivative on moduli space, because it is equivalent to breakdown of BRST invariance, and this can only be due to a BRST operator passed around b(z) (the reparametrization ghost), generating an insertion of the energy momentum tensor T(z) (which is equivalent to a total derivative on moduli space). In several recent calculations [6-9], the (gauge and gravitational) anomaly was evaluated, and shown to yield indeed a total derivative.

It is by now well established that the anomaly is a total derivative on moduli space. What was not shown yet is that it vanishes. After all, many (modular invariant) amplitudes are given by such total derivatives and fail to vanish. One example is the BF^n coupling (D=2n+2) at zero momentum calculated in [12,13], which we'll discuss later, but there are many others. Even the k=0 four point function in D=10 heterotic (and type 2) strings, which is the one loop contribution to the effective action (contributing four particle couplings) is given by a total derivative w.r.t. τ. In these examples there is a finite contribution to the integral from the boundary at $\tau = i\infty$, which does not have to vanish because of modular invariance.

Indeed, when one looks at the results of [6-8], the following structure emerges: at zero momentum the anomaly integrates to a finite result (in the modular invariant case) proportional to TrJ^{n+2} (the fermion polygon contribution). For other values of the momenta the boundary term from infinity diverges because of momentum conservation (which is neccessary for modular invariance). There is no region of the kinematic invariants in which the total derivative integrates to zero. This is obviously a problem.

An additional problem, which exists regarding the anomaly is what happens in the modular non invariant case. It is well known that the finiteness of closed superstring theories relies heavily on modular invariance. When this is gone one has to make sense out of the resulting divergent amplitudes. A way to do this was suggested in [9], [10] (we will discuss it later). The conclusions reached in these papers are however opposite: while [9] obtained a zero anomaly but all other amplitudes zero too (and therefore no unitarity), [10] claimed that full modular invariance is not neccessary for finiteness and anomaly cancellation. We'll try to clarify this issue later.

Another interesting development was the clarification of the connection between full modular invariance and the possibility to cancel the polygon anomaly by a B term contribution. It was shown in [11] that in a modular invariant theory the traces of currents are related in precisely the right way for anomaly cancellation to be possible. In [12],[13] it was shown that a B coupling to photons and gravitons is indeed generated at one loop by the string, with the right form to cancel the anomaly. The coupling was shown to be related to the chiral gauged partition sum. We will see that the same functions govern the anomaly amplitude too, and show how the results of [12],[13] emerge from the study of the anomaly amplitude.

The purposes of this paper are :
1)To evaluate the anomaly taking care of boundary terms and show it vanishes for a modular invariant theory.
2)To relate the string result to the low energy field theory, and show the correspondence between various terms in the string amplitudes and low energy ones.
3)To calculate the anomaly for modular non invariant theories and show that it is in general non zero.

In section 2 we discuss the n+2 gauge boson amplitude in a modular invariant theory. It is shown that in order to calculate the amplitude one has to do one of two things: either split the amplitude into a sum over channels (different t orderings of the emmitted particles on the torus), each of which is evaluated at different momenta, thus breaking modular invariance, or add to it contact terms like the ones discussed in [14]. After this is done, one gets a finite result in which the two low energy contributions (of the fermion loop and B term) are manifest. When one passes to the anomaly, the two contributions cancel with no leftover boundary terms.

In section 3 the general case is considered. Under broad assumptions, the theory is still invariant under a subgroup of the modular group Γ of finite index. We can thus define all amplitudes on an extended fundamental region F' of Γ'. We explain in what sense the amplitudes defined on F' are divergent and discuss several ways of making sense out of them. A particular regularization, which gives finite amplitudes is discussed. We show how the anomaly calculated using this regularization gives a sum of the fermion loop and a B term similar to the one calculated in [12],[13].

Section 4 contains examples of calculations of the anomaly in the modular non invariant case. We show that the result is in general non zero and gives zero in the case when the massless spectrum is anomaly free.

§2 The anomaly: modular invariant case

In this section we will compute the anomaly amplitude for n+2 gauge bosons in the Cartan Subalgebra (C.S.A.) of the D=2n+2 gauge group. The gravitational case is similar and will not be discussed. The anomaly was considered previously in several papers (e.g. [6-9]), and therefore we will stress only the points which seem important for the understanding of the result.

The amplitude to be calculated is the n+2 point function on the torus in the periodic-periodic sector of the fermions ψ^{μ} (in the heterotic string; all the essential points go through for type 2 strings).

As explained in [7], [9], due to the presence of supermoduli the vertices that should be used in the PP sector are n+1 V_0's (namely gauge boson vertices in the picture with ghost charge 0) and one V_{-1} (which has ghost charge -1). To cancel the ghost charge of V_{-1} (the total ghost charge on the torus should be zero), an insertion of $\{Q_{BRST}, \xi\}$ (in the notation of [4]) is used, leading after evaluating the ghost correlator to an insertion of the superconformal generator

$$G(z) = \psi^\mu \partial_z X_\mu + T_F^{int}(z) \tag{2.1}$$

The amplitude should not depend on the position z of the insertion (2.1) , thus we can integrate over it without modifying the result (up to the volume of integration: $Im\tau$).

The vertices we use are

$$V_0(z,\bar{z}) = \xi^I_\mu \left[\partial_z X^\mu + ik\cdot\psi\psi^\mu\right] J^I(\bar{z}) e^{ik\cdot X(z,\bar{z})} \tag{2.2}$$

$$V_{-1}(z,\bar{z}) = \xi^I_\mu \psi^\mu(z) J^I(\bar{z}) e^{ik\cdot X} \tag{2.3}$$

Here $J^I(\bar{z})$ are the C.S.A. left moving Kac Moody currents (I=1..rank(left moving gauge group)); X^μ, ψ^μ are the space time fields.

There are various approaches to calculating the anomaly. Most use the simplifications which occur when one considers directly the divergence of the gauge field (namely $\xi = k$ for the appropriate gauge boson)[6,7,9]. For our purposes (which is to understand the origin of the different terms) it is more appropriate to calculate the full amplitude first, identify the terms corresponding to different field theory contributions and then take the divergence (as in [8]).

Thus, consider the amplitude for scattering n+2 gauge bosons in the PP sector

$$A = \int \frac{d^2\tau}{(Im\tau)^2} \prod_{i=1}^{n+2} \int d^2 z_i \int d^2 z (G(z) V_{-1}(z_1) V_0(z_2) \ldots V_0(z_{n+2})) \tag{2.4}$$

If we want to obtain a Bose symmetric expression for the amplitude, we have to *symmetrize* (2.4) . Alternatively, we could work with (2.4) as it stands, in which case conservation should hold independently for V_{-1} and V_0. This way of treating the problem is analogous to considering V_{-1} as representing an "axial current" and V_0 a "vector" one, the vector currents being automatically symmetrized.

The anomaly is related to the impossibility to conserve both currents simultaneously. In the form we use, the "vector current" V_0 is conserved automatically (without any requirements on the theory). Indeed, replacing ξ^i by k^i for $i \geq 2$ (see (2.2)), we get an expression with 2n+2 ψ's in (2.4) . Because we are in the PP sector, the ψ's have zero modes and thus have to contribute them to the amplitude to create an ϵ tensor. The remaining $\partial_z X^\mu$ from the insertion G(z) (see (2.1)) then turns the whole correlator into a total derivative of an elliptic function, whose z integral obviously vanishes. In principle one has to be careful about using equations of motion inside correlators, but here this is allowed, as we will verify explicitly later. In [9] it was shown that the V_0 currents are indeed not anomalous, however the contribution from V_{-1} was not considered.

We return to the calculation of (2.4) . Due to the zero modes of ψ^μ in the PP sector, 2n+2(=D) ψ's contribute only the zero modes to create an ϵ tensor. This still leaves two ψ fields uncontracted. The amplitude (2.4) can then be expressed (here and in general overall constants are ignored):

$$A = \int \frac{d^2\tau}{(Im\tau)^2} \prod_{i=1}^{n+2} \int d^2 z_i \int d^2 z B(\{z_i\}) B_J(\{\bar{z}_i\}) \tag{2.5}$$

where

$$\begin{aligned} B(\{z_i\}) =& \{ \sum_{i=2}^{n+2} \sum_{j=1} \epsilon \left[\mu \xi_1 \prod_{l=2}^{n+2} \prod_{l \neq i} k_l \xi_l \right] (k_i \cdot k_j \xi_{i\mu} - k_j \cdot \xi_i k_{i\mu}) S(z - z_i) \Delta(z - z_j) \\ &+ \sum_{j=2}^{n+2} \epsilon \left[\prod_{l=2}^{n+2} k_l \xi_l \right] \xi_1 \cdot k_j S(z - z_1) \Delta(z - z_j) \} \\ & \frac{1}{(Im\tau)^{n+1}} \exp \left[\sum_{i<j} \frac{1}{2} k_i \cdot k_j G_{ij}(z_{ij}) \right] \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} B_J(\{\bar{z}_i\}) =& \langle J(\bar{z}_1) .. J(\bar{z}_{n+2}) \rangle \\ =& \partial_{s_1} ... \partial_{s_{n+2}} \exp \left[\frac{1}{2\pi^2} \sum_{i<j} s_i \cdot s_j \tilde{\Delta}'(\bar{z}_{ij}) \right] A(\bar{q}, s, 0)|_{s=0} \end{aligned} \tag{2.7}$$

(2.6) is a sum of the different ways to choose the two ψ fields which are to be contracted; S is the $\langle \psi\psi \rangle$ correlator and Δ is the $\langle \partial X X \rangle$ one. B_J is obtained by using Wick's theorem and the definitions in Appendix A and [13]. $\tilde{\Delta}'$ is the oscillator contribution to $\langle J(z) J(w) \rangle$. $A(\bar{q}, s, 0)$ is defined and discussed in Appendix A.

It is worthwhile to discuss the structure of various correlators in (2.6), (2.7).

The boson correlator G(z) is given by

$$G(z) = \ln|\frac{\Theta_1(z,\tau)}{\Theta'(0,\tau)}| - \pi\frac{(Imz)^2}{Im\tau} \tag{2.8}$$

$$\Delta(z) = \frac{1}{z} - z\hat{G}_2(\tau) - \bar{z}\frac{\pi}{Im\tau} - \sum_{k=1}^{\infty} z^{2k+1}G_{2k+2}(\tau) \tag{2.9}$$

G_n are the Eisenstein functions and $\hat{G}_2(\tau) = G_2(\tau) - \frac{\pi}{Im\tau}$.

The fermion correlator S(z) in the PP sector should be an elliptic function in z, behave as $\frac{1}{z}$ for small z (this should be its only pole) and be holomorphic in z. All the requirements are incompatible by a general theorem. Thus, to define S, we have to relax one of them. The easiest requirement to drop is holomorphicity. If we are interested in a modular form of weight one, which is periodic in z with periods 1 and τ and behaves as $\frac{1}{z}$ for small z, $\Delta(z)$ in (2.9) is a possible candidate[7]. We will therefore use $S(z) = \Delta(z)$ in what follows (note that adding a constant to S in (2.6) does not change the amplitude). Another way to fix S is to use (global) 2D SUSY: start with the ∂XX correlator, rewrite the ∂X as a SUSY current encircling a ψ and deform the contour so that it surrounds the X, thus contributing another ψ. The result is that $S = \Delta$.

In the J correlator (2.7) the situation is simpler. The correlator $\langle J(z)J(0)\rangle$ is fully determined by :

1)The periodicity as $z \to z+1,\ \ z \to z+\tau$.

2)The behaviour as $z \to 0$ (which is a $\frac{1}{z^2}$ pole).

3)Modular covariance : it is a modular form of weight 2.

4)Meromorphicity in z.

5)Holomorphicity in τ.

The only function up to an additive modular form of weight 2 $F(\tau)$ satisfying 1) - 5) is the Weierstrass function

$$P(z) = \frac{1}{z^2} + \sum_{k=1}^{\infty}(2k+1)z^{2k}G_{2k+2}(\tau) \tag{2.10}$$

Thus we have

$$\langle J(z)J(0)\rangle = P(z) + F(\tau) \tag{2.11}$$

(2.11) differs from the space time boson correlator by $\hat{G}_2$:

$$\Delta'(z) = \langle \partial X(z)\partial X(0)\rangle = P(z) - \frac{1}{2}\hat{G}_2(\tau) \tag{2.12}$$

Comparing (2.11) and (2.12) we can conclude that the contribution of the oscillators (which is, by definition, the same in both cases) to (2.11) , (2.12) is

$$\tilde{\Delta}'(z) = P(z) - \frac{1}{2}G_2(\tau) \tag{2.13}$$

and the contribution of the zero modes is $\frac{1}{2}G_2 + F(\tau)$ for the compact boson and $\frac{1}{2}\frac{\pi}{Im\tau}$ in the space time case. $F(\tau)$ is determined by the particular theory through its particle content and the corresponding traces over the zero mode of J. It will have poles at cusps (if non zero). It is easy to see that F is non zero in supersymmetric theories.

Note that this choice of correlators (which arises naturally in operator formalism), makes the amplitude (2.5) - (2.7) almost modular invariant. The information about non invariance under the full modular group is concentrated in the behaviour of the partition sum $A(\bar{q},0,0)$ and its gauged generalization $A(\bar{q},s,0)$.

The integral over the position of the insertion z is easily performed by using the fact that S is equal to Δ. The result of the z integration is (see Appendix B):

$$\begin{aligned} A = \int d^2\tau \prod_{i=1}^{n+2} \int \frac{d^2 z_i}{Im\tau}\{ \sum_{i=2}^{n+2}\sum_{j=1} \epsilon \left[\mu\xi_1 \prod_{l=2\, l\neq i}^{n+2} k_l\xi_l\right] \\ (k_i\cdot k_j\xi_{i\mu} - k_j\cdot\xi_i k_{i\mu})\left[\partial_\tau + \frac{Imz_{ij}}{Im\tau}\partial_{z_i}\right] G(z_{ij}) \\ + \sum_{j=2}^{n+2} \epsilon\left[\prod_{l=2}^{n+2} k_l\xi_l\right]\xi_1\cdot k_j\left[\partial_\tau + \frac{Imz_{1j}}{Im\tau}\partial_{z_1}\right] G(z_{1j}\}\exp\left[\frac{1}{2}\sum_{i<j} k_i\cdot k_j G(z_{ij})\right] B_J \end{aligned} \tag{2.14}$$

(2.14) has to be evaluated cautiously as we will see later. It is a formal expression which diverges for all values of the momenta.

In the low energy limit, there are two contributions to (2.14) that will interest us:

1)The B terms. These are poles in the amplitude appearing at $k_i \cdot k_j = 0$. In (2.14) they appear as follows: when $z_i \to z_j$ there is a $\frac{1}{z_{ij}^2}$ coming from B_J, a factor $\frac{\bar{z}_{ij}}{z_{ij}}$ from $Im z_{ij} \partial_{z_i} G(z_{ij})$ and $|z_{ij}|^{\frac{1}{2} k_i \cdot k_j}$ from the exponential in (2.14). The behaviour for small z_{ij} is therefore $\frac{\xi_i \cdot k_j}{|z_{ij}|^{2-\frac{1}{2} k_i \cdot k_j}}$ which integrates to a pole $\frac{1}{k_i \cdot k_j}$ with the right coefficient $(\xi_i \cdot k_j)$. Since we contract two J's in order to get the B contribution, this term is proportional to the lower traces over the massless fields: $Tr J^n$ and lower. This is of course the expected behaviour of this term.

2)The fermion loop. This appears from the τ integration when $Im\tau \to \infty$. In this limit all the J's contribute their zero modes (see Appendices A, B). We rewrite $z_i = x_i + \tau y_i$, $(x_i, y_i \in [0,1])$ and expand (2.14) for $Im\tau \to \infty$. The y_i are related in a simple manner to the Feynmann parameters in the point field theory loop.

This can be seen by expanding G(z) (2.8) for $\tau \to i\infty$ as a function of x and y. The assymptotic expansion is

$$G(x,y) = \pi Im\tau y(1-y) \ln\left(1 - e^{2\pi i z}\right) \tag{2.15}$$

The ln part of (2.15) is the B pole contribution, which exists for all values of τ in particular $\tau \to i\infty$. To discuss the fermion loop we disregard it and concentrate on the y(1-y) part of (2.15) . Its contribution to (2.14) is an exponential of:

$$\frac{\pi}{2} Im\tau \sum_{i<j} k_i \cdot k_j y_{ij}(1 - y_{ij}) \tag{2.16}$$

We divide the y integration region into domains with a specific ordering of the y's (e.g. $y_1 < y_2 < \ldots$). This corresponds to discussing one planar diagram in the field theory limit at a time.

Defining the natural kinematic variables of the fermion loop

$$p_{lm} = \sum_{l=i<j}^{m} k_i \cdot k_j \tag{2.17}$$

one can show that the exponent in (2.16) is:

$$\frac{\pi}{2} Im\tau \sum_{i>j} k_i \cdot k_j y_{ij}(1 - y_{ij}) = \frac{\pi}{2} Im\tau \sum_{l<m} \alpha_l \alpha_m p_{lm} \tag{2.18}$$

where

$$\alpha_l = \begin{cases} y_{l+1} - y_l & 1 \le l \le n+1; \\ 1 - y_{n+2} + y_1 & l = n+2; \end{cases} \tag{2.19}$$

The α parameters are essentially the Feynmann parameters, since $0 \le \alpha_i \le 1$ and $\sum \alpha_i = 1$. The integrand in (2.14) is thus proportional to

$$\prod_{i=1}^{n+2} \int d\alpha_i \delta(\sum \alpha_i - 1) \exp\left[\frac{\pi}{2} Im\tau \sum_{l<m} \alpha_l \alpha_m p_{lm}\right] \tag{2.20}$$

(2.20) integrated over τ gives the usual behaviour of a loop diagram: massless thresholds. We remark that the imaginary part of the loop is correctly reproduced if the integration region over $Im\tau$ (the Schwinger proper time) reaches infinity. In order to reproduce also the formal field theory prescription for the real part one needs $Im\tau \in (0, \infty)$. Any other choice of the integration region in τ which does not include $Im\tau = 0$ corresponds to an effective ultraviolet cutoff.

To pass to the anomaly, we have to replace one of the ξ_i's by k_i. For $i \ne 1$ (2.14) then yields zero identically. In other words, the V_0's in (2.4) are conserved[9] as remarked above. When $k_1 = \xi_1$ is taken, however, a non zero result is obtained:

$$A = \epsilon \int d^2\tau \prod_{i=1}^{n+2} \int \frac{d^2 z_i}{Im\tau} \left\{ \left(\partial_\tau + \sum_i \frac{Im z_i}{Im\tau} \partial_{z_i} \right) \exp\left[\frac{1}{2} \sum_{i<j} k_i \cdot k_j G(z_{ij}) \right] \right\} B_J \tag{2.21}$$

where $\epsilon = \epsilon_{\mu_1\nu_1 \ldots \mu_{n+2}\nu_{n+2}} k_1^{\mu_1} \xi_1^{\nu_1} \ldots k_{n+2}^{\mu_{n+2}} \xi_{n+2}^{\nu_{n+2}}$ and the derivative part in brackets acts on the exponent only, as emphasized by the { }; the integral over τ can be restricted to a fundamental region of the modular subgroup under which the theory is invariant.

The next stage in the argument would be to show that (2.21) is a boundary integral over τ. This would have been so if the derivative (.....) in (2.21) acted on the whole expression. To see this, we change variables to $z_i = x_i + \tau y_i$ $(x_i, y_i) \in [0,1]$, which satisfy: $\frac{d^2 z_i}{Im\tau} = d^2 x_i$, and

$$\partial_\tau + \sum_i \frac{Im z_i}{Im\tau} \partial_{z_i} = \partial_\tau |_{x_i, y_i = const}$$

Since the region over which x and y are integrated is independent of τ, the ∂_τ can then be pulled out of the x integrals in (2.21) and then the anomaly is indeed a total derivative.

Therefore, in (2.21), we would like to commute the $\partial_\tau + \sum \frac{Im z_i}{Im\tau}\partial_{z_i}$ through $B_J(\bar{z}_i)$. There are two problems encountered in the process:

1)Boundary terms in z, related to passing the ∂_{z_i} through inverse powers of $\bar{z}_{ij}$ (as is well known, δ functions can occur in the process).

2)Boundary terms in the τ integration. Here too, naively $\partial_\tau B_J(\bar{z}_i) = 0$, however this assumes that B_J is analytic in the region over which τ is integrated. As will be explained later, this is not true in modular non invariant theories.

We start with problem 1) first; this is basically a kinematic problem: is it possible to find a region of the kinematic invariants, in which terms of the form

$$\exp\left[\frac{1}{2}\sum k_i \cdot k_j G(z_{ij})\right] \partial_{z_r} B_J(\bar{z}) \tag{2.22}$$

vanish? It seems at first that such a region does not exist, at least when the calculation is done on shell, since then $\sum_{i<j} k_i \cdot k_j = 0$ and terms like (2.22), which behave like $|z_{ij}|^{k_i \cdot k_j} \delta^2(z_{ij})$ diverge at least for one pair ij. This is however not so.

To explain the reason for this, it is helpfull to discuss a related problem of the on shell calculation for massless particles. Expressions like (2.21) are actually highly divergent in the $Im\tau \to \infty$ limit. The reason is that at $Im\tau \to \infty$, G can be expanded as in (2.15):

$$G(x, y) = \pi Im\tau(y - y^2)\ln(1 - e^{2\pi i z}) \tag{2.23}$$

By computing $\exp\left[\sum_{i<j} k_i \cdot k_j G(z_{ij})\right]$ it can be shown that for all values of the kinematic invariants (except, of course zero), there is some region of the y integration that contributes a $e^{+Im\tau}$ divergence to (2.14), (2.21) (and other amplitudes). The reason for this divergence is rather obvious: the string amplitude is a sum of all the field theory diagrams of the same order, in particular a sum over channels. The sum over channels is implemented in the string by summing over the t(=Im z) orderings of the vertices. For each channel, there is a kinematic region in which there are no pole and threshold singularities and then integral representations like (2.14), (2.21) are convergent (e.g. for the ordering $y_i < y_{i+1}$ $i = 1..n+1$ choosing $k_i \cdot k_{i+1} > 0$ is such a choice). The amplitudes are then continued analytically to the physical momenta. However for external particles with positive or zero mass square there is no common region of kinematical variables where all the orderings are convergent.

The same problem occurs, of course, already at tree level. There the amplitude is calculated in pieces, each of which converges at *different* values of the momenta. The resulting Γ functions are then glued together by analytic continuation.

A similar solution can be used to solve problem 1) in our case.Instead of calculating (2.21) , with the z_i integrated over the whole torus, we split it into a *sum* of terms in each of which the z_i's are integrated over in one particular order (of Imz_i). For each such term, a range of the momenta can be found, such that:
a)The $\tau = i\infty$ limit is finite.
b)There are no poles encountered when passing the z derivatives through B_J. These momenta are actually precisely the ones for which there are no singularities in the field theory diagrams in the appropriate channel. It is possible by requiring finiteness of (2.21) , (2.23) at $\tau = i\infty$ to calculate these ranges, but this will not be needed here.

There is an exception to the discussion presented here, namely the evaluation of (2.21) in D=4. Then n=1 and (2.21) is a three point function of massless particles. In this case, all kinematical invariants are fixed to be zero and a different treatment is neccessary. The formalism needed to evaluate such expressions was developed in [14], and it involves adding contact terms to the calculation or calculating slightly off shell, taking the limit in the end of the calculation (as in [15], [16]).

To summarize the above discussion, we saw that (2.21) has to be split into a sum of [illegible] (in each of which the z's are ordered), which are calculated at different momenta, [illegible] result is then continued analytically. When this is done, in each term separately the ∂_{z_i} can be passed through B_J without producing boundary terms (alternatively, one can use contact terms as shown in [14]).

For a modular invariant theory, the second problem (of $\partial_\tau B_j \neq 0$) does not occur, since the partition sum is analytic on the fundamental region. The only possible singularity is at $\tau = i\infty$, however this is taken care of by the kinematical choice we use (a more precise description of this will be presented in the next section).

Therefore the procedure to calculate (2.21) in practice is the following. We split the expression into a sum of convergent terms (different orderings):

$$A = \epsilon \sum_{orderings} \int_{B(F)} d\bar{\tau} \prod_{i=1}^{n+2} \int \frac{d^2 z_i}{Im\tau} \exp\left[\frac{1}{2}\sum_{i<j} k_i \cdot k_j G(z_{ij})\right] B_J \tag{2.24}$$

In (2.24) we already rewrote the integral (2.21) as an integral $d^2\tau$ of a total derivative and then as an integral $d\bar{\tau}$ over the boundary of the fundamental region denoted by B(F). Each of the terms in (2.24) is calculated at different values of the momenta. Note that the sum in (2.24) is not modular invariant anymore, because of this difference of the momenta. The contribution to (2.24) from the boundary at $\tau = i\infty$ is made to vanish by the kinematic choice for each of the terms separately (this is in fact how the different momenta are chosen). $\tau \to \tau + 1$ is still a symmetry *in each of the terms separately*, since this transformation does not act on the z's. Therefore the contributions from the two lines $Re\tau = \pm 1$ cancel too. We are left with

$$A = \epsilon \sum_{orderings} \int_{|\tau|=1} d\bar{\tau} \prod_{i=1}^{n+2} \int \frac{d^2 z_i}{Im\tau} \exp\left[\frac{1}{2}\sum_{i<j} k_i \cdot k_j G(z_{ij})\right] B_J \tag{2.25}$$

(2.25) should be calculated in each piece of the sum at the appropriate momentum. If we are interested in the anomaly at a certain value of the momenta, then the different pieces have to be continued to this value analytically from the region where the integral representations converge. But (2.25) is a sum of finite integrals which are well behaved as a function of the kinematic invariants. Thus we can now safely take the momenta to some common value (e.g. k=0) and evaluate the anomaly (2.25) there. Of course we can take non zero k too. (2.25) vanishes then also because of our ability to calculate all the terms at the same k. If all the terms in (2.25) are calculated at the same momenta, the expression is modular invariant again, and thus vanishes. The value at k=0 is the actual anomaly. Contributions at $k \neq 0$, which correspond to higher powers in α', can be removed (in field theory) by adding local counterterms.

After taking k=0, the integrals are easy: the sum over orders gives as usual an integral over the z's over the full torus, and thus can be evaluated using the results of Appendix

B. The result for the anomaly is finally

$$A = \epsilon \int_{|\tau|=1} d\bar{\tau}\partial_{s_1}..\partial_{s_{n+2}}\bar{A}(\bar{q},s)|_{s=0} \tag{2.26}$$

(2.26) vanishes obviously since it is an integral of a weight two modular form, which is *finite and well defined* on the line of integration (this is an important point as we will see later). Thus (2.26) goes to minus itself under $\tau \to -\frac{1}{\tau}$ and therefore vanishes.

In the modular invariant case the anomaly vanishes. We can trace back this vanishing (at least to lowest order in α') to the cancellation of the two field theory terms discussed in (2.15) - (2.20) . The point is that the limit $\alpha' \to 0$ in the expression for the amplitude (2.14) is equivalent to the limit $\tau \to i\infty$ which was considered before. Indeed, consider

$$\chi(c,w) = e^{G(z,\tau)} = \exp\left[\frac{(\ln|c|)^2}{2\ln|w|}\right] \cdot |\frac{1-c}{c^{\frac{1}{2}}}| \cdot |\prod_{m=1}^{\infty} \frac{(1-w^m c)(1-\frac{w^m}{c})}{(1-w^m)^2}| \tag{2.27}$$

where $c = e^{2\pi i z}$, $w = e^{2\pi i\tau}$. The amplitude (2.14) contains a product of terms of the form

$$\chi(c,w)^{\alpha' k_i \cdot k_j}$$

In the limit $\alpha' \to 0$ the first term in (2.27) contributes since it is large in the $\tau \to i\infty$ limit. The second has to be kept since it diverges at $z \to 0$ (for all α'). The third term (the infinite product) on the other hand is 1 at z=0 and finite for all z in the integration region (which is $|Re z| \leq \frac{1}{2}$, $|Im z| \leq \frac{1}{2} Im\tau$ appropriately tilted). Therefore, when $\alpha' \to 0$ we can use

$$\chi(c,w) = \exp\left[\frac{(\ln|c|)^2}{2\ln|w|}\right] \cdot |\frac{1-c}{c^{\frac{1}{2}}}| \tag{2.28}$$

changing from χ to G and from c,w to z, τ, we see that for small α', G(z) is given approximately by (2.15) , which is precisely the $\tau \to i\infty$ limit. We have already established that the two terms in (2.28) are responsible for the fermion loop and B terms.Therefore, the anomaly can also be calculated by calculating the sum of the two terms. Dividing the region of integration over z_i into the whole region minus a small disk around $z_i = z_j$ (for each j) and the disk, we can ignore the contribution of the pole in (2.28) to the amplitude integrated over the former. The amplitude that remains is just the fermion loop contribution considered above. It can be calculated from (2.14) and (2.20) by using field theory techniques. The integration over the vicinity of the pole $z_i = z_j$ gives a pole at $(k_i + k_j)^2 = 0$. The B contribution is obtained by factorizing (2.14) on that pole (this is the only contribution that persists, since all other contributions vanish when α' is taken to zero (see (2.28)). After factorizing on the pole, the calculation reduces to evaluating the $BTrF^n$ coupling considered in ref. [13]. The relevant integrals appear in Appendix B. The coupling of B to n photons is *not* (only) an additional local coupling at low energy. Indeed, if we factorize the scattering amplitude (2.14) on the B pole, we get a sum of terms proportional to

$$\frac{\xi_n \cdot k_m}{(k_n + k_m)^2} \int \frac{d^2\tau}{(Im\tau)^2} \exp\left[\frac{1}{2}\sum_{i<j} k_i \cdot k_j G(z_{ij})\right] B'_j(\bar{z}) \tag{2.29}$$

where now i,j run over the B and n gauge bosons and

$$B'_j(\bar{z}) = \langle J(\bar{z}_1)..J(\bar{z}_n)\rangle \tag{2.30}$$

(2.29) was expanded in [13] in the invariants. Strictly speaking, this is not the whole amplitude: as we can see from the preceeding analysis, the amplitude has cuts as $k \to 0$. Using (2.20) , it is easy to show that (2.29) is proportional to

$$\frac{\xi_n \cdot k_m}{(k_n + k_m)^2} \int d\alpha_1..d\alpha_{n+1}\delta(\Sigma\alpha_i - 1)\left[f(\alpha, k)\ln f(\alpha, k) - f(\alpha, k) + const\right] \tag{2.31}$$

where

$$f(\alpha, k) = \sum_{i<j} \alpha_i\alpha_j p_{ij}$$

(see (2.16) - (2.20)). *Since the amplitude under consideration is one* appearing in the anomaly, we are assured that the k=0 limit of (2.31) is finite (as can be easily verified). The field theory loops giving the behaviour (2.31) vanish at zero momentum and the only remaining contribution in the string amplitude is the local coupling.

When the anomaly is calculated, the B contribution is (from (2.31)) a term proportional to $\frac{k^2}{k^2}$. This is similar to the situation encountered in [15] , [16] in a different context. We also see that the imaginary parts cancel when we sum over all n,m: the B contribution to the anomaly is a total derivative by itself. In the field theory, the imaginary part of the B amplitude is cancelled by a contribution of another loop. The coupling of B to the appropriate fermion current J, is through a term (dB-AF)J in the low energy action. Thus, in addition to the BF^n fermion loop which contributes to the n+2 gauge boson coupling, there is a loop arising from the four particle vertex AFJ. In the string, this coupling is of course absent, and its contribution arises from massive states running in the loop as usual.

§3 The modular non invariant case

In a theory in which full modular invariance is absent, the discussion of section 2 has to be reexamined. (2.21) is in this case defined with the τ integration extending over a fundamental region for Γ', a modular subgroup of finite index. In order to rewrite (2.21) as a boundary integral we have to reconsider problem 2, namely what does the τ derivative do to B_J. From (2.7) we see that B_J depends on $\bar{\tau}$ only. However, the partition sum $A(\bar{q}, 0, 0)$ and thus the gauged partition sum $A(\bar{q}, s, 0)$ which appears in (2.7) (and in fact all modular functions), have poles on the boundary of moduli space for all subgroups of Γ. For example, there is always in the canonical fundamental region of such a subgroup a point with $Im\tau = 0$. At such a point, the lattice sum described in Appendix A diverges since there is no exponential dumping factor for large w_L^2. Examples of this will be given below. Therefore, in general, one cannot pass the τ derivative through B_J without taking care of boundary terms.

We regulate the integral by excluding from the integration region (the fundamental domain) neighbourhoods of the poles limited by arcs of radius r. After calculating the integral, r is taken to zero. This is a well known procedure in the case of modular forms with poles at cusps. For such forms, boundary integrals are performed over the trajectory in Fig. 1: Of course, when this is the trajectory, the integral of a modular form of weight 2 from ρ to ρ' is no more zero, although $\tau \to -\frac{1}{\tau}$ is still a symmetry. The resulting boundary terms are for example the reason that $\frac{G_6}{G_4}|_{\bar{q}^0 \ term} \neq 0$ (see Appendix A). In the modular invariant case all this means that a theory has to be both modular invariant and to have $A(\bar{q}, 0, 0)$ (chiral partition sum), which is regular in the fundamental region (except perhaps infinity). Since this is usually the case, this is not interesting.

In theories with partial modular invariance, on the other hand, these boundary terms are an inevitable consequence of the structure of the fundamental region. It is these boundary terms that cause the anomaly.

To summarize, it is possible to set $\partial_\tau B_j = 0$ only if small areas around poles of B_J are excluded. In the end of the calculation the size of these areas is to be taken to zero.

After the discussion of the boundary terms in τ in the general case, we may now continue the calculation from (2.21). We consider a theory, which is not modular invariant, but is invariant under a subgroup Γ' of the modular group Γ of finite index. This is usually the case in string models which are based on a finite number of Kac Moody representations. We rewrite (2.21) as a boundary integral

$$A = \epsilon \int_{F'} d^2\tau \partial_\tau \left\{ \prod_{i=1}^{n+2} \int \frac{d^2 z_i}{Im\tau} \exp\left[\frac{1}{2} \sum_{i<j} k_i \cdot k_j G(z_{ij}) \right] B_J(\bar{z}_i) \right\} \tag{3.1}$$

where the z integration is divided into regions such that for the appropriate choice of the kinematical invariants for each region the integral converges. The domain of integration over τ is as specified above.

Since the τ integral is independent of the fundamental region chosen, we will choose a specific one for definiteness, namely a region symmetric around the $Im\tau$ axis and extending to infinity (of course, this doesn't define F' uniquely). Such a region is bounded by the lines $|Re\tau| = \frac{m}{2}$ $(m = 1, 2...)$ and some curve defining a UV cutoff, which we will denote by γ (e.g. $|\tau| = 1$ for a fully modular invariant theory; other examples will be given below.)

First, irrespectively of modular invariance the $Im\tau = \infty$ boundary term vanishes. The reason is that in each of the terms (representing one channel), the $k \cdot k$'s are arranged so that (see (2.23)):

$$\sum k_i \cdot k_j G(z_{ij}) = -Im\tau H(y) \tag{3.2}$$

where $H(y) \geq 0$. B_J behaves simply as $\tau \to i\infty$ (see (2.7) and appendix B):

$$B_J(\bar{z}_i) = \partial_{s_1} .. \partial_{s_{n+2}} \exp\left\{ \frac{1}{2\pi^2} \frac{1}{Im\tau} \left[f(s)\delta^2(x_{ij}) + .. \right] \right\} \bar{A}(\bar{q}, s) \tag{3.3}$$

The exponent in (3.3) contains δ functions which do not contribute as explained before, because of the kinematical region chosen and because of the $\frac{1}{Im\tau}$ damping. Thus B_J behaves for large $Im\tau$ as $\partial_{s_1}..\partial_{s_{n+2}}\tilde{A}|_{s=0}$, which tends to a constant (TrJ^{n+2} over the massless fermions). The exponential in (3.1) then causes the full integral to vanish because of (3.2) .

We still have to evaluate the contribution of the "finite boundaries" $Re\tau = \pm\frac{m}{2}$ and the UV region. It is obvious that by separating the z integrals into regions of definite order and calculating them at different k's we have spoiled modular invariance. However $\tau \rightarrow \tau + m$ is still a symmetry, since under this operation the order of the z's is not changed. This is similar to what happens in the modular invariant case. Thus for each of the terms in the "sum over channels", the above boundaries cancel each other.

The only surviving contribution is the one from the UV cutoff γ (e.g. the integral from ρ to ρ' in fig. 1). The result for the anomaly is given by a generalization of (2.25) (now we explicitly write the sum over channels):

$$A = \sum_{orderings} \epsilon \int_\gamma d\bar{\tau} \prod_{i=1}^{n+2} \int \frac{d^2 z_i}{Im\tau} \exp\left[\frac{1}{2}\sum_{i<j} k_i \cdot k_j G(z_{ij})\right] B_J \qquad (3.4)$$

(3.4) is the formal result for the anomaly. As we can see, it is given by the contribution from the UV cutoff introduced by the string only and the different terms in B_J (2.7) give terms proportional to various traces of J. The TrJ^{n+2} term is the fermion loop contribution, while the lower traces represent the B coupling , as we will soon see.

The first thing that we would like to show is that (3.4) diverges. Note that the contribution is non zero only from the vicinity of the poles (in τ) of the integrand. The reason is that the boundary goes to minus itself under some modular transformation S' in the subgroup Γ', under which the partition sum is invariant. Therefore, for all points on the curve γ on which the integrand in (3.4) is finite, we can calculate all the terms in the sum over orderings at the same value of the momenta (as in section 2, because the integrals are finite), and then use the transformation S' to cancel the integrals over the different parts of γ. This argument (as shown in Appendix A for the case of modular forms with poles at cusps) does not work at the points of γ where there are poles in τ. In the example considered there, these poles gave rise to finite contributions obtained from surrounding them. In our case, this gives rise to infinities.

The simplest way to see this is the following. Consider a transformation U', which takes the pole $a \in B(F')$ to the fundamental region F of Γ and maps it to $\tau = i\infty$. U' is obviously in Γ/Γ'. The reason U' exists is that we can certainly find a $U' \in \Gamma/\Gamma'$, which takes a into F. Because the amplitude is potentially divergent at a, it will have to diverge at the image of a in F. This image thus has to be $\tau = i\infty$ since this is the only point in F where the amplitude is potentially divergent.

In the integration over γ in (3.4) we deform the curve around a, to avoid the pole. One can easily convince oneself that a suitable "cutoff" curve is one which is transformed by U' to the line

$$\gamma' = \left\{\tau; Im\tau = b\ ,\ |Re\tau| \leq \frac{1}{2}\right\} \tag{3.5}$$

b is then the UV cutoff introduced near a. After calculating (3.4) on γ defined near a by $(U')^{-1}\gamma'$, we have to take b to infinity (remove the cutoff).

After transforming the integral (3.4) near a to γ'_a (by defining $\tau = (U')^{-1}\tau'$), we have an integral of the form

$$A = \sum_{(orderings)} \sum_{(poles\ a)} \epsilon \int_{\gamma'_a} d\bar{\tau} \prod_{i=1}^{n+2} \int \frac{d^2 z_i}{Im\tau} \exp\left[\frac{1}{2}\sum_{i<j} k_i \cdot k_j G(z_{ij})\right] B_J \tag{3.6}$$

Consider a specific term in the sum (3.6). Suppose that the momenta appearing in it are already fixed by the requirement that there be no pole at $\tau = i\infty$ in the original region $(B(F'))$, so that we *do not* allow different choices of the momenta for different values of τ, and therefore the invariants are all fixed at this stage. If $U'(\tau) = \frac{a\tau+b}{c\tau+d}$, then in passing from (3.4) to (3.6), z is also transformed by

$$z' = \frac{z}{c\tau + d} \tag{3.7}$$

Obviously, if U' takes a point near zero to infinity, $c \neq 0$. (3.7) implies that under U' the z's are permuted so that if we started with $y_1 < y_2 < \ldots$, the z' are arranged in some other way (which is not even a different ordering by the value of y'_i). As we saw in the analysis of section 2, this means that the correspondence between different terms in (3.6) and channels of field theory is ruined (the α_i of (2.19) are no more in $[0,1]$)and therefore (3.6) diverges (In the same way that a Feynmann parameter representation of a loop amplitude in field theory diverges if we choose momenta in which cuts appear).

The next question is how does one make sense out of the divergent integral (3.6). The first thing that is plausible to try along the lines of section 2 is to redivide the integration region over z *and* τ to make (3.6) finite again. A way to do this was essentially proposed (of course without using this language) in [9], [10]. There, the momenta were chosen differently around each pole $a \in \gamma$ and the division of the z integration in (3.6) was made so that after applying U', the z' of (3.7) were arranged in the planar fashion described above. If such a division is used, the poles on γ do not contribute to the anomaly, since they are eliminated by the same mechanism as the contribution at $\tau = i\infty$ (see Section 2). Then, by the arguments presented above, the anomaly amplitude vanishes. In [9] it was shown (and we'll verify it in the next section), that this doesn't contradict the fact that the low energy field theory *is* anomalous. The point is that using this regularization, the correspondence between the string and low energy field theory amplitudes is lost. Thus the theory is not unitary to one loop, but it has nothing to do with an anomaly. The illness appears already in four point functions in light cone gauge.

The reason is also obvious in our language. By using the peculiar kinematic choice, this regularization treats the cuts near $\tau = 0$ as additional fermion loops. The "particles" that contribute to it can be obtained by transforming $A(\bar{q},0,0)$ to $i\infty$ by using the transformation U':

$$A_{eff}(\bar{q},0,0) = A[(U')^{-1}\bar{q},0,0)]$$

Because of the transformation properties of $A(\bar{q},0,0)$, these additional loops contain different circulating particles from those present at tree level(some of the new particles may even have negative norm[9]). These particles obviously break unitarity.

In such a situation, the anomaly is of course a totally irrelevant quantity. We would like to suggest a different regularization scheme, namely a different way of treating the poles of $A(\bar{q},0,0)$ on γ, which will avoid the above problem of unitarity, so that the anomaly can be considered in a meaningfull context.

A way to do this (motivated by the operator formalism) is to use (3.6) with the *same* Imz ordering and corresponding $k \cdot k$ choice all over the fundamental region over which τ is integrated. We then have to introduce the cutoff b defined above (3.5) . The question now is how does one treat the $b \to \infty$ limit. One way to do this is to take the cutoff b to infinity and keep only the terms of (3.6) which are independent of the cutoff. These are finite as we will see. This is the regularization that we will use.

In the limit $b \to \infty$ we can expand G(z) in (3.6) using (2.15) with $Im\tau = b$. The exponential factor in (3.6) then contains combinations like $bk_i \cdot k_j f(y_{ij})$. The only term independent of b is obviously the one obtained by taking b(or k)=0. Thus the right prescription in (3.6) is to set $k \cdot k$ to zero. Integrating $B_J(\bar{z}_i)$ over the whole torus (summing all the terms in (3.6)) (see Appendix B and [13] for details), we get the final result for the anomaly:

$$A = \int_\gamma d\bar{\tau} \partial_{s_1} .. \partial_{s_{n+2}} \bar{A}|_{s=0} \tag{3.8}$$

where γ is the curve connecting the $Re\tau = \pm \frac{m}{2}$ lines, avoiding the "dangerous" points along its trajectory where $\bar{A}$ diverges in the manner explained above . (3.8) is nothing but a straightforward generalization of (2.25) .

It is perhaps worthwhile to pause before discussing how one calculates (3.8) and showing examples, and discuss why this expression is natural and in some sense trivial. First note, (see Appendix A) that $\partial_{s_1} .. \partial_{s_{n+2}} \bar{A}$ is a sum of terms, each of which contains a different trace of J, from TrJ^{n+2} and lower. We will now show that in general the different terms arrange into a fermion loop contribution and a B term. Consider the TrJ^{n+2} term in (3.8) . It is given by

$$A^f = \int_\gamma d\bar{\tau} \partial_{s_1} .. \partial_{s_{n+2}} A(\bar{q}, s)|_{s=0} \tag{3.9}$$

The integrand is a holomorphic function which is analytic on γ + the other boundaries of the fundamental region (remember that the points where it is divergent are excluded by γ). Therefore the integral over the boundary of the fundamenatl region of (3.9) is zero (by Cauchy). Since the $Re\tau = \pm \frac{m}{2}$ lines cancel ($\tau \to \tau + m$ is a symmetry of (3.9) ; see Appendix A), the integral over γ is equal to minus the integral over $Im\tau = \infty$. If γ

extends from $Re\tau = -\frac{m}{2}$ to $+\frac{m}{2}$, then

$$A^f = m\partial_{s_1}..\partial_{s_{n+2}}A|_{s=0;q^0 \ term} = mTrJ^{n+2} \tag{3.10}$$

The trace in (3.10) is over the massless fermions. Up to the normalization m in (3.10), which is overall, this expression is just the anomaly contribution from the regulator of a fermion loop in field theory. To see that the remainder of (3.9) is a B term, we rewrite it as

$$A^b = A - A^f = \int_{B(F')} d\bar{\tau}\partial_{s_1}..\partial_{s_{n+2}}\bar{A}|_{s=0} \tag{3.11}$$

where $B(F')$ is the whole boundary of the fundamental region F'. It can be rewritten (up to constants) as:

$$A^b \propto \int_{F'} d^2\tau\partial_\tau\partial_{s_1}..\partial_{s_{n+2}}\bar{A}|_{s=0} \propto \int_{F'} \frac{d^2\tau}{(Im\tau)^2}\partial_{s_1}..\partial_{s_{n+2}}\bar{A}|_{s=0} \tag{3.12}$$

(3.12) is a simple generalization of the result of [13] for the B coupling to n gauge bosons at one loop (at zero momentum). The low energy field theory says nothing about the value of this term. In fact, there it is put in by hand to cancel the anomaly. As shown in [13], in modular invariant theories this term is generated by the string loop and is precisely equal to $-TrJ^{n+2}$ due to trace identities which follow from modular invariance, and thus cancels the polygon anomaly. Modular non invariant theories are in general anomalous, and thus the B term is not well defined in field theory. (3.12) is the $BTrF^n$ coupling which is generated by the string in such theories. It indeed is not (in general) enough to cancel the anomaly. In some cases this coupling vanishes[9].

Another trivial point we should make is that (3.8) vanishes in the modular invariant case: γ is then the $|\tau| = 1$ line (fig. 1), and since there are no singularities on this curve at ρ, i, ρ', the usual argument holds (under $\tau \to -\frac{1}{\tau}$ the line goes into minus itself, while the integrand is invariant). As we will see in the next section, in modular non invariant cases (3.8) need not, and in general does not vanish.

Before passing to examples, we would like to explain how one actually calculates (3.8) for a modular non invariant theory. One way to do this is to use (3.10) and (3.12). The

B term can be calculated by projecting the integral over F' into F :

$$\int_{F'} \frac{d^2\tau}{(Im\tau)^2} G(\tau) = \int_{F} \frac{d^2\tau}{(Im\tau)^2} \sum_{i=1}^{\mu} G(\tau)|W_i \tag{3.13}$$

where W_i are coset representatives of Γ/Γ' and μ is the index of Γ' in Γ. The fermion contribution is taken from (3.9) , (3.10) .

It can be shown that this way of calculating the B term actually corresponds to our previous prescription of treating the singularity, namely excluding it from the integration region and then taking the exclusion radius to zero .

§4 Examples of anomaly calculation

To demonstrate the general ideas discussed above, we consider several theories invariant under a certain subgroup of Γ. Consider theories of heterotic strings in D=10, with the internal coordinates lying on an odd self dual lattice[10]. Such theories are obviously invariant under $\tau \to -\frac{1}{\tau}$, but are not invariant under $\tau \to \tau + 1$, but only under $\tau \to \tau + 2$. The subgroup Γ' generated by these two transformations has index $\mu = 3$ in Γ, and a fundamental region (F') can be chosen e.g. as in Fig. 2a or 2b.

Both regions contain points with $Im\tau = 0$ where the divergences discussed above occur. Therefore in boundary integrals, we exclude these corners of F'. The anomaly (3.8) can then be calculated by the prescription (3.13) , which using F' of Fig. 2a gives just

$$G(\tau) + G(\tau+1) + \tau^{-n} G(\frac{\tau - 1}{\tau}) \tag{4.1}$$

We will consider two examples, one non anomalous and one anomalous. The corresponding lattices are:

1)$\Lambda_r + \Lambda_v$ of SO(32)(= Z_{16})

2)$(\Lambda_r + \Lambda_v)_{SO(16)} \times \Gamma_8$ (SO(16)$\times E_8$ symmetry).

The first theory (discussed in [10]) is non anomalous since the massless spectrum is the same as in $\frac{Spin(32)}{Z_2}$ theory. The second is anomalous since the s (spinor) of SO(16), needed for $E_8 \times E_8$ anomaly cancellation is absent. We will apply the procedure outlined above to the two theories.

The theta series for the Z_{16} lattice is:

$$A(q,s)=\frac{1}{\eta^{24}(\bar\tau)}\Theta_3(\frac{s}{2\pi i},\bar\tau)^{16} \tag{4.2}$$

To calculate (4.1) for it, it is enough to consider how $A(\bar q,0)$ transforms, since as was shown in [11], the only effect of s in (4.2) is introducing a phase $\propto s^2$ in the transformation law which is cancelled by the explicit exponential factor $e^{\frac{\pi}{Im\tau}s^2}$ in $\bar A$ (Alternatively, one can consider gravitational anomalies, where only $A(\bar q,0,0)$ appears).

If we rewrite $A(\bar\tau,0,0)$ as o+v then $A(\tau+1,0,0)$=o-v and $A(-\frac{1}{\tau}+1,0,0)$=s+c . Thus (4.1) is in this case

$$A'(\bar\tau)=A(\bar\tau)+A(\bar\tau+1)+\bar\tau^{-4}A(\frac{\bar\tau-1}{\bar\tau})=\frac{1}{\eta^{24}}\sum_{w\in 2o+s+c}\bar q^{w^2} \tag{4.3}$$

The integral over the whole boundary of F' in Fig. 2 is thus

$$\int_{B(F')}d\bar\tau\partial_{s_1}..\partial_{s_6}\bar A|_{s=0}=\int_{B(F)}d\bar\tau\partial_{s_1}..\partial_{s_6}\bar A'=\partial_{s_1}...\partial_{s_6}A'|_{s=0;q^0\ \ term}=2\partial_{s_1}..\partial_{s_6}A|_{s=0} \tag{4.4}$$

The last equality in (4.4) follows since the s+c of SO(32) have vectors of length squared≥ 4 and thus do not contribute to (4.4) . The factor of two is due to the two o's in (4.3) . To get the anomaly, we have to substract the contribution from the infinite boundary in Fig. 2 (see (3.10)). But this is precisely equal to (4.4) . Thus the total anomaly vanishes.

In this theory the singularities from the edges of the fundamental domain cancelled. We will see in the second example that this is not a general feature (quite the contrary!). In fact, even if we took the $\frac{Sp(32)}{Z_2}$ (or any other modular invariant theory for that matter), and defined it on F' instead of F, the same boundary integral would not vanish. Of course, this still does not mean that the anomaly would be non zero. The contributions from the above edges would be in that case cancelled by the kinematic choice: this choice is such that the $\tau=i\infty$ term is zero; the term from the edge ($\tau=1$ in Fig. 2a) is equal to the $\tau=i\infty$ one through the symmetry $\tau\to-\frac{1}{\tau}+1$. It is precisely the lack of this symmetry in the models under consideration, that led to the anomaly expression in section 3.

The second example we will consider is that of a SO(16)×E_8 theory with the SO(16) in o+v representations. Again, we have to perform the sum (4.1) , which is very similar to (4.3) and can be seen to give

$$A'(\tau) = \frac{1}{\eta^{24}} \sum_{w \in \Lambda} q^{w^2} \tag{4.5}$$

where $\Lambda = (2o+s+c)_{SO(16)} \times \Gamma_8$ In this case, there are additional massless states generated by the proccess (4.1) . Calculating the anomaly again, we get

$$\partial_{s_1}..\partial_{s_6} A'|_{s=0;q^0 \ term}$$

and the full anomaly is again

$$A = \left[\partial_{s_1}..\partial_{s_6} A' - 2\partial_{s_1}..\partial_{s_6} A\right]|_{s=0;q^0 \ term} = Tr_{s+c} J^6 \tag{4.6}$$

We get a net anomaly as expected from the low energy field theory.

Another example which is simple is the class of models discussed in [9]. There, the lattice on which the compact bosons live is even but not self dual. For these models it is shown in [9] that

$$A' = 0 \tag{4.7}$$

Therefore, the B term in such a model vanishes, by (4.7) . The whole anomaly then comes from the $\tau = i\infty$ region and is given by TrJ^6 over the adjoint representation.

§5 Discussion

We showed that the heterotic string anomaly has its origin in the divergences of the chiral partition function A(q,0,0) on the boundaries of moduli space. There is always a potential divergence at $\tau = i\infty$, which can be removed by a choice of the kinematic invariants. In modular non invariant theories there is at least one additional boundary point where an independent divergence occurs (namely one which is not a reflection of the behaviour at infinity by some symmetry transformation of the theory). By regularizing this other divergence in a particular way, we were able to obtain an expression for the anomaly and saw that it is non zero in general. We also saw that the result we get can be interpreted naturally as a sum of a polygon anomaly with a B coupling which was calculated in [13] generalized to the case of partially modular invariant theories.

Although our approach gives a more accurate description of the anomaly in the modular invariant case too, its main features are apparent in the modular non invariant case. This case was previously considered in ref. [9] and [10]. Since we get different results then these papers, it is neccessary to explain the origin of the differences. It will also help to clarify the ideas presented here.

In [10] the first example of section 4 was considered. It was claimed that in this theory the anomaly vanishes because in the fundamental region of fig. 2b, γ goes to minus itself under $\tau \to -\frac{1}{\tau}$, and the integrand is invariant. As we saw, this invariance is in general not enough to ensure vanishing of the integral,due to the poles of the integrand on γ. This was extensively discussed in the text and in Appendix A. Of course, as we saw, the specific example of [10] gives indeed zero (after all, we know that SO(32) is not anomalous), but this is accidental and has nothing to do with $\tau \to -\frac{1}{\tau}$. The second example of section 4 shows this.

The second argument of [10], regarding finiteness of amplitudes, is more interesting. The idea is to examine the $\tau = 1$ singularity by transforming it to $\tau = i\infty$ through $\tau' = \frac{1}{1-\tau}$ (this is a special example of U' discussed in section 3). Then, as we saw in section 4, the partition sum $A(\tau, 0)$ transforms to $A'(\tau, 0)$ which has simply different particle content (s+c of SO(32) instead of o+v), but finiteness seems to persist. A similar argument was made more precise in [9]. There, the amplitudes are defined by folding into F (the fundamental region of Γ). In a theory like those considered in [10], this would mean that an amplitude is calculated by using (3.13) . It was shown then that in many cases all amplitudes vanish after summing over cosets.

As we saw, (3.13) is a formal expression, which has to be split into several parts in order to be calculated. Now, as shown above, in each of the terms $G(\tau)|W_i$ in (3.13) , the singularity at $\tau = i\infty$ is due to one of the singular points on the boundary of $F'(= B(F'))$. We saw that each W_i spoils the y ordering of the positions of the vertices. In order for $G(\tau)|W_i$ to be well defined at $\tau \to i\infty$, the kinematical invariants and ordering of the z integrals should be chosen so that in $G(\tau)|W_i$ they correspond to a region where the integral over τ converges. Returning to the fundamental region of the subgroup F' this means:

1)The kinematic choice and the division of the z integration has to be made differently near various poles $a \in \gamma$ of the amplitude.

2)For a given pole a in $B(F')$, the z integration is not divided with respect to the values of Im(z), but with respect to some combination of Re(z) and Im(z).

For example, in the special case of section 4, the $\tau = i\infty$ part of the F' integration has to be performed at a certain ordering of the y's, and near $\tau = 1$ a different choice of the division of the z integration has to be made; in particular the division in this region is not with respect to the size of y_i.

Up to now we explained what the choice of [9] and [10] for the regularization of the amplitudes means in terms of our discussion. One may now ask which construction is more natural. The definition of [10], [9] has the basic feature that unitarity is broken already in simple light cone gauge amplitudes like the *four point* (*massless*) function. New particles propagate in the loops[9], some of which can in general have negative norm, and thus there is no need to even get to the anomaly problem. This prescription is inconsistent at a much more basic level.

It is perhaps understandable that the above regularization should run into this kind of problems. After all, it involves in *some regions* of the integration over τ, ordering the particles on the torus not with respect to their Im(z). Since Im(z) is related to time, this prescription does not correspond to calculating a time ordered product of the fields. It is hardly surprising that one gets nonunitary results. In operator formalism such a prescription would never arise, because of the Im(z) ordering.

We have not considered in detail general amplitudes using our regularization. It would seem from the preceeding discussion (and from operator formalism) that properly regularized amplitudes in our formalism should be well behaved; the sickness will then show up in the anomaly as shown in this paper.Of course, this will lead to other sicknesses too.

If the model is only partially modular invariant, but the massless spectrum is anomaly free, it is plausible that there exists a regularization in which unitary amplitudes are obtained for all proccesses, which reduce to the correct low energy limit of gauge theory.

ACKNOWLEDGEMENTS

I would like to thank A. Schwimmer and N. Seiberg for numerous helpful discussions, suggestions and critical reading of the manuscript. This work was supported by a BSF grant.

APPENDIX A

In this Appendix, we define various functions used in the text and discuss some simple facts which show the origin of anomalies in the modular non invariant case. It is convenient to use the covariant lattice language [17] to define various gauged partition sums[11,13]. The gauged partition sum of the model, which contains all the information about traces in the chiral sector is defined by:

$$A(\tau,s,R)=\frac{1}{\eta^{24}(\tau)}\exp\left[\sum_{k=1}^{\infty}\frac{1}{4k}\frac{1}{(2\pi i)^{2k}}G_{2k}(\tau)Tr\left(\frac{iR}{2\pi}\right)^{2k}\right]\sum_{(w_L,w_R)\in\Lambda}{}' q^{w_L^2}e^{w_L\cdot s} \quad (A.1)$$

where the primed sum is over vectors describing space time fermions ($w_R=w_R(s)$ $w_R^2=2$). R is the gravitational curvature and s the skew eigenvalues of F (for notational details see [13]).

It was shown in [11], that A(q,s,R) is not a modular form; instead it transforms in a modular invariant theory as:

$$A(\tau',s',R')=\exp\left[\frac{ic(TrF^2-TrR^2)}{32\pi^3(c\tau+d)}\right](c\tau+d)^{-n}A(\tau,s,R) \quad (A.2)$$

where

$$\tau'=\frac{a\tau+b}{c\tau+d}\ ,\ s'=\frac{s}{c\tau+d}\ ,\ R'=\frac{R}{c\tau+d}\ ,\ D=2n+2$$

It is possible to cancel the anomalous term in the transformation (A.2) and define a modular form out of A in two ways(basically):

1)Preserving the holomorphicity in τ which is manifest in (A.1) (this is convenient for proving trace identities):

$$\tilde{A}(q,F,R)=A(q,s,R)\exp\left[\frac{1}{64\pi^4}G_2(\tau)(TrF^2-TrR^2)\right] \quad (A.3)$$

where

$$\frac{1}{64\pi^4}TrF^2=-\frac{1}{4\pi^2}s^2 \quad (A.4)$$

2)Breaking holomorphicity (this is the choice chosen by string amplitudes as we saw):

$$\bar{A}(q,F,R) = \exp\left[\frac{1}{64\pi^4}\frac{\pi}{Im\tau}(TrF^2 - TrR^2)\right] A(q,s,R) \tag{A.5}$$

$\bar{A}$ entered the discussion in the text; here we will consider $\tilde{A}$. From (A.2) and (A.3) it is obvious that $\tilde{A}$ is a holomorphic modular form of weight -n. In [11] it was shown that the fact that

$$F(\tau) = \partial_{s_1}..\partial_{s_{n+2}}\tilde{A}|_{s=0} \tag{A.6}$$

is a form of weight 2, causes its q^0 term to vanish. From definitions (A.1) and (A.3) it then follows that TrJ^{n+2} (J is some Cartan Subalgebra (CSA) generator of the gauge group) is proportional to a combination of lower traces of J (the trace is over the massless chiral fermions). This trace identity causes the anomaly to vanish[1] . Thus the vanishing of the q^0 term in $F(\tau)$ is important for anomaly cancellation. We will check when it happens by considering an elementary proof.

Consider a general dimension two holomorphic modular form $F(\tau)$. If it does not have poles in a region, then

$$\partial_{\bar{\tau}}F = 0 \tag{A.7}$$

in that region. Consider the region G in Fig. 3 bounded by $|Re\tau| = \frac{1}{2}$, $|\tau| = 1$ and $Im\tau = b > 1$. Now integrate $\partial_{\bar{\tau}}F$ over G, or equivalently integrate F over the boundary of G. The integrals over $|Re\tau| = \frac{1}{2}$ cancel each other by $\tau \to \tau + 1$ symmetry of F. The integral over $|\tau| = 1$ vanishes by $\tau \to -\frac{1}{\tau}$. The $Im\tau = b$ integral leaves the q^0 term in F only. Since the integral is identically zero by (A.7), we conclude that :

$$F(\tau)|_{q^0\ term} = 0 \tag{A.8}$$

Of course, (A.8) does not hold for any two form F. It is easy to see when it breaks: if F has a pole somewhere in G, (A.7) does not hold anymore and (A.8) is spoiled. A trivial example is the weight two form $\frac{G_6}{G_4}$ (the G's are Eisenstein functions), which does not

[1] Taking derivatives w.r.t. R generates similar relations for the gravitational case. To avoid repetition, this case is ignored from now on.

satisfy (A.8) because G_4 has a zero on the boundary $|\tau| = 1$. and therefore the boundary integral of $\frac{G_4'}{G_4}$ has to be performed over the trajectory of *Fig.* 1 *instead of that of* Fig. 3, thus producing finite boundary terms. For modular invariant theories[11], however, (A.8) holds since $F(\tau)$ defined by (A.6) is regular on G for any *finite* b (the *only* possible pole is at infinity).

The discussion presented above is useful for analyzing the modular non invariant case. Then, to prove a trace identity such as (A.8), we would have to modify G so that the boundaries we use are those of the fundamental region for a subgroup of Γ under which (A.6) is covariant. Here, however, we encounter a problem; for any subgoup, all fundamental regions contain points where the partition sum (A.6) diverges, e.g. with $Im\tau = 0$ (see sections 3,4). These singularitites, while not causing infinitites, produce finite contributions from the vicinity of the poles, thus spoiling (A.8). The contributions of these poles might be zero in specific cases but in general they are certainly non zero, as we saw in section 4.When such contributions vanish, the anomaly is zero too. In general they contribute and this means that the polygon anomaly can not be cancelled by a B contribution.

APPENDIX B

There are several integral formulas which we use in the text to evaluate the z integrals of correlators: as shown in [13]:

$$\int d^2z \Delta(z-w) = 0 \tag{B.1}$$

$$\int d^2z \Delta'(z-w) = 0 \tag{B.2}$$

where Δ and Δ' are defined in (2.9) and (2.12) respectively. This essentially follows from the relations:

$$\begin{aligned} \Delta(z) &= \partial_z G(z) \\ \Delta'(z) &= \partial_z \Delta(z) \end{aligned} \tag{B.3}$$

where G is the boson correlator (2.8) . Since $G(z+1) = G(z+\tau) = G(z)$, (B.1), (B.2) give zero after z integration. Since

$$\begin{aligned} \tilde{\Delta}'(z) &= \Delta'(z) - \frac{\pi}{2Im\tau} \\ \int d^2z \tilde{\Delta}'(z-w) &= -\frac{1}{2}\frac{\pi}{Im\tau} \cdot Im\tau \end{aligned} \tag{B.4}$$

Another usefull integral is [7]:

$$\frac{1}{2\pi i}\int \frac{d^2z}{Im\tau}\Delta(u-z)\Delta(u-w) = \left[\partial_\tau + \frac{Imz}{Im\tau}\partial_z + \frac{Imw}{Im\tau}\partial_w\right] G(z-w) + const \tag{B.5}$$

which can be verified by differentiating (B.5) with respect to z or w.

We also use in the text the assymptotic expansion for Δ'. In [13] it was shown that Δ' is given in terms of x,y ($z = x + \tau y$) by:

$$\Delta'(x,y) = -\frac{\pi}{2Im\tau}\sum_{k,m}{}' \frac{m-k\bar{\tau}}{m-k\tau} e^{2\pi ikx} e^{2\pi imy} \tag{B.6}$$

where the primed sum is over all pairs $(k,m) \neq (0,0)$. Taking $\tau \to i\infty$

$$\Delta' = -\frac{\pi}{2Im\tau}\left[2\delta(y) - \delta(x)\delta(y) - 1\right] \tag{B.7}$$

FIGURE CAPTIONS

Fig. 1: Integration contour for modular forms with poles at cusps.

Fig. 2: Possible choices for the fundamental region for Γ'.

Fig. 3: The fundamental region F for Γ.

References

1. M. B. Green, J. H. Schwarz and E. Witten, Superstring Theory, Cambridge Press (1987)

2. M. B. Green and J. H. Schwarz, Phys. Lett. 151B (1984), 21; Nucl. Phys. B255 (1985), 93

3. D. J. Gross, J. A. Harvey, E. Martinec and R. Rohm, Phys. Rev. Lett. 54 (1985), 502; Nucl. Phys. B256 (1985), 253; B267 (1986), 75

4. D. Friedan, E. Martinec and S. Shenker, Nucl. Phys. B271 (1986), 93

5. J. Polchinski and Y. Cai, UTTG-09-87

6. A. Sugamoto and H. Suzuki, Phys. Rev. Lett. 57 (1986), 1665; KEK TH 132 (1986) H. Suzuki, KEK TH 126 (1986)

7. K. Miki, Nucl. Phys. B291 (1987), 349

8. S. Yahikozawa, Nucl. Phys. B291 (1987), 369

9. D.J. Gross and P.F. Mende, Nucl. Phys. B291 (1987), 653

10. M. Sakamoto, KYUSHU-87-HE-1 (1987)

11. A.N. Schellekens and N.P. Warner, Nucl. Phys. B287 (1987), 317

12. W. Lerche, B.E.W. Nilsson and A.N. Schellekens, Nucl. Phys. B289 (1987), 609

13. W. Lerche, B.E.W. Nilsson, A.N. Schellekens and N.P. Warner, Nucl. Phys. B299 (1988), 91

14. M.B. Green and N. Seiberg, *IASSNS-HET-87/38*

15. J.J. Atick, L.J. Dixon and A. Sen, *Nucl. Phys.* **B292** (1987), 109

16. M. Dine, I. Ichinose and N. Seiberg, Nucl. Phys. **B293** (1987), 253

17. W. Lerche, D. Lust and A.N. Schellekens, Nucl. Phys. **B287** (1987), 477

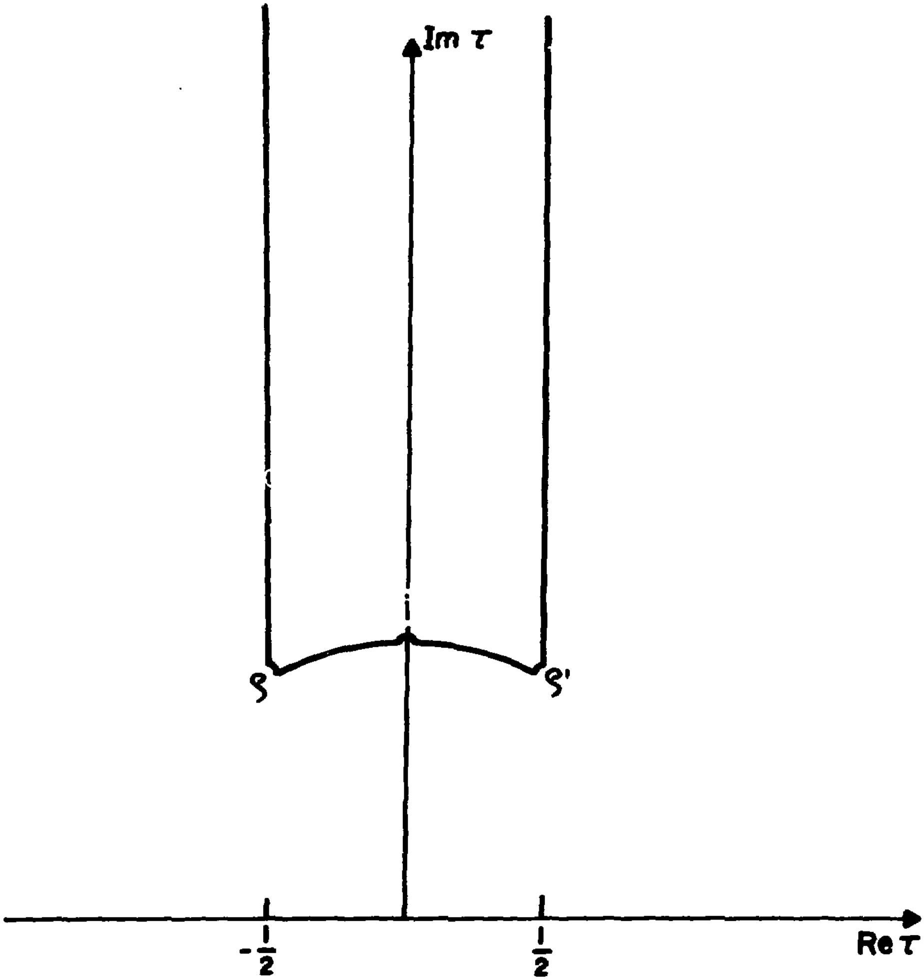

FIG 1

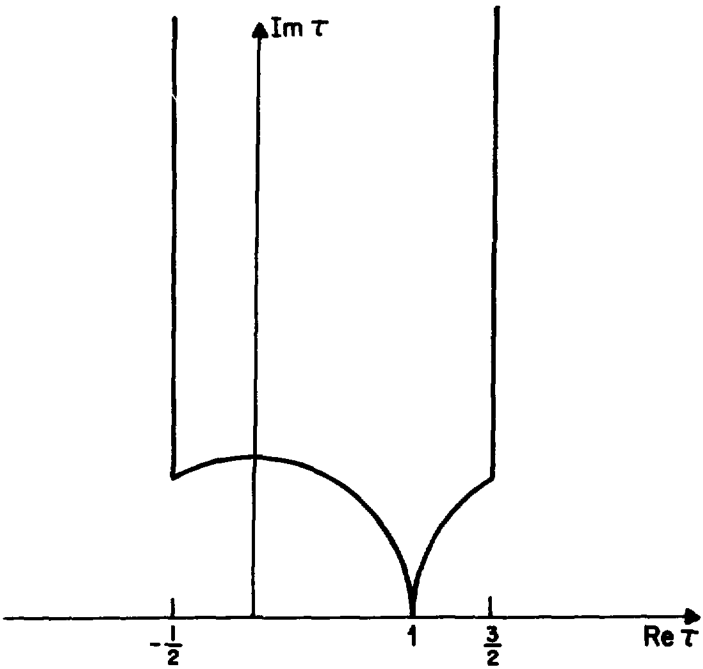

FIG 2a

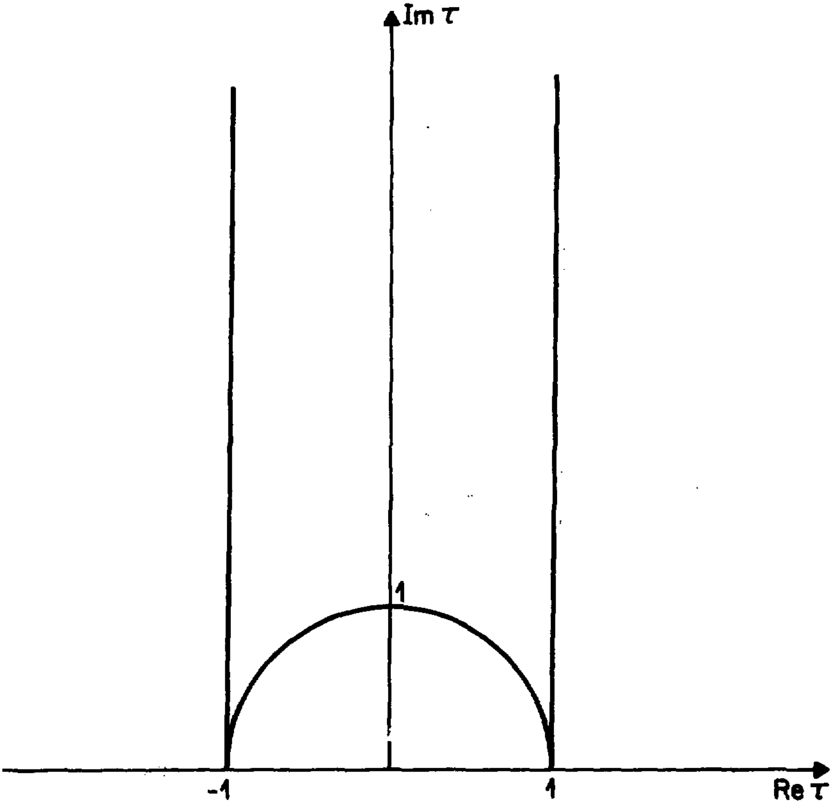

FIG 2b

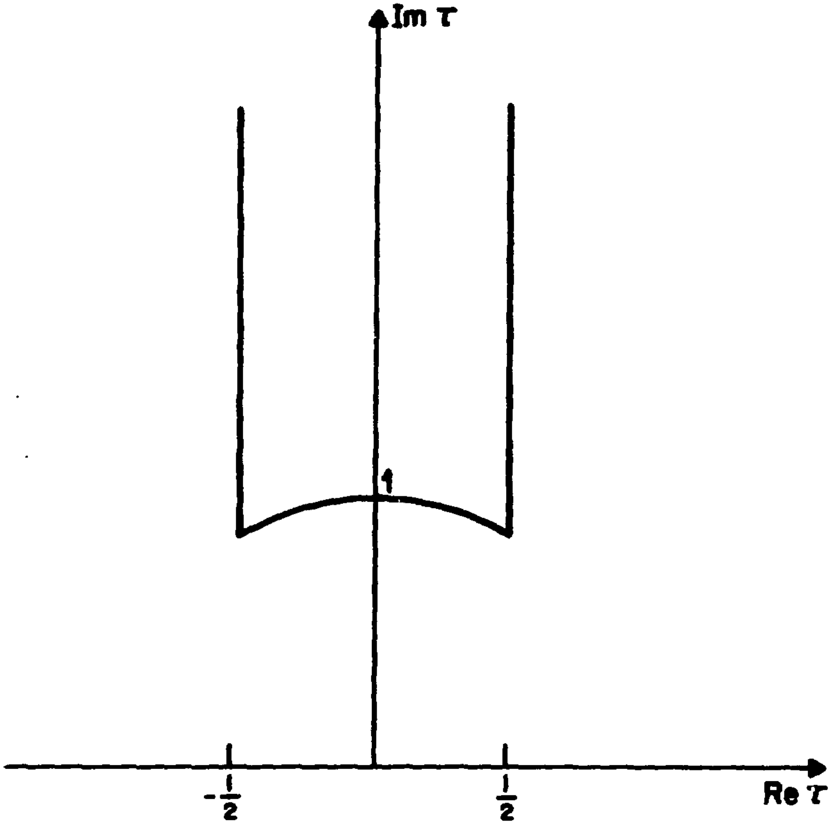

FIG 3