921

STRINGS AND SOLITONS IN GAUGE THEORIES

by

NEIL GEOFFREY TUROK

A thesis presented for the Degree of
Doctor of Philosophy of the University of London
and Diploma of Membership of Imperial College

Department of Physics,
Blackett Laboratory,
Imperial College,
London SW7 2BZ

May 1983

STRINGS AND SOLITONS IN GAUGE THEORIES

N.G. Turok

ABSTRACT

This thesis covers two topics in nonabelian gauge theories.

The first is the existence, in some spontaneously broken gauge theories of topologically stable 'string' solutions. A comprehensive classification of those grand unified theories yielding these solutions is given. The cosmological consequences of strings produced at the grand unification phase transition are investigated and it is shown how spinning loops of string may lead to the formation of galaxies.

The second topic is the connection between gauge theories and integrable systems, notably the Toda systems. These are known to possess soliton solutions, both topological and non-toplogical, and this fact is related in an essential way to their integrability. It is shown how self-dual gauge field configurations give rise to the Toda equations, and a complete classification of Toda systems is given. A new simple algebraic proof of the integrability of Toda systems is given which clarifies the role of an underlying 'Kac-Moody' algebra for the Toda lattice equations. This also gives a clue as to the quantisation of these systems, and a link with the 'Quantum Inverse Scattering' method.

PREFACE

The work presented in this thesis was carried out in the Blackett Laboratory, Imperial College, under the supervision of Dr. D.I. Olive. The material presented is original except where otherwise stated and has not been submitted for a degree of this or any other university.

Section 1.2 and Chapter II and III represent work done in collaboration with Dr. D.I. Olive, part of Section 1.4 with Prof. T.W.B. Kibble and Section 1.6 with P. Bhattacharjee and Prof. T.W.B. Kibble.

It is a pleasure to thank Dr. D.I. Olive for his constant help and invaluable guidance throughout this work. I would also like to thank Prof. T.W.B. Kibble similarly. I thank P. Bhattacharjee for useful collaboration. Lastly, I would like to thank all the other members of the Theoretical Physics Group who made my stay at Imperial College so enjoyable.

I also thank the S.E.R.C. for financial support.

Parts of this thesis have been or are to be published in the following articles; Phys. Lett.116B(1982)141 (with T.W.B. Kibble), Phys. Lett.117B(1982)193 (with D.I.Olive), Phys. Lett.119B (1982)95 (with P.Bhattarcharjee and T.W.B.Kibble), Phys. Lett.123B (1983)387, Phys. Lett., to be published. Nuc. Phys.B215(1983)470 (with D.I.Olive), Nuc. Phys., to be published (with D.I.Olive).

To Ben and Mary

CONTENTS

INTRODUCTION

This thesis is written as one of the principal predictions of the non-abelian gauge theory of the electroweak interactions, the existence of a massive charged vector boson at 80 GeV, has just been reported confirmed. However, the theory has been believed by most theoretical physicsts for some time and the result caused little real surprise. The reason for this lies in the remarkable and unique mathematical and physical properties of non abelian gauge theories (or Yang-Mills theories) invented by C.N. Yang and R.L. Mills in 1954 [1]

Indeed the impact at the present time of gauge theories on both mathematics and physics can hardly be overestimated. Some of their special properties for physics are:

i) They are the only known consistent, renormalisable local quantum field theories of the strong and electroweak interactions.

ii) They are asymptotically free - a property conjectured to be necessary [2] for any nontrivial quantum field theory (which Quantum Electrodynamics and ϕ^4 theory fail). More concretely the property is essential for a theory of the strong interactions which at least explains the success of the parton model.

iii) They predict a natural unification at very high energies of the strong and electroweak gauge interactions into a single 'Grand Unified Theory' (GUT). This would describe in a simple, aesthetic way the physics of the very early universe, the high degree of symmetry being spontaneously broken as the universe cooled.

Many problems remain with gauge theories, however, mainly because their complex nonlinear structure frustrates attempts to understand the quantum theory properly i.e. what the particle states are, at least partly by allowing localised non-dissipative solutions which are topologically stable and persist in the quantum theory. In the Euclidean theory, one may attempt build the quantum theory in a semiclassical way upon classical instanton solutions. Indeed current theories of quark confinement give a key role to these solutions and others, 'monopoles' and 'flux lines', which are sometimes generically referred to as 'solitons'. The quotation marks indicate the as yet poorly understood nature and role of these topological objects in an unbroken gauge theory like Quantum Chromodynamics. The classification of all instantons by Atiyah, Drinfeld, Hitchin and Manin[3] has also led to far-reaching new developments in mathematics.

Similar solutions, though rather better defined, result when a grand unified theory is spontaneously broken by the Higgs mechanism[4]. These are grand unified monopoles, strings and domain walls. These may have drastic consequences for cosmology - the 'monopole problem' for instance which occurs in most GUTs, is that the theories predict more monopole matter than any other, in flagrant disagreement with observation. Similarly, domain walls, although not predicted by most GUTs, are cosmologically unacceptable[5].

This thesis is directed towards both these problems - towards understanding the quantum field theory of gauge theories and their cosmological predictions.

Chapter I is devoted to the occurence in some GUTs of topologically stable solutions with string-like form, 'strings'. Their spherical analogues, magnetic monopoles, have been extensively investigated in the literature and occur as something of a problem for GUTs, as explained above. As I hope to show, 'strings' in grand unified theories could play a rather more positive cosmological role - that is, they may seed the process of galaxy formation.

Section 1.1 describes what strings are and gives criteria for their existence in a GUT. Section 1.2 describes what seems the most physical examples, using techniques of Lie group representation theory, providing a fairly complete analysis of realistic GUTS. Section 1.3 describes the behaviour of strings in the very early universe - how they yield the correct density fluctuations for galaxy formation under certain assumptions. Section 1.4 describes the production of non-self-intersecting loops in the very early universe, substantiating the previous analysis. Section 1.5 outlines in more detail how galaxies would form, and how large scale filamentary structures could also be caused by strings. Section 1.6 shows how collapsing strings could yield the universe's baryon number.

Chapter II is devoted to understanding the connection between gauge theories and integrable systems, notably the Toda systems [6] These are known to possess soliton solutions, both topological and non-topological, and this fact is due in an essential way to their integrability. The Toda molecule equations have already arisen in the study of spherically symmetric monopoles and instantons - it is shown in Section 2.1 how the Toda lattice equations also arise in self-dual Yang-Mills configurations. The Toda equations and their symmetries are discussed in Sections 2.2-2.6 using Dynkin diagrams - in particular a procedure for obtaining all Toda equations from the simplest ones is given, which promises to be useful in the analysis of self-dual monopoles and other Toda systems. Chapter III is devoted to an understanding of the algebraic structure underlying the classical integrability of all Toda systems. A purely Lie algebraic operator, called $\mathbb{P}$ is found from which the Lax Pair for the one-dimensional Toda systems may be constructed, and which guarantees the complete integrability of the systems. It is shown how this operator depends in a uniform way on the root system of the underlying algebras - the simple Lie algebras for the Toda molecule systems, and the affine Kac-Moody algebras for the Toda lattice systems. It includes a new simple proof of the integrability of all Toda systems. Interestingly, it also gives a strong hint of how to quantise the theories, providing a link with the 'Quantum Inverse Scattering Method'.

CHAPTER I

GRAND UNIFIED STRINGS AND GALAXY FORMATION

1.1 WHAT STRINGS ARE AND CRITERIA FOR THEIR EXISTENCE

A string is a localised topologically stable solution to the coupled field equations of a Yang-Mills-Higgs system with cylindrical form and finite energy per unit length. The simplest examples of strings are the flux lines observed in superconductors. Their occurrence in spontaneously broken gauge theories was first pointed out by Nielsen and Olesen [1].

The system of fields for a Yang-Mills-Higgs system is defined by the Lagrangian density

$$\mathcal{L} = -\tfrac{1}{4}(F^a_{\mu\nu})^2 + \tfrac{1}{2}(D_\mu \Phi^\alpha)^2 - V(\Phi) \tag{1.1.1}$$

where $$F^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - e f_{bc}{}^a W^b_\mu W^c_\nu$$

is the gauge field strength, W^a_μ the gauge potential, e the coupling constant and $f_{bc}{}^a$ the structure constants of the algebra of generators T_a of the gauge group G, defined by $[T_b, T_c] = i f_{bc}{}^a T_a$. Φ is the higgs field - the index α runs over the dimension of the representation in which it lies. $V(\Phi)$ is a fourth order polynomial (higher orders are non-renormalisable) invariant under gauge transformations

$$\Phi^\alpha(x) \longrightarrow \Phi'^\alpha(x) = D(g(x))^\alpha{}_\beta \, \Phi^\beta(x) \tag{1.1.2}$$

where $D(g)^\alpha{}_\beta$ is the matrix representative of the group element $g \in G$, the gauge group, in the representation Φ is in. The covariant derivative of Φ is given by

$$(D_\mu \Phi)^\alpha = \partial_\mu \Phi^\alpha + ie\, D(T_a)^\alpha{}_\beta W_\mu^a \Phi^\beta$$

We are interested in cylinder-like solutions. Imagine we are in the rest-frame of a 'string' which is locally straight. Let us look for criteria for the existence of static solutions with cylindrical symmetry. For static configurations the Hamiltonian is proportional to the Lagrangian, so any solution must be a stationary point of the Hamiltonian, in this case the energy per unit length,

$$\begin{aligned} E/L &= \int d^2x \, \tfrac{1}{2}\Big(\underbrace{(E_i^a)^2 + (B_i^a)^2}_{T_W} + (D_i \Phi^\alpha)^2\Big) + V(\Phi) \\ &= \qquad\qquad T_W \qquad + \quad T_\Phi \quad + \quad V \end{aligned} \tag{1.1.3}$$

where i runs over spatial indices,

$$E_i^a = -F_{oi}^a \qquad B_i^a = -\tfrac{1}{2}\epsilon_{ijk}F_{jk}^a,$$

and the length L comes from integration along the axis of cylindrical symmetry. This expression should be stationary under simple scale transformations of the fields

$$x'^\mu = \lambda x^\mu, \qquad \Phi'(x') = \Phi(x), \qquad W'_\mu(x') = \frac{1}{\lambda} W_\mu(x)$$

Under these transformations, the three terms in (1.1.3) scale as

$$\begin{array}{ccc} T_W & T_\Phi & V \\ \lambda^2 & 1 & \lambda^{-2} \end{array}$$

and it is clear there exists no stationary point if any of the three terms is absent. The T_Φ term is necessary because it is only through this term that the gauge potential and higgs field are coupled.

One may compare this analysis with ball-like spherically symmetric solutions, 'monopoles', and wall-like solutions symmetric in two directions, 'domain walls';

	T_W	T_Φ	V
monopoles	λ	λ^{-1}	λ^{-3}
domain walls	λ^3	λ	λ^{-1}

It is immediately clear that strings are actually the most complex configurations - for in the case of monopoles, one can still have static solutions with $V(\Phi)$ put equal to zero (this is called the Bogomolny-Prasad-Sommerfeld or self-dual limit) and in the case of domain walls one does not need the T_W term, nor actually gauge fields at all.

However, in Grand Unified Theories one does have all three terms, and the higgs potential cannot be zero (this would give unobserved massless scalar particles.) There is only one quantity with the dimensions of a mass in the expression (1.1.3) - it is the vacuum expectation value $\langle\Phi\rangle$ of the higgs field which is non zero when the gauge symmetry is broken. In fact $|\langle\Phi\rangle|$ is also approximately the temperature at which the symmetry breaking phase transition occurs.[2] By scaling all dimensional quantities out of the action,

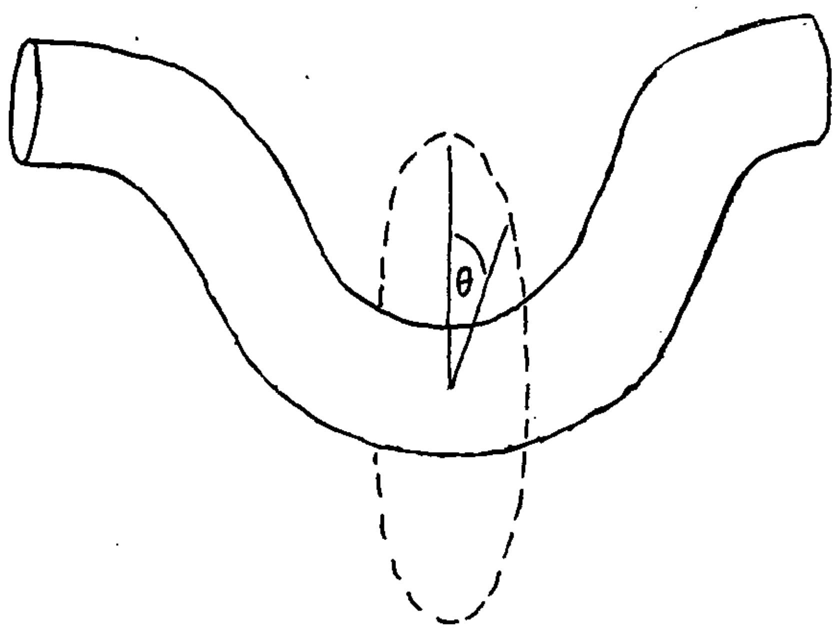
$A_\mu = a_\mu |\langle\Phi\rangle|$, $x = \frac{1}{e|\langle\Phi\rangle|} y$, $\Phi = |\langle\Phi\rangle| f$ we find (1.1.3) reduces to

$$E/L = |\langle\Phi\rangle|^2 \tau(\alpha/e^2, \beta/e^2 \ldots)$$

where α and β are arbitrary dimensionless constants in the higgs potential and τ is a dimensionless topological integral.

We obtain $\mu \equiv E/L \simeq |\langle\Phi\rangle|^2$ (1.1.4) as our estimate of the mass per unit length. If $|\langle\Phi\rangle| \sim 10^{15}$ GeV as in Grand Unified theories, this yields $\mu \approx 10^{19}$Kg m^{-1}!

Of course, dimensional analysis is hardly enough - we need a more powerful criterion for the existence of strings. A very simple one exists and is related to the topology of the group considered. Consider a stringlike configuration of fields:



At large distance from the string, for a finite energy per unit length Φ must minimise the potential. In fact as we wind around the string Φ as a function of θ (see diagram) provides a map from the circle S_1 into $\mathcal{M}_0 \equiv \{ \Phi : \frac{\partial V}{\partial \Phi} = V(\Phi) = 0 \}$ being the set of minima of the higgs potential, which is chosen positive semi-definite. It is well known that if there is no accidental degeneracy in the potential,

$$\mathcal{M}_0 \approx G/H \qquad (1.1.5)$$

where H is the subgroup of G leaving Φ invariant. In mathematical language, at large distance from the string we have

$$\Phi(\theta) \in \pi_1(G/H)$$

where π_1 is the first homotopy group of G/H . If $\Phi(\theta)$ is not deformable continuously to a trivial map i.e. there does not exist a $f(\theta,t)$ such that $f(\theta,0)) = \Phi(\theta)$ and $f(\theta,1)$ = constant, independent of θ, then the string solution cannot decay away continuously to the vacuum. It is then topologically stable. An example is provided by the occurrence of flux lines in superconductors. Here the U(1) of ordinary electromagnetism is broken completely by the complex scalar field Φ describing Cooper pairs. Thus G = U(1) and H $= \mathbb{1}$ so $\pi_1(G/H) = \mathbb{Z}$, being the group of maps from circles to circles, labelled by a winding number $0 \leq n \leq \infty$.

In grand unified theories, in the simplest case we are interested in the spontaneous breaking of a simple group G (i.e. the symmetry group of a gauge theory with a single coupling constant) to some subgroup H εG. Now an arbitrary simple (or even semi-simple) group G is of the form $\tilde{G}/g(G)$ where g(G) is a subgroup of the centre of $\tilde{G}$, the universal covering group of G which is simply connected. If $\tilde{G}$ has a subgroup $\tilde{H}$ leaving Φ invariant, then G has a subgroup H obtained by identifying those elements of g(G) contained in $\tilde{H}$. In fact since the gauge group is G not $\tilde{G}$, we know that g(G) leaves Φ invariant, so $\tilde{H}$ contains all of g(G). Thus $H = \tilde{H}/g(G)$. Clearly then

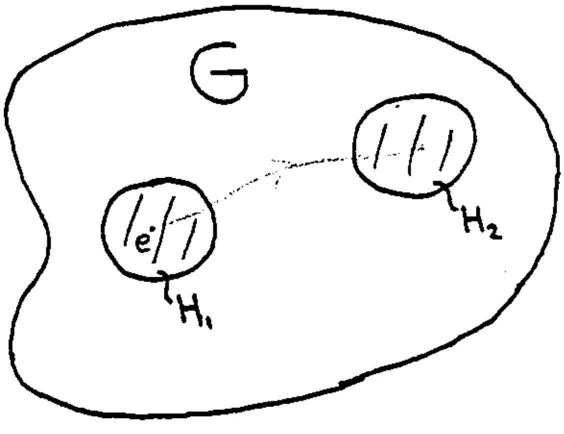
$$G/H \approx \tilde{G}/\tilde{H} \tag{1.1.6}$$

since both involve taking $\tilde{G}$, identifying g(G) and H , and these operations clearly commute.

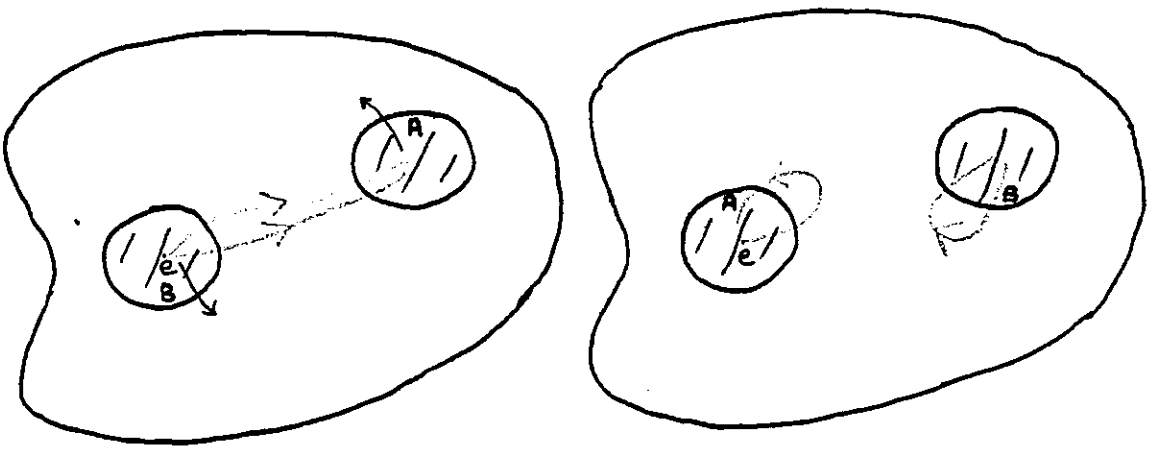
We may thus restrict ourselves to the case where $G = \tilde{G}$ is simply connected, so that $\pi_1(G) = \pi_0(G) = 0$ and homotopy theorems yield

$$\pi_1(G/H) \approx \pi_0(H) \qquad (1.1.7)$$

which is nontrivial if H has several disconnected components. The following picture illustrates how this works if H has two disconnected pieces, H_1 and H_2



Then the path shown in red is clearly nontrivial in G/H i.e. cannot be deformed to the identity. However, repeating the path twice in G/H , we find



A and B differ by an element of H , but can be deformed by elements in G as shown to obtain the second configuration. The loops can then be shrunk to zero. Thus $\pi_1(G/H) = \mathbb{Z}_2$.

1.2 Z_2 STRINGS IN GRAND UNIFIED THEORIES

The purpose of this section is to point out that Z_2 strings arise very naturally in grand unified theories as a consequence of the sort of symmetry breaking which seems to be needed to give masses to fermions. This generalises to any semisimple group, for example, E_8, E_7 and E_6, a result recently obtained for SO(10) [3]. In such theories the chiral fermions are placed in a fundamental representation of the semisimple gauge group G (in fact the 248, 56, 27 and 16 respectively for the groups mentioned). This representation has as highest weight a "fundamental weight" associated with a point of the Dynkin diagram of G. It seems that typically G is broken to the subgroup whose Dynkin diagram is obtained by deleting this same point from the original Dynkin diagram. Thus in the examples E_8 is broken to E_7, E_7 to E_6, E_6 to SO(10) and SO(10) to SU(5).

We shall show that Z_2 strings result from a natural choice of Higgs mechanism which achieves these two goals of giving some of the fermions masses and of yielding the desired breaking. For E_8 this Higgs is a complexified 27,000, for E_7 a complexified 1463, for E_6 a 351, and for SO(10) a 126.

Our result is very general, being valid for any semisimple Lie grou and any fundamental representation of that group. In fact we shall also show how to obtain Z_n strings. The proof uses Lie algebraic techniques recently developed in connection with monopole theory [4].

Let the gauge symmetry G of the Lagrangian be broken to a subgroup H_Φ by the Higgs field Φ. Its vacuum manifold consists of the coset space G/H_Φ so that a loop in space may "capture" a string if $\pi_1(G/H_\Phi)$ is nontrivial. If G is connected, as we shall assume, then/as shown above without loss of generality it can be taken to be simply connected and homotopy theorems assure us that

$$\pi_1(G/H_\Phi) \cong \pi_0(H_\Phi) \cong H_\Phi/H_c \qquad (1.2.1)$$

where H_c is the component of H_Φ connected to the identity. By continuity H_c is an invariant subgroup of H_Φ so that the topological quantum numbers (1) actually from a finite group. Topologically distinct strings thus correspond to the disconnected components of H_Φ.

In the original examples of Nielsen and Olesen[1] H_Φ was a finite group but this is impossible in a physical GUT theory since H_Φ must contain U(3), the exact gauge symmetry of nature.

The structure of the strings resulting depends critically on the choice of the Higgs field. For the reasons explained above we shall concentrate on choices of Φ yielding an H_Φ containing as a subgroup the group K obtained by exponentiating the generators of G corresponding to the Dynkin diagram D(K) obtained by deleting one point (and its links) from the Dynkin diagram of G, D(G). By definition K is connected and it can be shown to be simply connected[4]. The points of D(G) correspon to simple roots of G[5,6] and the simple root corresponding to the deleted point is denoted α_ϕ. If α_i are the simple roots of G its fundamental weights λ_i are defined by

$$2\lambda_i.\alpha_j/(\alpha_j)^2 = \delta_{ij} \quad i,j = 1....\text{rank } G \qquad (1.2.2)$$

λ_ϕ then denotes the highest weight of the fundamental representation, D, carrying the chiral fermions. The fermion mass term transforms as $(D \times D)_{sym}$ so that any Higgs yielding fermion masses must lie in the symmetric part of the Clebsch-Gordon series DxD. (By the same token so must any Higgs which is a difermion condensate). The highest dimensional such representatio automatically has highest weight $2\lambda_\phi$ We shall now show that the correspondir. state $|2\lambda_\phi\rangle$ in the representation space of that representation is annihilate by precisely those generators of G which are generators of K.

That $|\mu\rangle$ has weight μ means simply that

$$H^i|\mu\rangle = \mu^i|\mu\rangle$$

for each Cartan subalgebra generator H^i. That $2\lambda_\phi$ is the highest weight of the representation means that there is no state with weight $2\lambda_\phi+\alpha$, for α a positive root of G. So $E_\alpha|2\lambda_\phi\rangle$ must then vanish. Now the α string through $2\lambda_\phi$ has length $2\alpha.(2\lambda_\phi)/\alpha^2$ (see Humphreys [6] p.114) and this vanishes for all roots α of G in which α_ϕ does not occur, i.e. roots of K. Hence the compact generators $(E_\alpha+E_{-\alpha})/2$ and $(E_\alpha-E_{-\alpha})/2i$ annihilate $|2\lambda_\phi\rangle$ if and only if α is a root of K, as does $\alpha.H$. This is also true if we consider the state $|n\lambda_\phi\rangle$ (n an integer) which is the highest weight state obtained by symmetrizing the Kronecker product of n representatio D. Thus a Higgs field whose vacuum expectation value is parallel to $|n\lambda_\phi\rangle$ automatically yields an H_Φ which we shall denote $H(n)$ whose generators are those of K.

Before explaining the global structure of H_{Φ} we shall consider a situation of interest in monopole theory when the Higgs field lies in the adjoint representation rather than the representations discussed above. If it assumes a direction along λ_{ϕ} in vacuo then [4,7]

$$H_{\Phi} = H' \cong U(1) \times K/Z_{\ell} \tag{1.2.3}$$

where K is as before and the U(1) is generated by $\lambda_{\phi}.H$. Division by the cyclic group Z_{ℓ} indicates that ℓ points of the centre of K actually lie in the U(1). Let us denote as V_0 the generator of this Z_{ℓ} (a precise formula appears later).

Our main result is that $H(n)$, the little group of $|n\lambda_{\phi}\rangle$ is

$$H(n) \cong Z_{n\ell} \times K/Z_{\ell} \tag{1.2.4}$$

Thus the U(1) occurring in (3) is broken to a cyclic group of order $n\ell$ whose generator will be an nth root of V_0. The topological quantum numbers for the strings are therefore given by (1) as

$$\Pi_1(G/H(n)) \cong Z_n \tag{1.2.5}$$

In particular, if the fermions are to acquire masses, n=2 and Z_2 strings result.

Notice that since H(n) is a subgroup of H' given by (3) the result (5) concerning strings holds even when there is an adjoint Higgs as described though the monopoles associated with the U(1) disappear. In grand unified models both sorts of Higgs mechanism are often thoughtto hold simultaneously.

Now let us prove (4). Consider the component of H(n) connected to the identity e and a point g_0 in it. Then $g^{-1}\partial g$ evaluated along the path joining e to g_0 annihilates $|n\lambda_\phi\rangle$ at each point and hence is a linear combination of the generators of K. Hence K is this component of H(n).

Consider any W in H(n) not necessarily in K: then by continuity

$$WKW^{-1} = K \tag{1.2.6}$$

Let C(K) denote the Cartan subalgebra of K.
Then $WC(K)W^{-1}$ is an equally good Cartan subalgebra which, by a general theorem [8], can be conjugated within K back to C(K). Thus we can choose k in K so that if $W' = kW$ then

$$W'C(K)(W')^{-1} = C(K) \tag{1.2.7}$$

W' and W lie in the same disconnected component (1) of H(n).

Now let us choose as Cartan subalgebra of G that spanned by C(K) and $\lambda_\phi . H$. Consider

$$W'\lambda_\phi . H(W')^{-1} - \lambda_\phi . H$$

It annihilates the state $|n\lambda_\phi\rangle$ and so belongs to the Lie algebra of K but, by (7) it commutes with each Cartan subalgebra generator of K and hence is a linear combination of them. Thus

$$W'C(G)(W')^{-1} = C(G) \tag{1.2.8}$$

$$\text{i.e.} \quad W'x.H(W')^{-1} = \sigma(x).H$$

where it was proven in [8] that σ is an element of the Weyl group of G, W(G). It follows from (8) and the one dimensionality of the vector space correspond to the highest weight of an irreducible representation [6] that

$$W'|n\lambda_\phi\rangle = \text{const}\,|n\sigma(\lambda_\phi)\rangle$$

But since W' lies in H(n), $\sigma(\lambda_\phi)=\lambda_\phi$. However the subgroup of W(G) with this property is precisely W(K). For, suppose that when σ is written as a minimal length product of reflections in simple roots of G (see Humphreys[6] p.51), σ_ϕ, the reflection in α_ϕ, appears:

$$\sigma = \sigma'\sigma_\phi\sigma''$$

with σ'' free of σ_ϕ. Then $\lambda_\phi = \sigma\lambda_\phi = \sigma'\sigma_\phi\lambda_\phi$. By the theory of the Weyl group $\sigma'\sigma_\phi\alpha_\phi$ is a negative root since $\sigma_\phi\alpha_\phi=-\alpha_\phi$ and no more sign changes can occur if σ is written in minimal length. Since the Weyl group respects scalar products this furnishes a contradiction since it implies that $2\lambda_\phi.\alpha_\phi/\alpha_\phi^2$ is negative when by (2) it equals unity. Thus σ_ϕ cannot occur in σ and the result is established.

Since W(K) can be realised by gauge transformations in K[8] we can choose an element k' of K so that if $W''=k'W'$,

$$W''x.H(W'')^{-1} = x.H$$

Hence W'' lies in the maximal torus of K obtained by exponentiating its Cartan subalgebra. We can choose an element k" of this torus so that

$$W'''=k''W''=e^{2iy\,\alpha_\phi.H/(\alpha_\phi)^2}$$

Then by (2) $W'''|\lambda\rangle=|\lambda\rangle$ for any fundamental weight λ of G distinct from λ_ϕ, while,

$$W'''|n\lambda_\phi\rangle = e^{iny}|n\lambda_\phi\rangle \qquad (1.2.9)$$

But W''' is an element of H(n) and so ny must be an integer multiple of 2π. When n=1, W''' must equal unity as it assumes this value on all states of all irreducible representations of G. Thus H(1) equals K and the result (4) is established for n=1. Indeed (2) implies

$$e^{4\pi i \alpha_\phi . H/(\alpha_\phi)^2} = 1 \qquad (1.2.10$$

This is a special case of the homomorphism [8]

$$\mu \rightarrow e^{4\pi i \mu . H}$$

from the coweight lattice of G onto the centre of its universal covering group. The kernel is the coroot lattice. It is instructive to expand $\alpha_\phi/(\alpha_\phi)^2$ in terms of coweights of U(1) x K (eqn (3):

$$\alpha_\phi/(\alpha_\phi)^2 = z\lambda_\phi/(\alpha_\phi)^2 + \mu^\vee \qquad (1.2.1$$

where z is the order of the centre of G divided by the order of the centre of K and $\mu^\vee$ is a coweight of K and so perpendicular to λ_ϕ[4]. It follows from (10) and (11) that

$$V_0 = e^{4\pi i z \lambda_\phi . H/(\alpha_\phi)^2} = e^{-4\pi i \mu^\vee . H} \qquad (1.2.1$$

This is the element V_0 mentioned earlier which generates the Z_ℓ subgroup common to U(1) and the centre of K in (3) as is clear from (12). ℓ is the smallest integer such that $(V_0)^\ell$ equals unity.

The preceding analysis showed that any element of H(n) could be written as an element of K times an integer power of $(V_0)^{1/n}$ where

$$(V_0)^{1/n} = e^{4\pi i z \lambda_\phi . H/n(\alpha_\phi)^2}$$

This generates the $Z_{n\ell}$ subgroup commuting with K in (4). Integer powers of V_0, eqn. (12) lie in K and this is why the Z_ℓ subgroup generated by V_0 must be divided out in the result (4) which is thereby established.

Notice that the component $|n\lambda_\phi\rangle$ is necessarily a complex component of the representation whose highest weight is $n\lambda_\phi$. Sometimes this representation may be a real representation and so possess no complex components. Then it is understood that we are talking about the complex representations formed by a pair of these real representations. This is why we specified the complexified 27,000 and 1,463 of E_8 and E_7 respectively.

There is some reason to think that there is a sequence of grand unified groups E_8, E_7, E_6, $E_5 \cong SO(10)$, $E_4 \cong SU(5)$ broken by successive adjoint Higgs producing U(1) factors [9]. Each of these U(1)'s can be broken by the $D^{2\lambda_\phi}$ Higgs discussed above (possibly a difermion condensate) to an effective Z_2 subgroup. Then instead of monopoles associated with each such U(1) factor Z_2 strings can arise which survive the subsequent symmetry breaking of this type. For these strings to seed galaxy formation they must be superheavy as must therefore be the fermions acquiring mass from the same Higgs. This seems to favour the breaking of E_6 and SO(10) by this mechanism since it is only then that just the unwanted components acquire masses.

1.3 STRINGS IN THE EARLY UNIVERSE

There is as yet no fully satisfactory theory of galaxy formation [10]. Zeldovich [11] first suggested that strings could provide the density fluctuations needed to give rise to galaxies. His argument was as follows. Strings move at typical velocities of the order of the speed of light (as will be seen in subsequent sections) so if strings move freely, by crossing and exchange of partners and subsequent annihilation one might expect of the order of one length of string to cross each particle horizon - that is (letting c = 1)

$$\rho_{string} \sim \mu t / t^3 \sim \mu / t^2 \tag{1.3.1}$$

In a radiation dominated universe, $\rho \sim \rho_{radiation} \sim 1/30\ Gt^2$. So the density fluctuation due to strings is

$$\frac{\delta\rho}{\rho} = \frac{\rho_{string}}{\rho} \sim 30 G\mu \sim 3.10^{-3} \tag{1.3.2}$$

if $G\mu \sim 10^{-4}$. Recalling $\mu \sim |\langle\Phi\rangle|^2 \sim Tc^2$, where Tc is the temperature at which the phase transition occurs, and $G = Mp^{-2}$ where Mp is the planck mass (letting $\hbar = 1$), we see $G\mu \sim (Tc/Mp)^2$ so (1.3.2) holds for $Tc \sim 10^{-2}\ Mp \simeq 10^{17}$ GeV, higher than in Grand Unified theories.

Why do we require (1.3.2)? Density perturbations only start to grow after the decoupling of matter and radiation, at $t \sim 10^{12}$s. At this time the Jeans length, the minimum scale on which fluctuations can start to grow, falls to a value much less than the horizon distance and galaxies can begin to form [10]. After 10^{12}s, $\delta\rho/\rho$ grows like $t^{2/3}$, so at recent epochs, $t \sim 10^{16}$s, when it is thought galaxies were formed,

$$\frac{\delta\rho}{\rho} \sim \left(\frac{10^{16}}{10^{12}}\right)^{2/3} \cdot 3.10^{-3} \sim 1 \tag{1.3.3}$$

Vilenkin [12,13] improved the idea by suggesting that closed loops of string would give $\delta\rho/\rho \sim 3.10^{-3}$ for Tc $\sim$ 10^{15} GeV, close to the value in most Grand Unified Theories. He made the following key assumptions:

(i) that there exist long-lived spinning loop solutions,

(ii) that loops are formed at a rate $\frac{dn}{dt} \sim 1/t^4$ corresponding to one loop of radius $\sim$ t produced per unit time per horizon (itself of radius $\sim$ t)

(iii) that the loops lose energy primarily by gravitational radiation.

By assumption (iii) the loops lifetime can be estimated ; a loop of radius r oscillates with frequency $\omega \sim 1/r$ (we use c = 1 everywhere) and loses gravitational energy at a constant rate

$$\dot{M} \sim -GM^2\omega^2 \sim -\mu^2 G \tag{1.3.4}$$

Its lifetime τ is thus of the order of

$$\tau \sim (G\mu)^{-1} r_0 \sim 10^8 r_0 \tag{1.3.5}$$

where r_0 is its initial radius. The smallest loops surviving till a time t were thus formed at $\sim G\mu t$.

Estimating the density of loops requires knowledge of the exact behaviour of their mass energy (which is not conserved) after they are formed. A simple estimate is that it is simply equal to their mass at formation, of order $\mu t'$ where t' is the time of formation.

Then (ii) yields

$$\rho_{loops}(t) \sim \frac{1}{R^3(t)} \int_{G\mu t}^{t} \mu t' \, R^3(t') \frac{1}{t'^4} \, dt' \tag{1.3.6}$$

where the factors of R account for the universes expansion, and $R \propto t^{1/2}$ in a radiation dominated universe. This gives

$$\rho_{loops}(t) \sim \mu/(G\mu)^{1/2} t^2 \tag{1.3.7}$$

so

$$\frac{\rho_{loops}}{\rho} \sim 30\,(G\mu)^{1/2} \sim 3.10^{-3} \tag{1.3.8}$$

as desired. So under all these assumptions, loops could provide the right density perturbation amplitude to account for galaxy formation. Similarly the number of loops per particle horizon (whose radius is of order t) is given by

$$\begin{aligned} N_{loops} &\sim \frac{t^3}{R^3(t)} \int_{G\mu t}^{t} R^3(t') \frac{1}{t'^4} \, dt' \\ &\sim (G\mu)^{-3/2} \sim 10^{12} \end{aligned} \tag{1.3.9}$$

compared to the number of galaxies observed today, estimated at 10^{11}.

We shall examine assumptions (i)-(iii) in subsequent sections, and find some interesting results along the way.

1.4 THE PRODUCTION OF STRING LOOPS IN AN EXPANDING UNIVERSE

In this section we deal with the main assumptions underlying the string theory of galaxy formation. In particular, we ask 'do there exist long-lived loop solutions?' and 'can these be produced at a rate $\frac{dn}{dt} \sim \frac{1}{t^4}$?'

First we search for long-lived closed loops. A possible criticism of the lifetime estimate (1.3.5) based on assumption (iii) is that other types of radiation may in fact be more important. However, strings do not carry any net electric or magnetic charges so any such process must be rather indirect. For the large scale strings we shall consider the proper acceleration of the string, equal to 1/r where r is the radius of curvature of the string in its rest frame, is vastly less than the mass of any particle coupling to it, so (1.3.5) and assumption (iii) are probably correct.

There is, however, another problem. A closed loop may from time to time intersect itself and break into two smaller loops, whose lifetime would be shorter. At first sight it seems plausible to suppose that for any random initial configuration there is some probability p per oscillation cycle of self-intersection, and that p should be independent of the actual size of the loop. If that were true, one should expect that a loop of length ℓ would break into two loops of length $\ell/2$ in a typical time ℓ/p . These in turn would break into yet smaller loops of length $\ell/4$ in a further time $\ell/2p$, and so on. Obviously this yields a geometric series, giving a total lifetime of order $2\ell/p$. Unless p were astonishingly small, this would reduce the number of loops to a level much less than required by Vilenkin.

This is the argument we propose to test. It is by no means conclusive, for it is not clear that it is reasonable to approximate the behaviour of a large number of loops by that of an "average" loop. It may happen that there are some initial configurations that lead to rapid self-intersection, but others that never intersect at all. Indeed, this is precisely what we shall show does happen. The remaining unanswered questions is what proportion of strings fall into this category.

Let us first recall the dynamical equations of strings (which may be derived from an action integral proportional to the invariant area of the world sheet swept out by the string). As shown by Goddard et. al. [14], it is possible to choose the parameterization of the string so that the equation of motion takes a particularly simple form. If t is the time and s the length parameter along the string proportion to the total mass of the string from a fixed point, the equation for the position $\underset{\sim}{r}(s,t)$ reduces to

$$\ddot{\underset{\sim}{r}} - \underset{\sim}{r}'' = \underset{\sim}{0} \qquad (1.4.1)$$

where $\dot{\underset{\sim}{r}} = d\underset{\sim}{r}/dt$ and $\underset{\sim}{r}' = \partial\underset{\sim}{r}/\partial s$. These quantities also satisfy the constraint equations

$$\dot{\underset{\sim}{r}} \cdot \underset{\sim}{r}' = 0, \qquad \dot{\underset{\sim}{r}}^2 + \underset{\sim}{r}'^2 = 1.$$

Of course the general solution of (1) is

$$\underset{\sim}{r} = \tfrac{1}{2}\, \underset{\sim}{a}(s-t) + \tfrac{1}{2}\, \underset{\sim}{b}\,(s+t)$$

where $\underset{\sim}{a}$ and $\underset{\sim}{b}$ are subject to the constraints

$$\underset{\sim}{a}'^2 = \underset{\sim}{b}'^2 = 1.$$

For a closed loop of invariant length L both $\underset{\sim}{r}$ and (in the centre of mass frame) $\dot{\underset{\sim}{r}}$ must be periodic with period L and therefore we must have

$$\underset{\sim}{a}(s+L) = \underset{\sim}{a}(s) \ ,$$

$$\underset{\sim}{b}(s+L) = \underset{\sim}{b}(s) \ .$$

In fact the period of the motion is really L/2 rather than L, since it is easily seen that

$$\underset{\sim}{r}(s+L/2,\ t+L/2) = \underset{\sim}{r}(s,t).$$

It remains only to examine these solutions for self-intersection One may suppose that loops are often formed in an initially almost static configuration. First, therefore, let us consider the case of an initially static string, $\dot{\underset{\sim}{r}}(s,0) = \underset{\sim}{0}$. Then $\underset{\sim}{a}'(s) = \underset{\sim}{b}'(s)$, and by suitable choice of an arbitrary constant vector we may take $\underset{\sim}{a}(s) = \underset{\sim}{b}(s)$. A half-period later we have

$$\underset{\sim}{r}(s,L/4) = \tfrac{1}{2}\ \underset{\sim}{a}(s - L/4) + \tfrac{1}{2}\ \underset{\sim}{a}(s+L/4).$$

However,

$$\underset{\sim}{r}(s+L/2,L/4) = \tfrac{1}{2}\ \underset{\sim}{a}(s+L/4) + \tfrac{1}{2}\ \underset{\sim}{a}\ (s+3L/4).$$

By the periodicity of a these are equal. Thus we have the remarkable result that any initially static string not merely self-intersects but actually collapses to a doubled loop after a half period. Presumably strings would then annihilate into particles. For other strings more complex process might occur.

At first sight this result might suggest that it is difficult to find non-self-intersecting strings, but in fact this is not the case. Indded there exists initial configurations that differ infinitesimally from a static one but which do not lead to self-intersection. Here we merely exhibit some simple examples.

For simplicity let us take the length L of the string to be 2π . The simplest type of solution for $\underset{\sim}{a}$ (or $\underset{\sim}{b}$) is given by

$$\underset{\sim}{a}'(s) = \underset{\sim}{e}_1 \cos s + \underset{\sim}{e}_2 \sin s \tag{1.4.2}$$

where $\underset{\sim}{e}_1$ and $\underset{\sim}{e}_2$ are orthogonal unit vectors. If both $\underset{\sim}{a}'$ and $\underset{\sim}{b}'$ have this form, we can choose the axis $\underset{\sim}{e}_1$ in the direction of the intersection of the circles traced out by $\underset{\sim}{a}'$ and $\underset{\sim}{b}'$ and by suitable choice of the zero of s write $\underset{\sim}{a}'$ as in (2), and $\underset{\sim}{b}'$ as

$$\underset{\sim}{b}'(s) = \underset{\sim}{e}_1 \cos s + (\underset{\sim}{e}_2 \cos\phi + \underset{\sim}{e}_3 \sin\phi) \sin s.$$

The corresponding solution is no longer initially static, but it is easy to check that

$$\underset{\sim}{r}(\pi-s,\pi/2) = \underset{\sim}{r}(s,\pi/2).$$

Hence the string collapses to a line after a half period.

It is easy to check that adding to (1.4.2) terms in cos2s and sin2s does not yield any extra solutions. The next simplest solutions involve cos3s and sin3s. (The only other possible single frequency addition is one containing cos5s and sin5s terms). By suitable choice of the zero point they may be written

$$\begin{aligned} \underset{\sim}{a}'(s) = \; & \underset{\sim}{e}_1\left[(1-\alpha) \cos s + \alpha\cos 3s\right] \\ + \; & \underset{\sim}{e}_2\left[(1-\alpha) \sin s + \alpha\sin 3s\right] \\ + \; & \underset{\sim}{e}_3 2\left[\alpha(1-\alpha)\right]^{\frac{1}{2}} \sin s \, , \end{aligned}$$

where α is an arbitrary parameter between 0 and 1. Taking $\underset{\sim}{a}'$ of this form and $\underset{\sim}{b}'$ as in (2) we find a perturbation of the initially-static string solution, namely

$$\underset{\sim}{r}(s,t) = \tfrac{1}{2}\, \underset{\sim}{e}_1 \Big[(1-\alpha)\sin s_- + \tfrac{1}{3}\alpha \sin 3s_- + \sin s_+\Big]$$

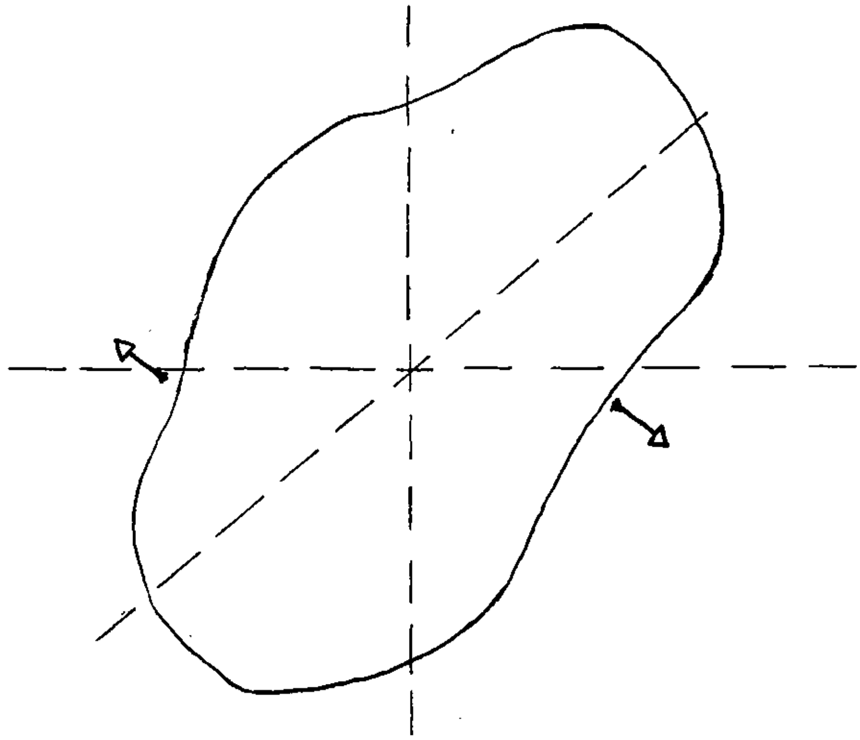
$$- \tfrac{1}{2}\, \underset{\sim}{e}_2 \Big[(1-\alpha)\cos s_- + \tfrac{1}{3}\alpha \cos 3\, s_- + \cos s_+\Big]$$

$$- \quad \underset{\sim}{e}_3 \big[\alpha(1-\alpha)\big]^{\frac{1}{2}} \cos s_- \tag{1.4.3}$$

with $s_\pm = s \pm t$. It is then straightforward to check that this string does not intersect itself for $0 < \alpha < 1$, i.e. there is no non-trivial solution of the equation $\underset{\sim}{r}(s,t) = \underset{\sim}{r}(s',t)$.

What is particularly noteworthy about this solution is that only a small perturbation away from one of the collapsing solutions is sufficient to make it non-self-intersecting. This suggests that there is in fact a large class of stably-oscillating solutions, so that the loops required by Vilenkin's scenario may indeed have long lifetimes.

There are of course several questions still unanswered. It has to be shown that other modes of radiation of strings do not lead to rapid decay of loops. Also it is not clear how special are the solutions we have found. It will be necessary to examine other perturbations of initially-static strings to determine whether the avoidance of self-intersection is a general feature or a fortuitous result of our special choice. Ideally one would wish to estimate what proportion of loops formed by self-intersections of long strings would be in a self-avoiding configuration. This is at present hard to do, but we have at least shown that the class is non-empty.



A STABLY SPINNING LOOP

Next we shall deal with the second major assumption, that the number of loops per unit volume, n, produced by the self-intersection and exchange of partners of lengths of string stretched by the expansion of the universe obeys

$$\frac{dn}{dt} \sim \frac{1}{t^4}$$

Obviously this equation is crucial to the whole scenario. Previously only a consistency argument (which will be discussed below) was given.[4] Here we shall show that an analysis of the behavior of exact solutions for waves on lengths of string yields precisely this result.

The consistency argument may be stated as follows. Strings would initially be formed[2] in a tangled Brownian configuration of persistence length of the order of the correlation length $\sim T_c^{-1}$. There would be around 10^4 lengths of string formed across each horizon. These would be conformally stretched by the expansion of the universe on scales larger than the horizon. If there were no mechanism for energy loss, the string density ρ_s would (ignoring subtleties which will be discussed below) scale as $\frac{1}{R^2}$, where R is the scale factor, compared to the radiation density scaling as $\frac{1}{R^4}$, and soon come to dominate the total energy density of the universe. In order for this not to happen, there must be some mechanism for decreasing the string density so that $\rho_s \propto \frac{1}{R^4}$ instead. In the radiation dominated very early universe, $R \propto t^{1/2}$ so, since no mechanism can operate on scales larger than the horizon, we assume one acts very efficiently inside the horizon, yielding $\rho_s \sim \frac{\mu}{t^2}$ corresponding to one length of string stretched across each horizon, where μ is the mass per unit length of the string and $\mu \sim T_c^{\,2}$.[2]

This will indeed be the case if there are $\sim \frac{1}{t^4}$ loops formed per unit volume per unit time of radius of order the horizon distance, and if these loops can then collapse and annihilate, gravitationally couple to the surrounding fluid and create turbulence or radiate away via gravitational radiation.

Let us see how this can happen. We consider a Friedmann-Robertson-Walker universe, with metric

$$ds^2 = dt^2 - R^2(t)d\underline{x}^2$$

Just as the action for a particle is proportional to the length of its world line, that for a string is proportional to the area of the world sheet it sweeps out:[14]

$$S = \mu\int dA = \mu\int d\sigma d\tau \sqrt{\left(\frac{\partial x}{\partial\sigma}\cdot\frac{\partial x}{\partial\tau}\right)^2 - \left(\frac{\partial x}{\partial\sigma}\right)^2\left(\frac{\partial x}{\partial\tau}\right)^2}$$

where $x^\mu(\sigma,\tau)$ are the space-time coordinates and σ,τ the parameters describing the sheet surface. We simplify the equations of motion by choosing a new time coordinate η such that $dt = Rd\eta$ and the metric becomes

$$ds^2 = R^2(d\eta^2 - d\underline{x}^2) \qquad (1.4.4)$$

Then we choose $\tau = x^o = \eta$ and define σ so that

$$\dot{\underline{x}}\cdot\underline{x}' = 0 \qquad (1.4.5)$$

for all η, where $\dot{\underline{x}} = \frac{\partial \underline{x}}{\partial \eta}$ and $\underline{x}' = \frac{\partial \underline{x}}{\partial \sigma}$. In this way we obtain the equation of motion

$$\frac{\partial}{\partial \eta}\left(\frac{\underline{x}'^2 \dot{\underline{x}}}{\sqrt{\underline{x}'^2(1-\dot{\underline{x}}^2)}}\right) + 2\,\frac{\dot{R}}{R}\left(\frac{\underline{x}'^2 \dot{\underline{x}}}{\sqrt{\underline{x}'^2(1-\dot{\underline{x}}^2)}}\right) = \frac{\partial}{\partial \sigma}\left(\frac{\underline{x}'(1-\dot{\underline{x}}^2)}{\sqrt{\underline{x}'^2(1-\dot{\underline{x}}^2)}}\right) \tag{1.4.6}$$

If the second (damping) term is small we can (by rescaling σ) apply the extra constraint

$$\dot{\underline{x}}^2 + \underline{x}'^2 = 1 \tag{1.4.7}$$

upon which Eq(1.4.7) becomes simply the wave equation

$$\ddot{\underline{x}} = \underline{x}'' \tag{1.4.8}$$

and Eq(1.4.7) is easily checked to be preserved under Eq.(1.4.8)

When is this possible? A general wave of coordinate amplitude a and wavelength λ yields $|\underline{x}''| \sim a(\frac{2\pi}{\lambda})^2$. Using Eq(1.4.7) we find that damping is small in Eq(1.4.6) if (since $|\dot{\underline{x}}|$ must be less than 1)

$$\frac{2\dot{R}}{R}\,(\frac{\lambda}{2\pi})^2\,\frac{1}{a} << 1$$

or, if $R \sim \eta^k$ ($k = 1$ in a radiation-dominated universe),

$$\frac{\Lambda^2}{A} < h \tag{1.4.9}$$

where h is the horizon distance, $h \sim t$ and A, Λ are the proper amplitude and wavelength of the wave respectively.

For Brownian strings we expect $\Lambda \sim A$ on scales larger than the persisten length and see that when A is larger than the horizon, the motion is heavily 'damped' i.e. $|\dot{\underline{x}}| << 1$ and the strings are conformally stretched, $A \sim \Lambda \propto R(t)$ and move slowly with respect to the surrounding matter. The horizon distanc $h \sim R\eta \sim t$ grows faster - when it catches up with A damping becomes small. Thereafter Eq.(1.4.9) applies and (as we shall see) large amplitude waves of constant comoving amplitude a (i.e. increasing proper amplitude $A = Ra$) can propagate inside the horizon.

Notice that small amplitude waves behave differently. They never obey Eq.(1.4.9) and damping is always important. Neglecting terms in a^2 we can write Eq.(1.4.6) for a small amplitude wave as

$$\ddot{\underline{x}} + \frac{2\dot{R}}{R}\dot{\underline{x}} = \underline{x}'' \tag{1.4.10}$$

If $R \sim \eta$ this becomes

$$(\eta \underline{x})^{\cdot\cdot} = (\eta \underline{x})'' \tag{1.4.11}$$

which has oscillating waves with $a \sim \frac{1}{\eta}$ as solutions.[13]

We see that waves of different structure show very different behaviour - for $A \sim \Lambda$ there are constant comoving amplitude solutions for which the energy per wavelength increases like R:

$$E = \mu \int \frac{R d\ell}{\sqrt{1-\dot{\underline{x}}_\perp^2}} = \mu R \int d\sigma \tag{1.4.12}$$

where $d\ell$ is the length element along the string, $\dot{\underline{x}}_\perp$ the perpendicular velocity and the last equality holds only if Eqs.(1.4.5) and (1.4.7) do. As Vilenkin [13] showed, for small amplitude waves, with $A \ll \Lambda$, the energy per wavelength increas like R for $\Lambda > h$ and decreases like $\frac{1}{R}$ after they fall inside the horizon. However he assumed that this result applied for Brownian strings, which it clearly does not, and used it to argue that the equation of state for these should be $\rho \propto \frac{1}{R^4}$ even without any loop formation. Our analysis shows the different behaviour of small and large-amplitude waves. If the latter occur (as they do in Brownian configurations), we are led inescapably back to the view that the formation of closed loops is crucial to the consistency of the string picture.

Let us see how these loops might be formed from Brownian waves ($A \sim \Lambda$). When a wave of given amplitude falls inside the horizon it starts moving freely. In general its motion will not be correlated over larger scales and we expect moving waves (in so far as they can be defined) to meet inside the horizon. A typical simple wave solution is

$$\underline{x}_1(\sigma,\eta) = \frac{a}{2}\,[\delta\tilde{\sigma}_-, \gamma\cos\tilde{\sigma}_- + \sin\tilde{\sigma}_+,\ \gamma\sin\tilde{\sigma}_- - \cos\tilde{\sigma}_+]\quad , \delta^2 + \gamma^2 = 1 \tag{1.4.13}$$

where $\tilde{\sigma}_{\pm} = \frac{1}{a}\,(\sigma \pm \eta)$ and a is the amplitude (and $\lambda = 2\pi a$). We have chosen for simplicity a wave with phase velocity v the speed of light - in genera a wave can have $-\infty < v < \infty$. Consider a situation where such a wave travellin along the x axis meets a similar one travelling in the opposite direction at $\eta = 0$. We shall add a frequency three component (frequency two is not allowed by the constraints (Eq.1.4.5,7) to this latter wave, which thus takes the form

$$\begin{aligned}\underline{x}_2(\sigma,\eta) = \frac{a}{2}\,[&\beta\tilde{\sigma}_+ - 2\sqrt{\alpha(1-\alpha)}\,\cos\tilde{\sigma}_-\ ,\\ &\sqrt{1-\beta^2}\,\cos\tilde{\sigma}_+ + (1-\alpha)\sin\tilde{\sigma}_- + \frac{\alpha}{3}\sin 3\tilde{\sigma}_-\ ,\\ &\sqrt{1-\beta^2}\,\sin\tilde{\sigma}_+ - (1-\alpha)\cos\tilde{\sigma}_- - \frac{\alpha}{3}\cos 3\tilde{\sigma}_-]\end{aligned} \tag{1.4.14}$$

We set

$$\begin{aligned}\underline{x},\dot{\underline{x}}(\sigma,0) &= \underline{x}_1,\dot{\underline{x}}_1(\sigma,0) \qquad \sigma \geq 0\\ &= \underline{x}_2 + \underline{c},\dot{\underline{x}}_2(\sigma,0) \qquad \sigma \leq 0\end{aligned} \tag{1.4.15}$$

where

$$\underline{c} = \frac{a}{2}[2\sqrt{\alpha(1-\alpha)}\quad ,\quad \gamma - \sqrt{1-\beta^2}\quad ,\quad -\frac{2\alpha}{3}]$$

is chosen to satisfy continuity. Then for $\eta > 0$ the solution is easily found to be

$$\begin{aligned}\underline{x}(\sigma,\eta) &= \underline{x}_1(\sigma,\eta) \qquad\qquad \sigma \geq \eta\\ &= \frac{1}{2}[2\sqrt{\alpha(1-\alpha)}\ ,\quad \gamma,\quad 1 - \frac{2\alpha}{3}\,]\\ &\quad + \frac{a}{2}[-2\sqrt{\alpha(1-\alpha)}\cos\tilde{\sigma}_-, \sin\tilde{\sigma}_+ + (1-\alpha)\sin\tilde{\sigma}_- + \frac{\alpha}{3}\sin 3\tilde{\sigma}_-\,,\\ &\quad -\cos\tilde{\sigma}_+ - (1-\alpha)\cos\tilde{\sigma}_- - \frac{\alpha}{3}\cos 3\tilde{\sigma}_-] \quad -\eta \leq \sigma \leq \eta\\ &= \underline{x}_2(\sigma,\eta) + \underline{c}\quad . \qquad\qquad \sigma \leq -\eta\end{aligned} \tag{1.4.16}$$

As η increases from 0 the solution winds out a loop. At $\eta = a\pi$ we find $\underline{x}(a\pi,-a\pi) = \underline{x}(-a\pi,a\pi) = \underline{x}(0,0)$ so the string meets itself at the origina meeting point. There are now two possibilities. The string could continue to wind out another loop exactly alongside the original one (since $\underline{x}_1(a\pi+\sigma,\ a\pi+\eta) = \underline{x}_1(\sigma,\eta)$ and $\underline{x}_2(-a\pi+\sigma,\ a\pi+\eta) = \underline{x}_2(\sigma,\eta)$ we are exactly, modulo the loop, in the original configuration). With a small perturbation however, the string will not merely touch but cross itself at right angles, and an exchange of partners (see below) might occur. This would result in a closed loop being formed, incidentally precisely the one shown in Ref. 5, never to intersect itself. Subsequent waves meeting on the string are likely to have larger amplitude (having been stretched longer) and so would produce a loop of larger size. We see that this process produces one loop of (coordinate) dimension a in coordinate time $a\pi$, or one loop of proper dimension t in a time $\sim t$, contributing

$$\frac{dn}{dt} \sim \frac{1}{t^4}$$

if there is one such process going on per horizon.

The above analysis is very approximate. Waves may move apart as well as collide, and we have only examined a special class. However by varying the parameters (e.g. v) it is easy to convince oneself that the production of loops is a general feature of colliding waves. Indeed there is really little else that can occur as the string piles up at the center of mass.

It is also clear that the simplest frequency contributions to the fourier spectrum of the solutions should be most important: as we have explained, in an expanding universe there is a cut-off as far as propagating waves are concerned at $\Lambda \sim t$, as longer wavelengths are damped. On scales larger than t, the initially Brownian strings have been conformally stretched so $A(\Lambda) \sim \Lambda$. As Λ falls inside the horizon waves begin to move freely and thereafter energy loss processes - annihilation, collision or gravitational radiation occur. Since the energy in a wave (like Eq. 1.4.13) is proportional to A/Λ, this means $A(\Lambda)$ is less than Λ is $\Lambda < t$ falling to zero at $\Lambda = 0$. Thus our simple examples are likely to be the most important contributors to the overall processes.

A major uncertainty remains - do strings really exchange partners when they cross, or do they simply pass through one another?

The magnetic flux component along the string is repulsive, and the repulsion is proportional to $\cos\theta$, where θ is the angle between the strings. However the higgs component is attractive, and independent of angle. Thus perpendicular strings can interpenetrate one anothers cores. Their attraction may then hold them together and allow them to exchange partners. Almost parallel strings would find it more difficult. There is some experimental evidence that such an exchange of partners does occur in flux lines in Type II superconductors,[15] where it is called "flux-cutting". If one takes estimates of the higgs mass in GUTS seriously, then it is less than the gauge field mass, and yields a Type I vacuum structure. That is, strings are attractive over a longer range $\frac{1}{m_H}$, and have a repulsive core of radius $\frac{1}{m_X}$, where m_H and m_X are the higgs and gauge field masses respectively. One might imagine flux cutting occuring more easily in these circumstances.

1.5 THE EVOLUTION OF COSMIC DENSITY PERTURBATIONS AROUND STRINGS

Here we shall follow a more direct approach than in Section 1.3 to the problem of galaxy formation by seeing whether each galaxy could have been the result of a single spinning loop of string. We shall calculate the gravitational field of a spinning loop or periodic wave in the weak field approximation in Minkowski spacetime, average it over time, and then use the result to calculate the growth of density fluctuations around moving string. This is valid provided the period of the loop (or its radius) or wave is much less than the expansion time (or horizon distance). The result is strikingly simple and shows that galaxies might well be formed today in the correct numbers by gravitational accretion around loops. We shall also see how linear structure may develop around lengths of string.

First we need the energy momentum tensor of the string. In its rest frame a segment $d\ell$ located at $\underline{r}$ contributes

$$\begin{aligned} dT_{00}(\underline{x}) &= \mu\, d\ell\, \delta^3(\underline{x}-\underline{r}) \\ dT_{33}(\underline{x}) &= \mu\, d\ell\, \delta^3(\underline{x}-\underline{r}) \end{aligned} \tag{1.5.1}$$

with all other components zero where $\mu \sim 10^{19}\,\mathrm{kg\,m^{-1}}$ for $T_c \sim 10^{15}\,\mathrm{GeV}$

This may be boosted to an observers frame in which the element has a velocity $\dot{\underline{r}}$ to obtain

$$\begin{aligned} dT_{00}(\underline{x},t) &= \mu\, d\sigma\, \delta^3(\underline{x}-\underline{r}(\sigma,t)) \\ dT^{\lambda}{}_{\lambda}(\underline{x},t) &= 2\mu/\gamma^2\, d\sigma\, \delta^3(\underline{x}-\underline{r}(\sigma,t)) \end{aligned} \tag{1.5.2}$$

wehre $d\sigma = \gamma d\ell$, $\gamma = (1-\dot{\underline{r}}^2)^{-1/2}$, $\dot{\underline{r}} = \frac{\partial \underline{r}}{\partial t}$ and $\underline{r}(\sigma,t)$ describes the trajectory of the string, its length being parametrised by σ. Its energy is thus given by

$$E = \mu \int d\sigma \tag{1.5.3}$$

In the weak field approximation, we write $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, $\eta_{\mu\nu} =$ diag (1, -1, -1, -1) and $|h_{\mu\nu}| \ll 1$. In the harmonic gauge Einstein's equations become

$$\partial^2 h_{\mu\nu} = -16\pi G \left(T_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu} T^{\lambda}{}_{\lambda}\right) \tag{1.5.4}$$

and we obtain the retarded solution to (1.5.2) and (1.5.4)

$$h_{00}(\underset{\sim}{x},t) = -4G\mu \int d\sigma \frac{\dot{\underset{\sim}{r}}^2(\sigma, t^{ret})}{|\underset{\sim}{x} - \underset{\sim}{r}(\sigma, t^{ret})|} \frac{1}{(1 - \dot{\underset{\sim}{r}}\cdot\underset{\sim}{\epsilon})} \tag{1.5.5}$$

where

$$\begin{aligned} t^{ret} &= t - |\underset{\sim}{x} - \underset{\sim}{r}(\sigma, t^{ret})| \\ \underset{\sim}{\epsilon} &= \frac{\underset{\sim}{x} - \underset{\sim}{r}(\sigma, t^{ret})}{|\underset{\sim}{x} - \underset{\sim}{r}(\sigma, t^{ret})|} \end{aligned} \tag{1.5.6}$$

and we have used the δ function to do the three dimensional integrals Notice that a static string produces no field (in this case, $\frac{1}{2}h_{00}$ is the Newtonian potential), a result first obtained by Vilenkin[16].

The force on nonrelativistic particles around the string is given by

$$\frac{d^2x^i}{dt^2} + \Gamma^i_{00} = 0 \quad ; \quad \Gamma^i_{00} = -\tfrac{1}{2}\left(2h_{0i,0} - h_{00,i}\right) \tag{1.5.7}$$

We shall time average (1.5.7) and so drop the time derivative term. For time averaging we use

$$dt = dt^{ret}(1 - \underset{\sim}{\epsilon}\cdot\dot{\underset{\sim}{r}}) \tag{1.5.8}$$

and calculate from (1.5.5) and (1.5.8)

$$\overline{h_{00}}(\underset{\sim}{x}) = -\frac{4G\mu}{T}\int_0^T dt \int d\sigma \frac{\dot{\underset{\sim}{r}}^2(\sigma,t)}{|\underset{\sim}{x} - \underset{\sim}{r}(\sigma,t)|} \tag{1.5.9}$$

where T is the period of the strings motion. The result is simple - the time-averaged field of a string is equal to that of the surface traced out by its motion with surface density proportional to $\dot{\underset{\sim}{r}}^2(\sigma,t)$.

In fact the total mass of this shell may be calculated. Using the equations for string

$$\ddot{\underline{\Sigma}} = \underline{\Sigma}'' , \quad \dot{\underline{\Sigma}}^2 + \underline{\Sigma}'^2 = 1 \tag{1.5.10}$$

where $\dot{\underline{\Sigma}} = \frac{\partial \underline{\Sigma}}{\partial t}$, $\underline{\Sigma}' = \frac{\partial \underline{\Sigma}}{\partial \sigma}$, we find the gravitationally effective mass in (1.5.9) to be, using (1.5.10) and integration by parts

$$\begin{aligned} \delta M_0 &= 2\mu \int d\sigma \int \frac{dt}{T} \dot{\underline{\Sigma}}^2(\sigma,t) \\ &= 2\mu\left(\int d\sigma - \int d\sigma \int \frac{dt}{T} \underline{\Sigma}'^2(\sigma,t)\right) \\ &= 2\mu\left(\int d\sigma + \int d\sigma \int \frac{dt}{T} \underline{\Sigma}''\cdot\underline{\Sigma}(\sigma,t)\right) \\ &= 2\mu\left(\int d\sigma + \int d\sigma \int \frac{dt}{T} \ddot{\underline{\Sigma}}\cdot\underline{\Sigma}(\sigma,t)\right) \\ &= 2\mu \int d\sigma - \delta M_0 \end{aligned} \tag{1.5.11}$$

So‡

$$\delta M_0 = \mu \int d\sigma \tag{1.5.12}$$

which is simply the total mass of the string.

This resolves certain paradoxes - if a static loop with no field collapses and passes through its Schwarzschild radius, it must form a black hole, with a definite field. The above analysis shows that the long range time-averaged field of an oscillating loop will be exactly that of a black hole the loop may form at some stage if it happens to pass through its Schwarzschild radius.

Some interesting features of the field calculated from (1.5.5) are the following; there is intense beaming of the gravitational field

‡ The boundary terms in (11) obviously vanish for a closed loop. They also cancel exactly for a periodic wave with unit phase velocity as in (1.4.13) . In general however they must be included - for periodic waves we find $0 < \delta M_0 < 2\mu\int d\sigma$, 0 being the result in the static limit.

produced by several powers of $(1-\dot{\underline{r}}\cdot\underline{\epsilon})^{-1}$ in the direction of motion of the string. Parts of the string may approach the velocity of light (for the solution (1.4.3), at $t=\pi/2$ and $\sigma=\pm\pi/2$) —this seems to favour the growth of two-armed structures about such strings, such as spiral galaxies. For near circular solutions, with very low angular momentum, which collapse to a small loop moving almost at the speed of light but expand outwards again, the string produces a disc-shaped field. This would give rise to oblate spheroidal structures of high density such as elliptical galaxies.

Here we confine ourselves to an overall estimate of the growth in mass of a density fluctuation around a loop after the decoupling time. Before this time the pressure is too high for galactic-scale fluctuations to grow. After this time, however, the effects of pressure may be neglected.[10] Obviously near the string the weak field, Newtonian approximation is invalid so we work on a comoving spherical surface surrounding the loop on which the velocity and density fluctuations about the mean values of the surrounding matter are small.

We work in an expanding Einstein-de Sitter Universe with metric

$$ds^2 = dt^2 - R^2(t)\,d\underline{x}^2 \tag{1.5.13}$$

and describe the matter around the string as a collisionless Newtonian fluid. We may neglect spatial curvature effects for times less that 10^{16}s or[10]. On the surface of our comoving shell, the relevant equations of motion of the fluid are

$$\frac{\partial M}{\partial t} = -\rho\int_S \underline{v}\cdot d\underline{S} \;;\quad \frac{\partial \underline{v}}{\partial t} + (\underline{v}\cdot\underline{\nabla})\underline{v} = -\underline{\nabla}\phi \tag{1.5.14}$$

where M is the total mass in the shell, $\underline{v}$ the fluid velocity, ρ its density and ϕ the gravitational potential. We set $M = M_0 + \delta M$, $\rho = \rho_0 + \delta\rho$ $\underline{v} = \underline{v}_0 + \delta\underline{v}$ and $\phi = \phi_0 + \delta\phi$ with the suffix zero labelling the unperturbed values $\underline{v}_0 = \dot{R}/R\,\underline{r}$ where $\underline{r}$ is the position of the shell, and $\rho_0 = 1/6\pi G t^2$ in a matter dominated universe. Then (1.5.14) give to first order in small quantities

$$\begin{aligned} \frac{\partial}{\partial t}\delta M &= -\rho_0 \int_S \delta\underline{v}.d\underline{S} \\ \frac{\partial}{\partial t}\delta\underline{v} + \dot{R}/R\,\delta\underline{v} + (\underline{v}_0.\underline{\nabla})\delta\underline{v} &= -\underline{\nabla}\delta\phi \end{aligned} \tag{1.5.15}$$

We integrate the second equation over S, on which $\underline{v}_0 = 0$, and differentiate the first to get

$$\frac{\partial^2 M}{\partial t^2} = \dot{\rho}_0/\rho_0 \frac{\partial \delta M}{\partial t} - \dot{R}/R \frac{\partial \delta M}{\partial t} + \rho_0 . 4\pi G \delta M - \rho_0 \int \delta\underline{v}.d\dot{\underline{S}}$$

using $\nabla^2\delta\phi = 4\pi G\delta\rho$. Finally, $d\dot{\underline{S}} = 2\dot{R}/R\,d\underline{S}$ and we find

$$\frac{\partial^2 \delta M}{\partial t^2} + \frac{4}{3t}\frac{\partial \delta M}{\partial t} - \frac{2}{3t^2}\delta M = 0 \tag{1.5.16}$$

with solutions $\delta M \propto t^{2/3}$ or t^{-1}. The growing solution yields

$$\delta M(t) = \delta M_0 (t/t_0)^{2/3} \tag{1.5.17}$$

where δM_0 is the initial value, equal to the mass of the string loop. Setting $t_0 \sim 10^{12}$s, the decoupling time, and $t_g \sim 10^{16}$s the time galaxies are thought to have formed, we find

$$\frac{\delta M}{M_0}(t_g) \sim 1 \tag{1.5.18}$$

where $M_0 = \frac{4}{3}\pi r_0^3 \rho(t_0) = \frac{2}{9}\frac{r_0}{G}\left(\frac{r_0}{t_0}\right)^2$ is the mass of a comoving sphere containing the loop, r_0 its radius at decoupling and $\delta M_0 \sim 2\pi\mu r_0$ its mass, if

$$r_0 \sim 10^{-2} t_0 \tag{1.5.19}$$

i.e. the loops radius is around one hundredth of the horizon distance at decoupling. This means it was formed at around 10^{10} seconds. The mass within a horizon was around $10^{17} M_{\odot}$ at decoupling so each of these loops would contain a mass of around $10^{11} M_{\odot}$ which is of course the size of a typical galaxy.

There would also be many smaller loops around at decoupling. These would contain smaller masses and would produce faster growing density inhomogeneities on smaller scales. The smaller number of loops larger than $10^{-2} t_0$ could presumably give rise to clusters of galaxies.

Above we assumed the mass of the loop to be constant. In reality however the loop slowly loses energy via graviational radiation. This gives it a lifetime of around $r/G\mu \sim 10^{8} r$ where r is its original radius. Loops formed at $10^{10}s$ might still be around, and would have radii of a hundred parsecs or so. Their effects would presumably still be visible in galactic cores. Indeed it is tempting to speculate that the very large internal sources of energy $(\sim 10^{53}$ or $10^{54} J)$ required to account for the intense radio emission from elliptical and Seyfert galaxies and quasars, could be provided by string loops. A loop of radius $10^{10}s$, for example has an energy of $10^{55} J$.

The behaviour of string after decoupling would be different - growing density perturbations could themselves bind gravitationally to lengths of string. Recall that whilst static lengths of string produce no field, periodic waves produce a mass per unit length of exactly μ and thus couple to the masses around the string. It is therefore unlikely that many loops would be formed after decoupling - the strings motion would be heavily weighed down by surrounding proto-galaxies.

An estimate of the extent of this effect is obtained by comparing the gravitational potential produced by the string to that produced by surrounding galaxies - the strings gravitational field would be strong enough to drag galaxies along with it out to a radius as given by

$$a \lesssim \sqrt{\mu/\pi\rho} \sim 0.7\, \Omega_0^{-1/2} \left(H_0/75\, \mathrm{km\, s^{-1} Mpc^{-1}}\right)^{-1} \mathrm{Mpc} \tag{1.5.20}$$

using $\rho = \Omega_0 \rho_c = \Omega_0 . 3H_0^2/8\pi G$. These long lengths of strings would move very slowly, the linear density of proto-galaxies in the universe exceeding that of string by a factor of 10^4 or so just after decoupling. Long wavelengths would thus tend to straighten out and would not move freely. This would lead to the occurence of filamentary structures on very large scales, with radius given by (1.5.20). Such structures have recently been observed[17] although their statistical significance seems at present difficult to assess.

According to the above picture, the motion of string on very large scales would be small, and would cause very little inhomogeneity of the microwave background on large scales (at present observational limits are $\delta\rho/\rho)_r < 3.10^{-4}$.

In this it compares favourably with the 'pancake' model of Zel'dovich and collaborators[18] which seems to require $\delta\rho/\rho)_r \sim 10^{-3}$ for galaxies to be formed at all[19].

If no more loops are produced after decoupling, then the string density scales as R^{-2} , more slowly than the matter density. It eventually comes to dominate at $t \sim 10^{22}$s - the onset of the string dominated era.

In conclusion, the string theory of galaxy formation has much to offer. N-body simulations of stars or galaxies moving around lengths or loops of string would give more concrete predictions as to the types of structures produced.

1.6 BARYON NUMBER FROM COLLAPSING COSMIC STRINGS

In section 1.4 we showed that any initially static loop collapses to a point or, more generally, a doubled loop (a configuration in which the string winds twice around the same loop) after a time L/4, where L is its initial length (and c=1). Z_2 strings would then presumably annihilate into particles. Z_3 strings might 'add' to form a string in the opposite direction with some release of particles. For other strings yet more complex processes could occur. We also showed that the lowest frequency mode non-static loop collapsed to a line at $\leq L/4$. Presumably such loops would simply annihilate into particles.

The purpose of this section is to show that these collapsing loops may also be a positive feature. The superheavy bosons released as the string annihilates would decay with some CP and baryon-number violation, as in the standard mechanism for generating baryon number in GUTs. This would occur well after the grand unification phase transition at $T_c \sim 10^{15}$ GeV. It is obviously an irreversible process, so that the requirement that the system be out of thermal equilibrium is automatically satisfied.

Strings formed at this transition are initially heavily damped by the surrounding matter [2]. A segment of radius of curvature r experiences an accelerating force μ/r per unit length, where $\mu \sim T_c^{\,2}$ is the string tension. This is opposed by the damping force $\sigma\rho v$, where ρ is the matter density, v the velocity of the string and σ the cross-sectional width for string-particle scattering which as Everett [20] has shown is very roughly (neglecting a logarithmic factor) of order 1/T. Thus the string reaches a terminal velocity $v_{ter} \sim \mu/\sigma\rho r$ Any kinks in a loop of string will tend to straighten out as it collapses (since $v_{ter} \propto 1/r$) and it will become roughly circular before disappearing, giving rise to very few superheavy bosons.

However this period of strong damping is quite brief. The largest relevant loops at time t are those of radius r ~ t. For these to acquire relaivistic speeds, we require $\mu/\sigma_\rho t \sim 1$, which occurs at

$$t = t_e \sim \frac{1}{30}\frac{m_p^3}{\mu^2} \sim 3.10^{-29}s$$

corresponding to a temperature of around 10^{11} GeV.

A more careful relativistic analysis defining t_e to be time at which energy loss through damping becomes small compared to the initial mass energy yields essentially the same result [20].

After this time, damping is negligible and strings move more or less freely. Collapsing loops may be formed in two ways - either when already existing loops enter the horizon, or by the self-intersection of longer strings. It is not at all clear what fraction of these may be formed in initially static or near-static configurations or in other collapsing configurations, but here we shall make what seems to be a not unreasonable assumption that the fraction is significant. These loops will then collapse to a doubled loop, a line or in special cases a point. These last will form black holes, while the others (at least in the case of Z_2 strings) annihilate and release their constitutuent boson quanta, both gauge bosons and Higgs particles. A rough estimate of the number of superheavy bosons released per unit invariant length of string is simply μ/m_X .

As argued above, the rate of creation of loops per unit volume is

$$\frac{dn}{dt} \sim 1/t^4$$

A typical loop created at time t has a radius of order t and therefore collapses, as explained above, at or before a time $t(1+\gamma/4)$ where γt is its length ($\gamma \sim 2\pi$). If it collapses to a doubled loop it gives rise to a net baryon number of order $\varepsilon\gamma t\mu/m_X$, where ε is the mean net baryon number produced in the decay of a superheavy boson.

We can now estimate the total baryon asymmetry produced by all loops collapsing after the damping period. The entropy density is

$$s = \frac{2\pi^2}{45} N T^3 \simeq 50 T^3$$

During the expansion the ratio n_B/s is constant except for the contribution from collapsing loops (or other baryon number-generating processes). Since the number of bosons decaying is small compared to the number of particles already present, their contribution to the entropy is negligible. Hence

$$\frac{d}{dt}\left(n_B/s\right) = \frac{1}{s}\left(\frac{dn_B}{dt}\right)_{\text{loops}}$$

Integrating from $t = t_e$ onwards (effectively to ∞ we find

$$\left(\frac{n_B}{s}\right)_{\text{final}} \sim 300 \; \frac{f\epsilon\mu^2}{m_X M_P^3} \; \sim 300 \; \frac{f\epsilon}{\alpha_G^2}\left(\frac{m_X}{M_P}\right)^3$$

where f is the fraction of loops produced that collapse in this way. If f is of order 10^{1} and $m_X \sim 5 \times 10^{14}$ GeV, we get typically $n_B/s \sim 10^{-8}\varepsilon$.

If following Nanopoulos and Weinberg [24], we assume that ε lies in the range between 10^{-2} and 1, we obtain a value in good agreement with the present observational bound $n_B/s \sim 10^{-9.8\pm1.7}$[22].

This is of course in addition to any baryon asymmetry created ealier. One interesting feature of the mechanism is that baryon number is not generated uniformly throughout space but in clumps around the collapsing strings. However, the scale of these clumps is too small to be of any relevance to galaxy formation.

Great uncertainties remain in this theoretical prediction. First, the number of superheavy bosons released per unit length of collapsing string is uncertain. It should be possible to calculate it, but a deeper understanding of the quantum or at least semi-classical theory of strings is needed. Second, the estimate of the time t_e at which the process of baryon-number generation by collapsing strings effectively begins is rather crude. In reality there is no sharp beginning. The process is a continuous one. Numerical calculations are presently under way to improve this estimate. But since $n_B/s \propto t_e^{-1/2}$, we do not expect the result to change very significantly. Lastly, the parameter ε is highly model-dependent and cannot at present be calculated from first principles.

What we have shown however is that this process of collapsing cosmic strings may be a significant contributor to the total net baryon number of the universe. Certainly in those GUTs that predict the appearance of stable strings it cannot be ignored.

CHAPTER II

GAUGE THEORIES AND INTEGRABLE SYSTEMS

2.1 SELF DUALITY AND TODA SYSTEMS

Work on finding exact solutions for sourceless Yang-Mills or Yang-Mills-Higgs systems has focussed to a large extent on 'self dual' configurations, for several good reasons. First, they are governed by first order equations which are easier to solve. Second, they describe in the Euclidean theory configurations of finite and stationary action, and are thus good candidate ground states for a semiclassical quantum theory to be built on, 'instantons'. In the case of magnetic monopoles, the self duality equations are equivalent to the Bogomolny equations, describing magnetic monopole configurations of minimum energy. As we shall see, under certain symmetry conditions the self duality equations describe completely integrable systems. This, one hopes, will give a clue as to how to quantise them.

In this section I will motivate the following two sections by showing how Toda systems[1] arise in the context of self-dual gauge fields. I will do this via the approach of Yang[22] which as I will show, explains and clarifies the seemingly ad-hoc approach of Leznov and Saveliev[2,3] which has proved so useful in monopole studies.[4]

Yang pointed out that the three equations

$$F_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta} \tag{2.1.1}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu]$ in Euclidean spacetime could be rewritten as

$$F_{yz} = F_{\bar{y}\bar{z}} = F_{y\bar{y}} + F_{z\bar{z}} = 0 \tag{2.1.2}$$

upon choosing

$$y = \tfrac{1}{2}(x_1 + ix_2) \qquad z = \tfrac{1}{2}(x_3 - ix_4)$$
$$\bar{y} = \tfrac{1}{2}(x_1 - ix_2) \qquad \bar{z} = \tfrac{1}{2}(x_3 + ix_4) \tag{2.1.3}$$

If through the use of some symmetry imposed on the solution one can eliminate the ∂y and $\partial\bar{y}$ derivatives, one can then define new potentials

$$a_z = A_z + iA_{\bar{y}}$$
$$a_{\bar{z}} = A_{\bar{z}} + iA_y \tag{2.1.4}$$

Then equations (2.1.2) guarantee that

$$f_{z\bar{z}} = \partial_z a_{\bar{z}} - \partial_{\bar{z}} a_z + i[a_z, a_{\bar{z}}] = 0 \tag{2.1.5}$$

and we have an integrable system.

The simplest conceivable symmetry is to set $\partial y = \partial\bar{y} = 0$; this guarantees (2.1.5) and as we shall see in (2.3.11) and (2.3.24) reading A_u for a_z and A_v for $a_{\bar{z}}$ shows that all Toda systems correspond to self-dual Yang-Mills configurations.

Up till now this was only known for the Toda molecule systems. This observation probably explains previous results suggesting a 'hidden Kac-Moody Symmetry' in self dual gauge field configurations obtained by Dolan and others.[23] It also raises the possibility of some interesting new solutions based on the Toda lattice equations, but that is a topic for future research.

I want to show next that the imposition of spherical symmetry on a Yang-Mills-adjoint Higgs system, described by the 'Bogomolny' equations'

$$B_i = D_i\Phi \,, \quad \partial_0 = 0 \,, \quad A_0 = 0 \tag{2.1.6}$$

leads, through Yangs approach, to the Leznov-Saveliev method of solution.

We consider (2.1.6) as a self duality equation (2.1.1) with the four euclidean dimensions consisting of the three spatial dimensions and a fictitious X_4, with $\Phi = A_4$. Then spherical symmetry is defined as

$$\begin{aligned} &[\underline{L} + \underline{t}, A_4] = 0 \\ &[L_i + t_i, A_j] = i\epsilon_{ijk} A_k \,, \quad i,j = 1,2,3 \end{aligned} \tag{2.1.7}$$

where $\underline{L} = -i\underline{r}\wedge\underline{\nabla}$ are the generators of angular momentum, and the t_i form an SU(2) subalgebra of the gauge group algebra involved. we choose a noncoordinate basis in which the metric is diag (1,1,1,1),

$$\begin{aligned} dx_1 &= r\,d\theta & \partial_1 &= \frac{1}{r}\partial_\theta \\ dx_2 &= r\sin\theta\, d\varphi & \partial_2 &= \frac{1}{r\sin\theta}\partial_\varphi \\ dx_3 &= dr & \partial_3 &= \partial_r \\ dx_4 &= dx_4 & \partial_4 &= \partial_4 = 0 \end{aligned} \tag{2.1.8}$$

We want to work only on the Z axis, and allow all angular dependence to be determined solely by (2.1.7). So we choose the axis $\theta = 0$

for our spherical polar coordinates to be the x axis. Further we choose a gauge in which $A_r = 0$. Then the transverse parts of (2.1.7) lead to

$$\partial_y A_\mu = \left[-\frac{t_-}{r}, A_\mu \right]$$
$$\partial_{\bar{y}} A_\mu = \left[+\frac{t_+}{r}, A_\mu \right] \qquad (2.1.8)$$

with $A_\mu = A_i$, $i = 1,2,4$, $t_\pm = t_1 \pm it_2$ and the radial part (using $\underline{r}.\underline{L} = 0$) leads to

$$[t_3, A_4] = 0$$
$$[t_3, A_{\bar{y} \atop y}] = \pm A_{\bar{y} \atop y} \qquad (2.1.9)$$

which are a statement of grading i.e. Ay is grade1, $A\bar{y}$ grade-1, and A_4 grade 0.

Yangs equations in this noncoordinatebase (where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu] - c_{\mu\nu}{}^\alpha A_\alpha$, $c_{\mu\nu}{}^\alpha$ being thestructure constants of the algebra of the basis i.e. $[\partial_\mu, \partial_\nu] = c_{\mu\nu}{}^\alpha \partial_\alpha$ and in this case $c_{yz}{}^y = c_{\bar{y}z}{}^{\bar{y}} = c_{y\bar{z}}{}^y = c_{\bar{y}\bar{z}}{}^{\bar{y}} = 1/r$) upon using (2.1.8) to eliminate terms in ∂_y and $\partial_{\bar{y}}$ read

$$\partial_z A_y + i\left[A_z, A_y + \frac{it_-}{r}\right] + \frac{1}{r} A_y = 0$$
$$\partial_{\bar{z}} A_{\bar{y}} + i\left[A_{\bar{z}}, A_{\bar{y}} - \frac{it_+}{r}\right] + \frac{1}{r} A_{\bar{y}} = 0$$
$$\partial_z A_{\bar{z}} - \partial_{\bar{z}} A_z + i\left[A_y + \frac{it_-}{r}, A_{\bar{y}} - \frac{it_+}{r}\right] + i[A_z, A_{\bar{z}}] + 2i\frac{t_3}{r^2} = 0$$

(2.1.10)

using $[t_+, t_-] = 2t_3$. Then it is easy to see that defining

$\widehat{A}_y = A_y + i\frac{t_-}{r}$, $\widehat{A}_{\bar{y}} = A_{\bar{y}} - \frac{it_+}{r}$, $\widehat{A}_{\substack{z\\ \bar{z}}} = A_{\substack{z\\ \bar{z}}} \pm \frac{it_3}{r}$ leads to

$$\partial_z \widehat{A}_y \quad + i \left[\widehat{A}_z, \widehat{A}_y\right] = 0$$
$$\partial_{\bar{z}} \widehat{A}_{\bar{y}} \quad + i \left[\widehat{A}_{\bar{z}}, \widehat{A}_{\bar{y}}\right] = 0$$
$$\partial_z \widehat{A}_{\bar{z}} - \partial_{\bar{z}} \widehat{A}_z + i\left[\widehat{A}_y, \widehat{A}_{\bar{y}}\right] + i\left[\widehat{A}_z, \widehat{A}_{\bar{z}}\right] = 0 \tag{2.1.11}$$

Now as in (2.1.4) we define new potentials a_z and $a_{\bar{z}}$ satisfing $f_{z\bar{z}} = 0$. It is easily checked that (note that $\widehat{A}_z = -\widehat{A}_{\bar{z}} = i\widehat{A}_4$)

$$a_r = \tfrac{1}{2}(a_z + a_{\bar{z}}) = \tfrac{i}{2}\left(\widehat{A}_{\bar{y}} + \widehat{A}_y\right)$$
$$a_4 = \tfrac{i}{2}(a_{\bar{z}} - a_z) = \widehat{A}_4 + \tfrac{1}{2}(A_{\bar{y}} - A_y)$$

provides the Lax pair of Leznov and Saveliev.[2,3,4] It is hoped that this more systematic understanding of the construction of integrability conditions for selfdual configurations will lead to new solutions with other symmetries - cylindrical symmetry for example.

2.2 THE TODA EQUATIONS

The Toda field equations in two dimensions (one space, one time),

$$\partial^2 \rho_a = -\sum_{b=1}^{n} C_{ab} e^{\rho_b} \quad a=1\ldots n \qquad (2.2.1)$$

have been the focus of much current research in mathematical physics. The equations are specified by the 'Cartan' matrix C_{ab} with integer entries which provides the connection with Lie algebras. There are three classes of Toda field equations - what we shall call the Toda Molecule (TM) equations, the Toda Lattice (TL) equations, and the generalised Toda Lattice (GTL) equations.

The simplest class, the TM equations have for C_{ab} the $r \times r$ ordinary Cartan matrix K_{ab} uniquely specifying a simple Lie algebra of rank r. Toda's original equations, where the ρ_a were functions of time only and represented the relative displacement of points on an infinite linear lattice, correspond to the case where K_{ab} is the Cartan matrix for SU(r+1), and r is taken to infinity.[1]

The Lie algebraic structure of the TM equations has only recently been appreciated, and used to find their general solution in terms of 2r arbitrary functions[2]. The simplest case is when r=1. Here, the Cartan matrix is simply the number 2, specifying the algebra SU(2). Then equation (2.2.1) reads as the Minkowski space version of the Liouville equation, $\partial^2 \rho = -2e^{\rho}$. The TM equations can thus be regarded as the relativistic multicomponent equations generalising the Liouville equation . As in the latter equation, there is no ground state for finite ρ_a since K_{ab} is a nonsingular matrix and so has no null vectors. However the TM equations do arise naturally in the study of spherically symmetric self-dual magnetic monopoles [3,4] and axially symmetric instantons [5,6].

The second class of Toda equations, the TL equations, have as C_{ab} the

$(r+1)\times(r+1)$ extended Cartan matrix $\overline{K}_{ab}$ associated with any simple Lie algebra of rank r. It is obtained from the Cartan matrix for the algebra by adding a row and a column in a way to be explained subsequently. Unlike the Cartan matrix, $\overline{K}_{ab}$ is singular and thus has a null vector. This means that a certain linear combination of the ρ_a satisfies the wave equation and can be consistently set zero, and that the equations then have a unique constant solution defining a ground state. The TL equations are therefore more interesting as a field theory than the TM equations.

The Lie algebraic structure of the TL equations ensures that they, like the TM equations, may be expressed as zero curvature conditions i.e. integrability conditions for linear equations. This allows their integration via the inverse scattering method.

The simplest case of the TL equations is when r=1 and so the algebra is SU(2), whose extended Cartan matrix is

$$\begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} \tag{2.2.2}$$

The corresponding one variable equation (taking $\rho_1+\rho_2$ to vanish and $\rho=\rho_1$) is

$$\partial^2\rho = -4\sinh\rho \tag{2.2.3}$$

This equation is related to the more familiar Sine-Gordon equation $\partial^2\varphi = -4\sin\varphi$ by the substitution $\rho = i\varphi$ which introduces the periodic function $\sin\varphi$ and leads to degenerate vacua and topological soliton solutions.

The final class of Toda equations, the GTL equations have for C_{ab} an $(n\times n)$ 'generalised' Cartan matrix $\hat{K}_{ab}$ as defined by Berman et. al. [7] not corresponding to any particular Lie algebra but obeying similar restrictions, and in particular possessing a null vector. The simplest case here is where n=2 and the matrix is

$$\begin{pmatrix} 2 & -4 \\ -1 & 2 \end{pmatrix} \tag{2.2.4}$$

For the vacuum we set $\rho_1+2\rho_2=0$ and $\rho=\rho_2$ to obtain

$$\partial^2\rho = -2e^{\rho} + e^{-2\rho} \tag{2.2.5}$$

which is the Bullough-Dodd equation [8].

In this paper we develop a systematic procedure for dealing with Toda equations. Firstly, we classify all of them, including those which seem to have been previously ignored. But secondly, we develop the procedure of reduction, which enables one to obtain from a given Toda equation, its integrability condition and general solution, by identifying variables in a systematic way, those for another Toda equation in fewer variables. Our procedure for reduction is based on the use of symmetries of Dynkin diagrams.

A Dynkin diagram $D(G)$ corresponding to the Lie algebra G uniquely encodes the structure of the Cartan matrix $K(G)$. The extended Dynkin diagram $\bar{D}(G)$ obtained by adding one point and its links to $D(G)$ similarly specifies $\bar{K}(G)$. Finally, the Generalised Cartan matrices also have associated Generalised Dynkin diagrams (GD's).

The symmetries of $D(G)$, and $\bar{D}(G)$ are denoted $\Gamma(G)$ and $\bar{\Gamma}(G)$ and constitute the automorphisms of the TM and TL equation as explained in section 2. $\Gamma(G)$ is simply the group of outer automorphisms of G according to a standard mathematical result. $\overline{\Gamma}(G)$ is much richer. It possesses an invariant subgroup $W_o(G)$ which corresponds to the subgroup of the Weyl group whose precise significance is explained in Section 5. We show there that

$$\overline{\Gamma}(G) / W_o(G) \cong \Gamma(G)$$

and

$$W_o(G) \cong Z(\widetilde{G})$$

where $Z(\widetilde{G})$ is the centre of the universal covering group $\widetilde{G}$ of G. These group theoretic results are interesting in their own right, being important in connection with vortex strings in Grand Unified Theories.

The reduction procedure explained in Section 6 is very simple and yields the following useful results. First, one obtains <u>all</u> TM equations, their integrability conditions and general solutions from only those corresponding to simply laced Dynkin diagrams (diagrams with only single lines), which specify symmetric Cartan matrices. Second, one obtains <u>all</u> TL equations and all GTL equations, their integrability conditions and general solutions from only those TL equations corresponding to simply laced extended Dynkin diagrams, again specifying symmetric extended matrices. This simplifies the task of the solution and analysis of the Toda equations considerably.

The reduction procedure is also a simple method for obtaining many special subalgebras of Lie algebras and thus extends the use of Dynkin diagrams which have as far as we know hitherto only been used for obtaining regular subalgebras.

2.3 CARTAN MATRICES, DYNKIN DIAGRAMS AND THEIR SYMMETRIES.

The TM and TL equations involving the matrices K_{ab} and $\bar{K}_{ab}$ can, as we shall see in Section 4, be understood as the compatibility conditions for certain first order linear equations whenever K_{ab} or $\bar{K}_{ab}$ constitute the ordinary or extended Cartan matrices of a simple Lie algebra . The GTL equations involve a Generalised Cartan Matrix $\hat{K}_{ab}$ not corresponding to any particular Lie algebra and are, as we shall show in Section 6, obtained naturally by a 'reduction' of the TL equations.

All these Cartan matrices are square nxn matrices C_{ab} with integer entries having in common the properties

a) All diagonal elements take the value 2

b) All off-diagonal elements are zero or negative

c) If C_{ab} vanishes then so does C_{ba}

Any such matrix can be conveniently represented by a Dynkin diagram obtained as follows. The diagram consists of n points with points a and b joined by $C_{ab}C_{ba}$ lines. When $C_{ab}C_{ba}$ exceeds one, and $|C_{ba}|$ equals one an arrow drawn on the lines pointing from a to b indicates that $|C_{ab}|$ equals that number of lines. If both $|C_{ab}|$ and $|C_{ba}|$ exceed one, then an arrow is drawn on $|C_{ab}|$ lines pointing from a to b, and one on $|C_{ba}|$ lines pointing from b to a. Thus C_{ab} and its Dynkin diagram can be reconstructed from each other. A Cartan matrix C_{ab} is called indecomposable if its Dynkin diagram is connected.

We shall consider two classes of such matrices. The ordinary Cartan Matric K_{ab} were discovered by Cartan and correspond to the simple Lie algebras. These are nonsingular and positive definite. Their Dynkin diagrams are tabulated in Table I. The Cartan matrices with left null vectors were calculated and listed by Berman, Moody and Wonenburger [7] and comprise the extended and generalised Cartan matrices tabulated in Tables II and III respectively. These tables presumably list all positive semidefinite Cartan matrices with properties and c above.

In searching for symmetries of the Toda equations, we look for permutatio P of the variables in equations (1.1):

$$p: \rho_a \rightarrow \rho_{p(a)} \qquad a = 1 \dots n \tag{2.3.1}$$

which leave the equations invariant in the sense that if the set $\{\rho_a\}$ is a solution then so is $\{\rho_{p(a)}\}$. This is guaranteed if

$$C_{ab} = C_{p(a)\,p(b)} \quad \forall a,b = 1 \dots n \tag{2.3.2}$$

which is true if and onlyif the permutation respects the structure of the Dynkin diagram corresponding to C.

Let us consider first the ordinary Dynkin diagrams D(G), shown in Table I. We have denoted certain points by a cross rather than a circle. These correspond to the 'minimal coweights' as explained in [9] and are relevant to studies of stable magnetic monopoles, and of symmetry breaking in Grand Unified Theories. What is important to us here is that a permutation of the points of D(G) respecti the structure of D(G) can be specified by its action on the crossed points and this fact will providea convenient notation for the symmetry operations.

The symmetry groups formed by the permutations respecting the structure of D(G) are denoted $\Gamma(G)$ and are tabulated in the last column of Table I. We denote as Z_n the cyclic group of order n and S_n as the permutation group on n objects The elements of $\Gamma(G)$ carry solutions of the corresponding TM equations onto other solutions of the same equations.

Next we examine the extended Dynkin diagrams $\overline{D}(G)$, corresponding to the TL equations. The extended diagram $\overline{D}(G)$ is obtained from the ordinary diagram $D(G)$ by adding one point, labelled 0 and its links. This corresponds to adding one extra row and column to the ordinary Cartan matrix as was noted above. The extended Dynkin diagrams $\overline{D}(G)$ for each simple Lie algebra G are tabulated in Table II. The extra point, labelled 0, is denoted by a cross. Note that the removal of any one of the crossed points yields $D(G)$ from $\overline{D}(G)$.

The symmetries of the TL equations are again given by the permutations of the points of $\overline{D}(G)$ which respect the structure of the diagram $\overline{D}(G)$ and form the group $\overline{\Gamma}(G)$ tabulated in the last column of Table II. These groups are somewhat richer in structure than $\Gamma(G)$. For example, $\overline{\Gamma}(SU(r+1))$, $r\geq 2$, is given by D_{r+1} , the dihedral group with $2(r+1)$ elements. This arises because D_{r+1} is the symmetry group of a regular $r+1$ sided plane figure, being generated by the permutation $(0,1,2\ldots(r+1))$,in cycle notation, and the nontrivial element of $\Gamma(SU(r+1))$, $(1,r)(2,r-1)\ldots(r/2,r/2+1)$ (if r is even) or $(1,r)(2,r-1)\ldots\left(\frac{r-1}{2},\frac{r+3}{2}\right)\left(\frac{r+1}{2}\right)$ (if r is odd), which of course in both cases leaves the point 0 fixed.

Again the symmetries in $\overline{\Gamma}(G)$ can be conveniently be specified by their action on the crossed points. For example $\overline{\Gamma}(SO(2r))$, $r>2$, equals D_4, formed by the permutations of the crossed points, in cycle notation:

$$D_4=\{(\,);(r-1,r);(0,1);(0,1)(r-1,r);(0,r-1)(1,r);(0,r)(1,r-1);\\ (0,r-1,1,r);(0,r,1,r-1)\} \tag{2.3.3}$$

$\overline{\Gamma}(E_6)$ consists of the six permutations of the points 0, 1 and 6 of $\overline{D}(E_6)$ and $\overline{\Gamma}(SO(8))$ of the 24 permutations of the points 0,1,3 and 4 of $\overline{D}(SO(8))$ and so on.

Notice in particular the nontrivial element of $\overline{\Gamma}(SU(2))\cong Z_2$ interchanging ρ_1 and ρ_2 , and hence reversing the difference ρ in equation (2.2.3)

This is the automorphism allowing ρ to be replaced by $i\varphi$, to obtain a real equation in a circular function, $\sin\varphi$, namely the Sine-Gordon equation.

Finally, we turn to the generalised Dynkin diagrams in Table III, corresponding to the GTL equations. The first four of these are seen to be related to four of the extended Dynkin diagrams by the reversal of arrows. We call the diagram obtained from $\overline{D}(G)$ by simply reversing the arrows $\overline{D}^T(G)$, since it is easily seen from the rules for obtaining the Cartan matrix above, that reversal of the arrows corresponds to taking the transpose of the Cartan matrix. The last two diagrams $GD(H_r)$ and $GD(BD)$ cannot be obtained in this way.

The GTL equations seem to have been hitherto ignored (apart from the BD equ (2.2.5)). They are not in particular related to any root system - this is immediate clear from H_r whose points would correspond to 'roots' of three different lengths whereas simple Lie algebras can have roots of at most two distinct lengths.

2.4 LIE ALGEBRAS AND THE INTEGRABILITY OF THE TODA EQUATIONS.

Let us consider a Cartan-Weyl basis for the simple Lie algebra G involving Cartan subalgebra generators H_i, i=1,2..... r = rank G and step operators e_α. These satisfy

$$[H_i, H_j] = 0 \quad ; \quad [H_i, e_\alpha] = \alpha_i e_\alpha \quad ; \quad [e_\alpha, e_{-\alpha}] = \alpha . H \qquad (2.4.1)$$

The r component vectors α are the roots of the algebra. We have further that

$$\begin{aligned} [e_\alpha, e_\beta] &= N_{\alpha\beta} e_{\alpha+\beta} && \text{if } \alpha+\beta \text{ is a root} \\ &= 0 && \text{if } \alpha+\beta \text{ is not a root and does not vanish} \end{aligned} \qquad (2.4.2)$$

The coefficients $N_{\alpha\beta}$ can be specified precisely but will not be necessary in this paper. We shall assume that the reader is familiar with the classic paper by Racah [10] or with one of the recent excellent books on Lie algebra theory [11,12].

A fundamental geometrical property of the root system of any simple Lie algebra G is that there exists a basis of simple roots

$$\Delta(G) = \{ \alpha^a ; \; a = 1,2 \ldots r = \text{rank}\, G \} \qquad (2.4.3)$$

such that either the positive or the negative of any root can be expressed as a sum of simple roots. Knowledge of the simple roots determines the Lie algebra up to an isomorphism, and the knowledge is embodied in the Cartan matrix (the "ordinary" Cartan matrix of the previous section),

$$K_{ab} = 2\alpha^a.\alpha^b/(\alpha^b)^2 \tag{2.4.4}$$

This is the matrix that occurs in the TM equation and we see immediately that $K_{ab}K_{ba}$ lies between 0 and 4, equalling 4 if and only if α^a and α^b are parallel. It is a property of roots systems that two roots are parallel only if $\alpha^a = \pm\alpha^b$. Also K_{ab} can only take integer values, two on the diagonal and negative or zero off the diagonal. As we explained,these features enable the Cartan matrix to be encoded in the Dynkin diagram which therefore represents the structure of the corresponding Lie algebra in a succinct and surprisingly useful way.

The Cartan-Weyl basis is useful because if its orthonormality properties with respect to the trace of a pair of generators. For algebraic purposes it is more convenient to introduce a slightly modified basis called the Chevalley basis:

$$E_\alpha \equiv \sqrt{2/\alpha^2}\, e_\alpha \qquad H_\alpha \equiv 2\alpha.H/(\alpha)^2 \tag{2.4.5}$$

Then

$$[H_\alpha, E_\beta] = 2\beta.\alpha/(\alpha)^2\, E_\beta \tag{2.4.6}$$

In particular if we concentrate on the simple roots (2.4.3) and denote

$$E_{\pm\alpha^a} = E_{\pm a} \ ; \ H_{\alpha^a} = H_a \tag{2.4.7}$$

we find

$$[H_a, H_b] = 0 \ ; \ [H_a, E_b] = E_b K_{ba} \ ; \ [E_a, E_{-b}] = \delta_{ab} H_b \quad (2.4.8)$$

In the last equation we have used the fact that by the definition of simple roots, $\alpha^a - \alpha^b$ cannot be a root. The significance of the Chevalley basis is that the structure constants are all integers. In practice in this paper we shall only use the commutators (2.4.8) and not the other unspecified commutators.

We are now in a position to formulate the TM equations

$$\partial^2 \rho_a = - K_{ab} e^{\rho_b} \quad (2.4.9)$$

as a zero curvature condition thereby establishing why the matrices K occurring in them have a Lie algebraic significance if the equations are to be integrable. The results we are about to explain are due to the contributions of several authors [13-17, 2].

Let us define the light cone variables

$$u = \frac{x+t}{2} \qquad v = \frac{x-t}{2} \quad (2.4.10)$$

so that $$\partial_u \partial_v = -\partial_t^2 + \partial_x^2 \equiv -\partial_\mu \partial^\mu$$

and the two dimensional gauge potentials A_μ whose light cone components are

$$A_u = \sum_{a\in\Delta}\left(\tfrac{1}{2}H_a\partial_u\phi_a + e^{\frac{1}{2}K_{ab}\phi_b}E_a\right) \qquad (2.4.11)$$
$$A_v = \sum_{a\in\Delta}\left(-\tfrac{1}{2}H_a\partial_v\phi_a + e^{\frac{1}{2}K_{ab}\phi_b}E_{-a}\right)$$

Then using the commutators(2.4.8) we find the curvature

$$\left[\partial_u + A_u, \partial_v + A_v\right] = \sum_{a\in\Delta}\left(\partial^2\phi_a + e^{K_{ab}\phi_b}\right)H_a \qquad (2.4.12)$$

The $H_1 \ldots H_r$ are linearly independent and K is nonsingular. Hence we can define

$$\rho_a \equiv \sum_{b\in\Delta} K_{ab}\,\phi_b$$

and conclude that the curvature 2.4.12) vanishes if and only if the TM equation (2.4.9) is satisfied by ρ_a . So the TM equations can be regarded as the compatibility or integrability conditions for the linear problems

$$\left(\partial_\mu + A_\mu\right)\Psi = 0$$

where Ψ is a nonsingular matrix.

Before discussing the TL equations $\partial^2\rho_a = -\bar{K}_{ab}e^{\rho_b}$ in a similar way we need to assemble some more Lie algebraic concepts.

Let us denote ψ the highest root of G, namely that root which exceeds each of the other roots by a sum of simple roots.

Let us also denote $\lambda^{a\vee}$, $a\in\Delta$, as the fundamental coweights of G where

$$2\lambda^{a\vee}\cdot\alpha^b = \delta^{ab} \qquad a,b\in\Delta \qquad (2.4.13)$$

Then we say that the fundamental coweight $\lambda^{a\vee}$ is minimal if

$$2\lambda^{a^\vee}.\psi = 1 \qquad (2.4.14)$$

Now let us denote

$$\alpha^0 = -\psi \qquad (2.4.15)$$

and consider the "extended system of simple roots".

$$\overline{\Delta}(G) = \{\alpha^0, \alpha^1 \ldots\ldots \alpha^r\} \qquad (2.4.16)$$

We see that $\overline{\Delta}$ possesses the same property as Δ that the difference of any two members cannot be a root. Hence

$$[E_a, E_{-b}] = \delta_{ab} H_b \qquad a, b \in \overline{\Delta} \qquad (2.4.17)$$

having extended the definitions of H_a and $E_{\pm b}$ to include a=0. Associated with $\overline{\Delta}$ is the extended Cartan matrix,

$$\overline{K}_{ab} = 2\,\alpha^a.\alpha^b/(\alpha^b)^2 \qquad (2.4.18)$$

and with this the extended Dynkin diagram $\overline{D}(G)$ tabulated in Table II. The extra point labelled 0, when compared to $D(G)$ corresponds to the root α^0.

The highest root ψ can be expanded as integer linear combinations in two ways; first in terms of simple roots:

$$\psi = \sum_{a\in\Delta} N_a \alpha^a \qquad (2.4.19)$$

By definition ψ exceeds any simple root so that the integer coefficients N_a each satisfy

$$N_a \geqslant 1 \tag{2.4.20}$$

If we define $N_0 = 1$ we can rewrite (2.4.19) using (2.4.15) as

$$\sum_{a \in \bar{\Delta}} N_a \alpha^a = 0 \tag{2.4.21}$$

It follows from the definition (2.4.18) that $\bar{K}$ has a left null vector N_a,

$$\sum_{a \in \bar{\Delta}} N_a \bar{K}_{ab} = 0 \tag{2.4.22}$$

Since K is nonsingular $\bar{K}$ has rank r and this is the unique null vector.

Note that N_a equals 1 only if a=0 or else if $a \in \Delta$ and the coweight is minimal. This is because by (2.4.13) and (2.4.19)

$$N_a = 2\lambda^{a^\vee} \cdot \psi$$

and the result follows by (2.4.14). In Table II the points of $\bar{D}(G)$ for which $N_a = 1$ are denoted by crosses rather than circles.

Alternatively ψ , being a weight, can be expanded as an integer linear combination of fundamental weights $\lambda^a = \lambda^{a^\vee}(\alpha^a)^2$

$$\psi = -\sum_{a \in \Delta} \bar{K}_{0a} \lambda^a \tag{2.4.23}$$

with the coefficients $\bar{K}_{0a}$ as indicated. This follows from (2.4.13) and (2.4.18). Unless α^a and α^0 are parallel $\bar{K}_{0a}\bar{K}_{a0}$ = 0,1,2,3 and so it follows that α^0 is a long root

$$\bar{K}_{a0} = -1$$

except for SU(2) which by (2.4.20) is the only algebra for which α^a and α^0 can be parallel. These two results provide the easiest way in practice of calculating $\overline{K}_{oa}$ and $\overline{K}_{ao}$ and hence $\overline{D}(G)$ as appearing in Table II. Notice that the arrows are now seen, for both $D(G)$ and $\overline{D}(G)$ to point from long roots to short roots i.e. 'downhill'.

Now we return to the integrability condition for the TL equations. Consider gauge potentials A_μ formally the same as the gauge potentials (2.4.11) used for the TM equations but with $\overline{\Delta}$ and $\overline{K}$ replacing Δ and K respectively,

$$A_u = \sum_{a\in\overline{\Delta}}\left(\tfrac{1}{2}H_a\partial_u\phi_a + e^{\frac{1}{2}\overline{K}_{ab}\phi_b}E_a\right)$$
$$A_v = \sum_{a\in\overline{\Delta}}\left(-\tfrac{1}{2}H_a\partial_v\phi_a + e^{\frac{1}{2}\overline{K}_{ab}\phi_b}E_{-a}\right) \qquad (2.4.24)$$

Then using the commutators (2.4.17) and (2.4.8) extended to include a=o we find

$$\left[\partial_u + A_u, \partial_v + A_v\right] = \sum_{a\in\overline{\Delta}}\left(\partial^2\phi_a + e^{\overline{K}_{ab}\phi_b}\right)H_a \qquad (2.4.25)$$

as before, equation (2.4.12). Now, however, the $H_a,\ a=0\ldots r$ are not linearly independent, because of (2.4.21) and their coefficients cannot be equated to zero when (2.4.25) vanishes.

Instead we have
$$\sum_{a\in\overline{\Delta}}\frac{2(\alpha^a)_i}{(\alpha^a)^2}\left(\partial^2\phi_a + e^{\overline{K}_{ab}\phi_b}\right) = 0$$

or equivalently,
$$\overline{K}_{ab}\left(\partial^2\phi_b + e^{\overline{K}_{bc}\phi_c}\right) = 0$$

Now let us define
$$\rho_a = \sum_{b\in\overline{\Delta}}\overline{K}_{ab}\phi_b$$

so that as a consequence of (2.4.22)

$$\sum_{a\in\bar{\Delta}} N_a \rho_a = 0 \tag{2.4.26}$$

Then the zero curvature condition stating that (2.4.25) vanishes reduces to

$$\partial^2 \rho_a + \bar{K}_{ab} e^{\rho_b} = 0 \tag{2.4.27}$$

which is the TL equation.

There appears to be one more variable ϕ_a occurring in (2.4.24) than the resultant number of variables ρ_a , namely r because of (2.4.26). This is illusory - in fact the potentials do not depend on the variable ϕ defined by

$$\phi_a = \sum_{a\in\bar{\Delta}} N_a (\alpha^a)^2 \phi$$

Before closing this section we want to discuss two features of the TL equations (2.4.27) which distinguish them from the TM equations.

The first is that the TL equations (2.4.27) possess a finite constant solution $\hat{\rho}_a$. This is only possible because $\bar{K}_{ab}$ is singular and has a right null vector $N_a(\alpha^a)^2$ to which $e^{\hat{\rho}_a}$ must be proportional, with condition (2.4.26) determining the constant of proportionality. We find

$$\hat{\rho}_a = \frac{1}{h} \sum_{b\in\bar{\Delta}} N_b \ln\left[N_a(\alpha^a)^2 / N_b(\alpha^b)^2 \right] \tag{2.4.28}$$

where $h = \sum N_a$ is the Coxeter number of G. Reference to Table II tells us that for $SU(N), (\alpha^a)^2 = (\alpha^b)^2,\ N_a = 1$ so $\hat{\rho}_a = 0$ for all a. In fact $\hat{\rho}_a$ minimise the Hamiltonian and so constitutes a ground state.

Finally it is possible to replace

$$E_0 \to z E_0 \quad ; \quad E_{-0} \to \frac{1}{z} E_{-0}$$

in the gauge potentials (2.4.24) and, providing z is an arbitrary constant, one finds that the condition (2.4.25) is independent of it. The significance of this is that the potentials (2.4.24) so obtained for different values of z are gauge inequivalent. So there is an infinite set of linear problems depending on the special parameter z which have the same TL equation (2.4.27) as their integrability condition.

The calculation leading from the potentials to the curvature hinged upon the fact that differences of roots in Δ or $\bar{\Delta}$ are not roots. Root systems with this property have been called admissible systems [15]. There are other admissible root systems which correspond to the generalised Dynkin diagrams listed in Table III, and these allow similar integrability conditions to be derived for the GTL equations. However, as we shall show, these are most economically obtained as the 'reductions' of TL equations, and so will not be considered further in this section.

2.5 SYMMETRIES OF THE GAUGE POTENTIALS.

In Section 2.3 we derived certain symmetries of the Toda equations. Now that these equations have been formulated as the compatibility conditions for a linear problem involving Lie algebras we shall analyse the symmetries of these linear problems and relate them to the previously found symmetries, thereby obtaining a deeper Lie algebraic understanding of them.

Note that the terms occurring in the gauge potentials can be said to belong to three subspaces of the Lie algebra; the Cartan subalgebra

$$H = \left\{ \sum_{i=1}^{n} x_i H_i \right\} \tag{2.5.1}$$

and

$$V_{\pm\Delta} = \left\{ \sum_{a\in\Delta} y_a E_{\pm a} \right\}$$

or

$$V_{\pm\bar{\Delta}} = \left\{ \sum_{a\in\bar{\Delta}} y_a E_{\pm a} \right\}$$

The derivation of the Toda equations as compatibility conditions (2.4.12) or (2.4.25) made use of the Lie algebra commutation relations. We shall seek the symmetries of the Lie algebra preserving the commutators (and linear structure) and the subspaces H, $V_{\pm\Delta}$ (or $V_{\pm\bar{\Delta}}$ as appropriate). The symmetries of the Lie algebra (without reference to the subspaces) are called automorphisms and form a group which we denote aut G. Conjugation of the generators X with respect to any element of the Lie group G, g say

$$X \to g X g^{-1} \tag{2.5.2}$$

always provides an automorphism, said to be 'inner'. Thus G itself provides a group of inner automorphisms which is a self conjugate subgroup in aut G. In fact it is a well known theorem that [11, p.87]

$$\mathrm{aut}\,G/G \cong \Gamma(G) \tag{2.5.3}$$

where the quotient group $\Gamma(G)$, the group of 'outer' automorphisms, is that finite symmetry group of the Dynkin diagram D(G) tabulated in Table I.

Let us consider the subgroup aut(G;H) of aut G which preserves the space H(2.5.1). It is not difficult to prove that its action on H is

$$\sum_{i=1}^{n} x_i H_i \longrightarrow \sum_{i,j=1}^{n} H_i A_{ij} x_j \tag{2.5.4}$$

where A_{ij} is a real orthonormal $r \times r$ matrix with the property that if α is a root then so is $A\alpha$. Such matrices form a group which we call $\mathrm{aut}\,\Phi$, where Φ denotes the root system of the Lie algebra with respect to the chosen Cartan subalgebra. Further the action on the step operators is

$$E_{\pm\alpha} \longrightarrow \text{const}\; E_{\pm A\alpha} \tag{2.5.5}$$

If T denotes the maximal torus of G, the maximal abelian subgroup of G, obtained by exponentiating the Cartan subalgebra, it is not difficult to prove that

$$\mathrm{aut}(G;H)/T \cong \mathrm{aut}\,\Phi \tag{2.5.6}$$

An important subgroup of aut Φ is furnished by the Weyl group W, that finite group generated by the reflections in roots i.e.

$$\sigma_\alpha(x) = x - 2\left(x.\alpha/(\alpha)^2\right)\alpha \tag{2.5.7}$$

In fact

$$A\sigma_\alpha A^{-1} = \sigma_{A\alpha} \tag{2.5.8}$$

so that W is a self conjugate subgroup of aut Φ, actually the subgroup of aut Φ provided by inner automorphisms (2.5.2), that is gauge transformations.

Let us define autΔ and aut$\bar{\Delta}$ to be the groups of real $r\times r$ orthogonal matrices preserving respectively Δ and $\bar{\Delta}$, the simple root system and the extended simple root system $(\bar{\Delta} \equiv \Delta \cup (-\psi))$.

It follows from the properties of root systems that

$$\mathrm{aut}\,\Delta \subset \mathrm{aut}\,\bar{\Delta} \subset \mathrm{aut}\,\Phi \tag{2.5.9}$$

i.e. the groups are subgroups as indicated. Using the fact that aut Δ and aut $\bar{\Delta}$ preserve scalar products and thus Cartan matrices, it is not difficult to prove that aut Δ and aut$\bar{\Delta}$ are respectively isomorphic to the groups Γ and $\bar{\Gamma}$ defined in Section 2 and tabulated in Tables I and II:

$$\mathrm{aut}\,\Delta \cong \Gamma \qquad \mathrm{aut}\,\bar{\Delta} \cong \bar{\Gamma} \tag{2.5.10}$$

Consider an element $p \in$ autΦ. Then

$$\Delta' = p(\Delta)$$

is another possible system of simple roots. It is a fundamental property of the Weyl group W that there exists a unique element $\omega \in W$ such that [11, p.51]

$$\omega\Delta' = \Delta$$

So

$$f(p) \equiv \omega p \quad \in \mathrm{aut}\,\Delta \tag{2.5.11}$$

Thus f defines a map from aut Φ onto its subgroup aut Δ. Further,

$$f(P_1)f(P_2) = \omega_1 P_1 \omega_2 P_2 = \omega_1 P_1 \omega_2 P_1^{-1} P_1 P_2 = f(P_1 P_2)$$

as W is a self conjugate subgroup (by (2.5.8)) of aut Φ and so $\omega_1 P_1 \omega_2 P_1^{-1} \in W$. Thus f is a homomorphism and its kernel is W. Hence the quotient group

$$\mathrm{aut}\,\Phi / W \cong \mathrm{aut}\,\Delta \cong \Gamma \qquad (2.5.12)$$

by a remark (2.5.10) above. This is a well known result [11, p.65]

All these statements are preliminaries to set the scene for our discussion of the Toda equations. Consider first the TM equations (2.2.2) with their associated potentials (2.4.11) involving terms in H, V_Δ and $V_{-\Delta}$ (equation (2.5.1)). We seek automorphisms of the Lie algebra preserving these subspaces. These form a subgroup of aut (G;H) which we denote aut (G;H,V_Δ). Then analogously to (2.5.6)

$$\mathrm{aut}(G;H,V_\Delta)/T \cong \mathrm{aut}\,\Delta \cong \Gamma \qquad (2.5.13)$$

This means that once we have fixed a gauge in the maximal torus T (which we have) we can only apply the outer automorphisms

$$H_i \to H_i A_{ij} \quad E_\alpha \to E_{A\alpha}, \quad A \in \mathrm{aut}\,\Delta \qquad (2.5.14)$$

This corresponds to our previous result that the symmetries of (2.4.9) were the symmetries of **D**, namely Γ.

A more interesting situation results in the case of the TL equations (2.4.27) with their associated potentials (2.4.24) involving terms in H, $V_{\bar\Delta}$ and $V_{-\bar\Delta}$ (equation (2.5.1)). We seek automorphisms of the Lie algebra preserving these subspaces. These form a subgroup of aut (G;H) which we denote aut (G;H,$V_{\bar\Delta}$)

Then analagously to (4.6),

$$\mathrm{aut}\,(G; H, V_{\bar{\Delta}})/_{T} \cong \mathrm{aut}\,\bar{\Delta} \cong \bar{\Gamma} \qquad (2.5.15)$$

Because of (2.5.9) this is a richer structure than (2.5.13). We can use the same homomorphism(2.5.11) that we used to map aut Φ onto aut Δ to map aut $\bar{\Delta}$ onto aut Δ . Its kernel is now

$$W_0 \equiv W \cap \mathrm{aut}\,\bar{\Delta} \qquad (2.5.16)$$

and so

$$\mathrm{aut}\,\bar{\Delta}/_{W_0} \cong \mathrm{aut}\,\Delta \cong \Gamma \qquad (2.5.17)$$

This means that aut $\bar{\Delta}$, on the right hand side of(2.5.15) is a semidirect product of W_0 and Γ . Thus the TL equations exhibit extra symmetries W_0 compared to the TM equations which are related to inner automorphisms (gauge transformations) permuting elements of $\bar{\Delta}$.

We have evaluated W_0 for each Lie group and found by computation that W_0 is isomorphic to $Z(\tilde{G})$, the centre of the universal covering group of G,

$$W_0 \cong W \cap \mathrm{aut}\,\bar{\Delta} \cong Z(\tilde{G}) \qquad (2.5.18)$$

It follows from (2.5.10), (2.5.17) and(2.5.18) that

$$\bar{\Gamma}(G)/_{Z(\tilde{G})} \cong \Gamma(G) \qquad (2.5.19)$$

which can easily be checked from Table I and II, at least once it is known how $Z(\tilde{G})$ fits into $\bar{\Gamma}(G)$. For example when $G = SU(N)$, $\bar{\Gamma} = D_N$, $Z = Z_N$ and so $\bar{\Gamma}/_Z \simeq Z_2 \simeq \Gamma$ indeed as it should.

These are the main results of this section, namely that the extra symmetries of the TL over the TM equations constitute a discrete group of gauge transformations isomorphic to the centre of $\widetilde{G}$.

The complete verification of (2.5.18) we shall relegate to an appendix presenting now the SU(N) case as a specimen of the argument.

The simple roots Δ of SU(N) can conveniently be expressed in terms of N orthogonal unit vectors $e_1 \ldots e_N$ [11,12]

$$\Delta = \{e_1 - e_2, e_2 - e_3, \ldots, e_{N-1} - e_N\}$$

whilst the highest root $\psi = e_1 - e_N$. Hence

$$\bar{\Delta} = \{e_N - e_1, e_1 - e_2, e_2 - e_3, \ldots, e_{N-1} - e_N\}$$

We see that the reflection in $\alpha_1 = e_1 - e_2$ interchanges e_1 and e_2 and it follows that the Weyl group of SU(N) permutes the $e_1 \ldots e_N$.

Consider $\sigma \in W$. Then $\sigma(e_1) = e_{k+1}$ for some k. Now $\sigma(\alpha_1) = \sigma(e_1 - e_2)$ $= e_{k+1} - \sigma(e_2) \in \bar{\Delta}$ if and only if $\sigma(e_2) = e_{k+2}$. Proceeding thus we find that $\sigma \in W_0 \equiv W \cap \operatorname{aut}\bar{\Delta}$ if and only if $\sigma(e_i) = \sigma(e_{i+k})$ where it is understood that $e_{i+N} = e_i$. Clearly such σ form the group Z_N which is indeed the centre of SU(N). Thus we have checked (2.5.18) and (2.5.19) for SU(N). All the other simple Lie algebras are treated in the appendix.

Finally let us discuss some gauge transformations in the maximal torus T. Consider the inner automorphism or gauge transformation (2.5.2) with

$$g = z^{-2\lambda^{a\vee}\cdot H} \tag{2.5.20}$$

where $\lambda^{a^\vee}$ is a fundamental coweight as defined in (2.4.13) (actually this g lies in the complexification of the original Lie group G). Using the commutators (2.4.1) it is easy to check

$$H_b \to H_b$$
$$E_b \to \begin{cases} \frac{1}{z} E_b \\ E_b \end{cases} \qquad E_{-b} \to \begin{cases} z E_{-b} & b = a \\ E_{-b} & b \neq a \end{cases} \tag{2.5.21}$$
$$E_0 \to z^{N_a} E_0 \qquad E_{-0} \to z^{-N_a} E_{-0}$$

where N_a is defined by (2.4.19) as the coefficient of α^a in the expansion of the highest root.

This is the basis of our assertion at the end of the preceding section that there is no spectral parameter in the TM case but that there is one for the TL case. Now we show that the spectral parameter is invariant with respect to the discrete gauge transformation associated with $W \cap \mathrm{aut}\bar{\Delta} \cong Z(\tilde{G})$.

Suppose

$$p\alpha^a = \alpha^{p(a)} \qquad a \in \bar{\Delta}, \quad p \in W \cap \mathrm{aut}\,\Delta$$

Then
$$E_{p^{-1}(0)} \to E_0 \qquad zE_0 \to zE_{p(0)} \tag{2.5.22}$$

Now apply the maximal torus gauge transformation (2.5.2), (2.5.20) choosing $a = p(0)$ so

$$E_0 \to z^{N_a} E_0 \qquad zE_{p(0)} \to E_{p(0)}$$

But if $a = p(0)$, $N_a = 1$ as we observed following equation (2.4.22). Hence overall $E_{p^{-1}(0)} \to zE_0$, $zE_0 \to E_{p(0)}$ i.e. z remains in conjunction with E_0.

We have seen that the TL equations (2.4.27) possess a ground state solution (equation (2.4.28)). Further compared to the TM equation (2.4.9) they possess extra symmetries corresponding to the subgroup of the Weyl group preserving the extended root system $\bar{\Delta}$

$$W_0 = W \cap \operatorname{aut} \bar{\Delta} \cong Z(\tilde{G})$$

For SU(2), $W_0 = Z_2$ and it is precisely this extra symmetry of the Sinh-Gordon equation compared to the Liouville equation which enables us to construct the Sine Gordon equation with periodic ground state structure.

We shall not pursue here this method of obtaining periodic equations (with degenerate vacua, and thus topological soliton solutions), but move on directly to the reduction procedure.

2.6 The reduction of the Toda equations

In this section we show how the symmetries of the ordinary and extended Dynkin diagrams discussed above can be used to obtain new integrable non-linear equations from any given TM or TL equation. In the case of the TM equations, we can obtain <u>all</u> the TM equations, together with their general solution and associated linear problem starting from those corresponding to simply laced Dynkin diagrams (ie. those with only single links). Using the same procedure, which we call reduction, following the nomenclature of Zakharov and Mikhailov in a more general context [18], we also obtain all the TL equations from those corresponding to simply laced extended Dynkin diagrams. In this case, however, we also obtain a whole new class of integrable equations which we call the generalised Toda lattice (GTL) equations. The simplest example of these is the Bullough- Dodd equation mentioned before, eqn.(2.2.5). In fact one obtains all GTL equations (as well as all the TL equations) by reduction of those TL equations corresponding to simply laced extended Dynkin diagrams.

The reduction procedure is perfectly simple. We noted earlier that if a permutation p(a) of the points of the Dynkin diagram yielded a symmetry of that diagram then the variables $\{\rho_{p(a)}\}$ satisfied the corresponding Toda equations if the variables $\{\rho_a\}$ did. It follows that if we equate ρ_a to $\rho_{p(a)}$ many of the equations become identical so that the original set is still consistently satisfied. Thus we have a set of equations in fewer independent variables which can be recast in the Toda form (2.2.1), and therefore corresponds to a new matrix C_{ab} with fewer rows and columns.

As an example consider the SU(4)TM equation. The only non-trivial symmetry of $D(SU(4))$ is the generator of the Z_2 symmetry that is $\Gamma(SU(4))$, (1,3) in cycle notation using the labelling in Table I. Correspondingly we set $\rho_1 = \rho_3 = \rho_{13}$, say, in the SU(4) TM equations and obtain

$$-\partial^2\rho_{13} = 2e^{\rho_{13}} - e^{\rho_2}$$
$$-\partial^2\rho_2 = 2e^{\rho_2} - 2e^{\rho_{13}}$$

together with the vanishing of $\rho_2 + 2\rho_{13}$. These reduced equations are the SO(5) TM equations.

The reduced equations always inherit the integrability property since we need only substitute $\phi_a = \phi_{\rho(a)}$ in the gauge potentials(2.4.11) (or (2.4.24)) in order to obtain the gauge potentials for the reduced equations.

Note that the V_Δ part of A_u(see equations(2.4.11) and(2.5.1)) is simply

$$\sum_{a\in\Delta} E_a e^{\frac{1}{2}\rho_a} \qquad (2.6.1)$$

Hence in our example, the coefficient of exp (ρ_{13} / 2) is $E_{13} = E_1 + E_3$. If we define $H_{13} = [E_{13}, E_{-13}] = H_1 + H_3$ we find that indeed $E_{\pm 13}$, $E_{\pm 2}$, H_2 and H_{13} satisfy equations(2.4.8) with the SO(5) Cartan matrix. Thus the reduced gauge potentials do indeed correspond to the gauge potentials for the reduced equations. This has worked because in the course of the reduction we have constructed an SO(5) sub-algebra of SU(4) which is not regular, i.e. its roots are not all roots of SU(4).

Notice too how the Dynkin diagram D(SO(5)) is obtained from D(SU(4)) by folding the points 1 and 3 onto each other and drawing an arrow towards them.

The general procedure is well illustrated by the above example providing we avoid reductions arising from a symmetry which transposes two distinct points which are directly linked in the original Dynkin diagram. Such reductions we call non-direct reductions and defer their discussion until later. The direct reductions (of which the illustration is an example) are sufficiant to yield all the TL and GTL equations from the TL equations for simply laced Dynkin diagrams in one and only one way.

In a direct reduction a point of the original Dynkin diagram D is identified either with itself or another point to which it is not directly linked. The diagram D* for the reduced equations is obtained by folding D in such a way that distinct identified points are superposed while a point identified with itself is left alone. If a pair of adjacent points of D are superposed with another pair we retain only one set of links between them in D*, deleting the others. All other links of D remain in D*, and arrows are drawn (if not already there) towards the points which are superposed and away from the points identified with themselves.

First we discuss the direct reduction of the TM equations, associated with ordinary Dynkin diagrams, Table I. The symmetries form the group Γ (G) listed in the fourth column of Table I. When $\Gamma(G) \cong 1$ only the trivial reduction is possible and is not listed. When $\Gamma(G) \cong Z_2$ one non-trivial reduction is possible and the result is listed in Table IV as obtained by the rules above. $\Gamma(SO(8)) \cong S_3$ and the result of the symmetry of order 2 is included in the SO(2r) series while the result of the symmetry of order 3, $D^* = D(G_2)$ is listed separately. Notice that our claim is substantiated - in the result column each non-simply laced D(G) is obtained once and only once. The simply laced D(G) can be regarded as the result of the trivial reduction corresponding to the identity.

We have not listed any non-direct reductions such as the Z_2 symmetry applied to SU(2r + 1) as these will be discussed later.

Now we consider direct reductions of the TL equations. Here $\overline{D}(G)$ has symmetries forming a group $\overline{\Gamma}(G)$ which in general is non-abelian and consists of both inner and outer automorphisms in the sense explained in the preceding section. Conjugate elements of $\overline{\Gamma}(G)$ yield the same reduction so that we need only consider one representative element from each conjugacy class. This can be seen as follows. If we consider a permutation $a \rightarrow p(a)$ the reduction consists in setting $\rho_a = \rho_{p(a)}$. The reduction corresponding to a conjugate permutation $c^{-1}pc(a)$ will set $\rho_{c(a)} = \rho_{pc(a)}$. But $c(a)$ labels all indices just as a does so the two reductions are identical.

Table IV indicates how each non-simply laced ordinary Dynkin diagram $D(G')$ can be obtained by a unique direct reduction of a simply laced Dynkin diagram $D(G)$ say based on a non-trivial element of $\Gamma(G)$, the symmetry group of $D(G)$. $\Gamma(G)$ is a sub group of $\overline{\Gamma}(G)$ as it corresponds to those symmetries of $\overline{D}(G)$ leaving the point 0 fixed, so the same element can be used for a reduction of the extended diagram $D(G)$. It can be checked that this always yields $\overline{D}(G')$. Hence Table IV can be immediately extended to extended diagrams and we shall not list these direct reductions separately.

As we explained in section 4 the symmetries of $\overline{D}(G)$, namely $\overline{\Gamma}(G)$, are richer than those of $D(G)$ and direct reductions based on the extra elements may also be considered. These are listed in Table V and it is seen that each of the generalised Dynkin diagrams of Table III is obtained in precisely one way. This is our main result since it established that the Toda equations (2.2.1) based on the generalised diagrams are also integrable, inheriting their integrability in the reduction procedure.

Direct reductions of the non-simply laced diagrams yield no new diagrams and are not listed (Note that $\overline{D}(SU(2))$ should be counted as simply laced as the two roots are of equal length). A direct reduction of $\overline{D}(SU(N))$ based on an element of the cyclic sub group Z_N of (SU(N)) of order a yields D(SU(N/a)).

Note that the Bullough-Dodd equation (2.2.5) corresponding to the diagram GD(BD) is obtained as a simple reduction of SO(8). It appears as the lowest member of the H_r series in Table III.

To see more clearly how integrability is inherited denote the sets of points identified by the symmetry considered as Σ, Σ', ...etc. Hence if $\rho_a = \rho_\Sigma$, $a \in \Sigma$, the coefficient of $\exp\frac{1}{2}\rho_\Sigma$ in expression (2.6.1) is

$$E_\Sigma = \sum_{a\in\Sigma} E_a \tag{2.6.2}$$

We see that

$$[E_\Sigma, E_{-\Sigma'}] = \delta_{\Sigma\Sigma'} H_\Sigma \tag{2.6.3}$$

where
$$H_\Sigma = \sum_{a\in\Sigma} H_a$$

Then using the fact that in a direct reduction $K_{ab} = 0$ for $a, b \in \Sigma$, $a \neq b$, we have

$$[H_\Sigma, E_\Sigma] = 2E_\Sigma$$

which means that H_Σ and E_Σ are correctly normalised. Then for Σ' distinct from Σ,

$$[H_\Sigma, E_{\Sigma'}] = E_{\Sigma'} K_{\Sigma'\Sigma}$$

defines the Cartan matrix associated with the reduced Dynkin diagram. The compatability condition can be evaluated just as in section 2.4 but using instead of H_a and $E_{\pm a}$ the H_Σ and $E_{\pm\Sigma}$ defined by equations (2.6.2) and (2.6.3) since we now have the appropriate generalisations of equations (2.4.8). We see that the reduced gauge potentials are the gauge potentials for the reduced Toda equations.

This completes our discussion of what we have called the direct reductions.

Less clear cut is the discussion of the non-direct reductions stemming from symmetries of the original Dynkin diagram in which linked points are transposed. The classic example, due to Fordy and Gibbons [19] and Mikhailov [20] is the reduction of the SU(3)TL equations based on $\overline{D}(SU(3))$. This has a Z_2 symmetry (12), using the labelling of Table II which leads us to set $\rho_1 = \rho_2 = \rho_{12}$ so that the equations reduce to

$$\partial^2\rho_0 = -2e^{\rho_0} + 2e^{\rho_{12}}$$
$$\partial^2\rho_{12} = -e^{\rho_{12}} + e^{\rho_0}$$

This is not in the form(2.2.1) as the coefficient of $\exp(\rho_{12})$ in the second equation does not equal minus 2, because of the non-direct nature of the reduction. Therefore we must set instead $\rho_1 = \rho_2 = \rho_{12} + \ln 2$ so that we now have

$$\partial^2\rho_0 = -2e^{\rho_0} + 4e^{\rho_{12}}$$
$$\partial^2\rho_{12} = -2e^{\rho_{12}} + e^{\rho_0}$$

which are the equations corresponding to the Cartan matrix(2.2.4). However the original variables satisfied the constraint(2.4.26).

$$\sum_{a\in\overline{\Delta}} N_a\rho_a = \rho_0 + \rho_1 + \rho_2 = 0$$

which yields for the new variables the constraint

$$\rho_1 + 2\rho_{12} = \ln 4$$

instead of zero which is the canonical value used above. Setting $\rho = \rho_{12}$ we obtain

$$\partial^2\rho = -2e^{\rho} + \tfrac{1}{4}e^{-2\rho}$$

which is not the Bullough-Dodd equation(2.2.5). This equation can be obtained after a further translation of ρ and a rescaling of the space-time variables,

$$\rho_1' = \rho + {}^2/_3 \ln 2 \qquad x^{\mu'} = 2^{-1/3} x^\mu$$

We conclude that the Bullough-Dodd equation in the form(2.2.5) is obtained far more naturally as a direct reduction of the TL equation for SO(8) as mentioned above rather than as a non-direct reduction of the TL equation for SU(3), which is how it was obtained previously [19,20] as the first example of the reduction procedure.

Notice that the coefficient of exp ($\frac{1}{2}\rho_{12}$) in(2.6.1) is

$$E_{12} = \sqrt{2}(E_1 + E_2)$$

and that

$$[E_{12}, E_{-12}] = H_{12} = 2(H_1 + H_2)$$

It can be checked that $E_{\pm 0}, H_0, E_{\pm 12}$ and H_{12} generate the algebra (2.4.8) with the Cartan matric(2.2.4). This illustrates the fact that non--direct reductions serve equally well in obtaining special sub algebras, but with extra normalisation factors arising compared with(2.6.2) and(2.6.3).

The above example typifies what happens in other non-direct reductions. The reduced equations correspond to Dynkin diagrams already discussed and can be put in the canonical form only after a rescaling of the space-time variables. Since no new diagrams result we shall not list the possibilities here but merely state the extra rule that must be added to the previous ones in order to obtain the reduced Dynkin diagram: if two linked points are superposed then leave the link (which has to be single) as a loop attached to the point. This loop then signifies that the number of lines pointing towards the point with the attached loop should be doubled. Finally the loop is deleted.

If the original Toda equation reads as(2.2.1) with $\sum N_\alpha \rho_\alpha$ vanishing where N_α is the left null vector of C then the reduced Toda equations will read

$$-\partial^2 \rho_\Sigma = C^*_{\Sigma\Sigma'} e^{\rho_\Sigma} \tag{2.6.4}$$

with

$$\Sigma N^*_\Sigma \rho_\Sigma = R$$

where C*, the Cartan matrix corresponding to the reduced diagram explained above, has left null vector N^*_Σ with positive integer components. R will be a constant depending on the reduction and will vanish only for direct reductions. The substitution

$$\rho_\Sigma \rightarrow \rho^*_\Sigma + R/(\Sigma N^*_\Sigma)$$
$$x^\mu \rightarrow \exp\left(R/2(\Sigma N^*_\Sigma)\right) x^\mu$$

will effectively reduce R to zero while maintaining (2.6.4).

APPENDIX

We saw that

$$\mathrm{aut}\,\overline{\Delta}/W_0 \cong \mathrm{aut}\,\Delta \tag{A.1}$$

and we now wish to show that the group W_0 consisting of those elements of the Weyl group which permute the extended system of simple roots, $\overline{\Delta}$ is isomorphic to the centre of the universal covering group. We regard this as a somewhat surprising result and can prove it only considering each simple Lie group in turn.

Since $\mathrm{aut}\,\Delta$ and $\mathrm{aut}\,\overline{\Delta}$ are isomorphic to Γ and $\overline{\Gamma}$ respectively, they can be read off the last column of Tables I and II. We see that $\Gamma \cong 1$ for $SU(2)$, $SO(2r+1)$ $r \geq 3$, $Sp(2r)$ $r \geq 2$, E_7, E_8, F_4 and G_2. It follows from (A.1) that $W_0 \cong \overline{\Gamma}$ in these cases and from Tables I and II one sees that $\overline{\Gamma} \cong Z(\tilde{G})$ and the desired result is established.

This leaves SU(r+1), $r \geq 2$; SO(2r), $r \geq 4$ and E_6. The result for SU(r+1) was established in the text while E_6 has $\overline{\Gamma} \cong S_3$ and $\Gamma \cong Z_2$ so that by (A.1) W_0 has order 3 and hence must equal Z_3 as that is the unique group of order 3. Since the covering group of E_6 has centre Z_3 the result is established leaving only the series SO(2r), $r \geq 4$. This is particularly interesting because for r=4 $\overline{\Gamma} \cong S_4$, $\Gamma \cong S_3$ while for $r \geq 5$ $\overline{\Gamma} \cong D_4$, $\Gamma \cong Z_2$ so that equation (A.1) tells us that W_0 is of order 4. There are two groups of order 4, $Z_2 \times Z_2$ and Z_4 and in fact $Z(\widetilde{SO(2r)})$ alternates between them for r even and odd respectively. We now show by computation that W_0 alternates similarly, thereby completing the proof of the desired result.

If $e_1, e_2, . e_r$ denote r orthogonal unit vectors the roots of SO(2r) can be expressed as

$$\Phi(SO(2r)) = \{\pm e_i \pm e_j \,,\; i \neq j\}$$

while, using the numbering of simple roots adopted in Tables I and II

$$\Delta(SO(2r)) = \{\alpha^0, \alpha^1, \alpha^2 \ldots \alpha^r\}$$

$$= \{-e_1 - e_2, e_1 - e_2, e_3 - e_4 \ldots e_{r-1} - e_r, e_{r-1} + e_r$$

A reflection in the root $e_i - e_j$ interchanges e_i and e_j while a reflection in $e_i + e_j$ reverses the signs of e_i and e_j. It follows that the Weyl group of SO(2r) consists of all permutations of the e_i combined with all possible even reflections.

The reflection σ_{1r} ,say, in the root $e_1 + e_r$ reverses e_1 and e_r thereby interchanging α^0 with α^1 and α^{r-1} with α^r whilst leaving the other simple roots fixed. It therefore constitutes an element of W_0 of order 2 which can be written using the notation of section 2 (whereby an element of $\overline{\Gamma}$ is denoted as a permutation of the crossed points in cycle notation)

$$\sigma_{1r} = (0, 1)(r-1, r) \tag{A.2}$$

Now we must distinguish the cases r even and odd. In the former case $e_i \to -e_{r+1-i}$ is a permutation of the e_i combined with an even reflectio and hence an element of $W(SO(2r))$, σ , say . We see that $\sigma(\alpha^a) = \alpha^{r-a}$ a = 0,1.....r so that σ is an element of W_0 of order 2 which, in the notation of section 2, is

$$\sigma = (0, r)(1, r-1)$$

Thus when r is even W_0 consists of the group $Z_2 \times Z_2$ generated by σ and σ_{1r} (eqn. A.2) and so

$$W_0(SO(2r)) \cong \{ (\,);\ (0,1)(r-1,r);\ (0,r)(1,r-1);\ (1,r)(0,r-1) \} \qquad r \text{ even}$$

and is indeed isomorphic to the centre as claimed.

On the other hand if r is odd consider instead

$$e_1 \to e_r,\ e_2 \to -e_{r-1},\ e_3 \to -e_{r-2},\ \ldots\ \ e_r \to -e_1$$

This is again a permutation of the e_i followed by an even reflection and hence an element of W, $\bar{\sigma}$, say. Then

$$\bar{\sigma}(\alpha^0) = \alpha^{r-1}\ ;\quad \bar{\sigma}(\alpha^1) = \alpha^r\ ;\quad \bar{\sigma}(\alpha^a) = \alpha^{r-a} \quad r \geq a \geq 2$$

so that $\bar{\sigma}$ is an element of W_0 which in cycle notation is $\bar{\sigma} = (0, r-1, 1$
Comparing with (A.2) we see

$$\bar{\sigma}^2 = \sigma_{1r}$$

so that $\bar{\sigma}$ has order 4 and hence generates Z_4 which must be W_0. Thus W_0 is again isomorphic to the centre of $\widetilde{SO(2r)}$ for r odd and

$$W_0 = \{ (\,);\ (0,r-1,1,r);\ (0,1)(r-1,r)\ ;\ (0,r,1,r-1) \}$$

The above analysis includes SO(8).

Table I: Ordinary Dynkin Diagrams

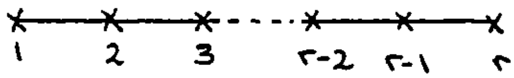
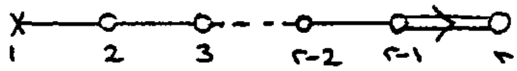
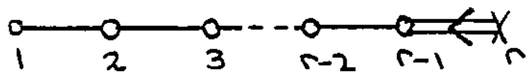
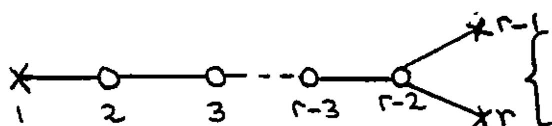
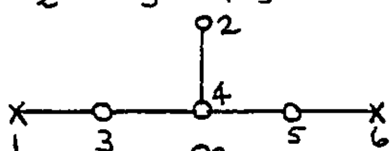
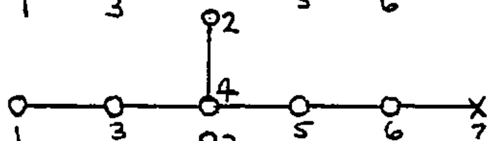
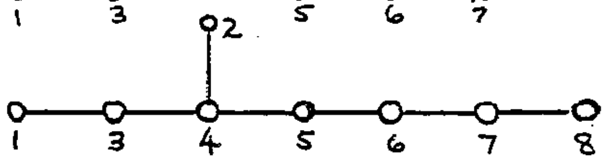
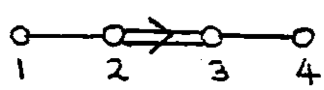
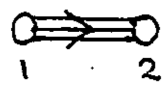
G	D(G)	Z(G)	Γ(G)
SU(r+1), $r \geq 1$		Z_{r+1}	r = 1; 1 $r \geq 2$; Z_2
SO(2r+1), $r \geq 2$		Z_2	1
Sp(2r), $r \geq 3$		Z_2	1
SO(2r), $r \geq 4$		r even; $Z_2 \times Z_2$ r odd; Z_4	r = 4; S_3 $r \geq 5$; Z_2
E_6		Z_3	Z_2
E_7		Z_2	1
E_8		1	1
F_4		1	1
G_2		1	1

Table II: Extended Dynkin Diagrams

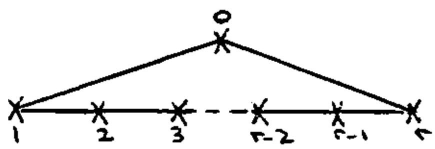
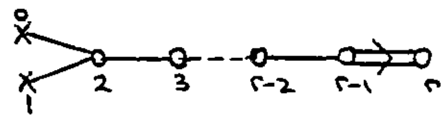
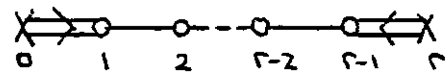
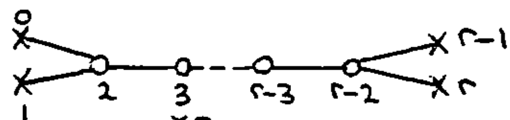
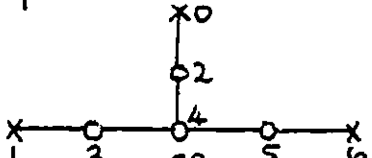
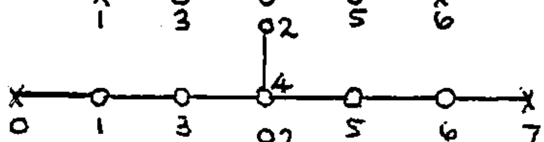
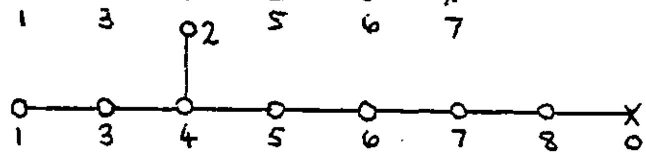
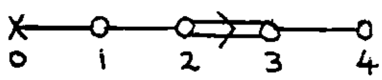
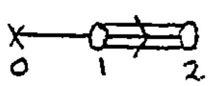
G	$\overline{D}(G)$	$\overline{\Gamma}(G)$
$SU(2)$		Z_2
$SU(r+1)$, $r \geqslant 2$		D_{r+1}
$SO(2r+1)$, $r \geqslant 2$		Z_2
$Sp(2r)$, $r \geqslant 3$		Z_2
$SO(2r)$, $r \geqslant 4$		$r = 4$; S_4 $r \geqslant 5$; D_4
E_6		S_3
E_7		Z_2
E_8		1
F_4		1
G_2		1

Table III: Generalised Dynkin Diagrams

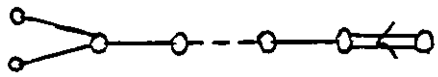
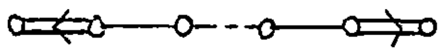
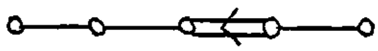
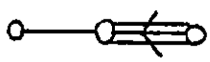
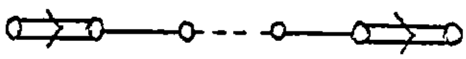
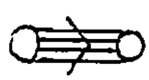
Name	GD	Symmetry Group
$\bar{D}^T(SO(2r+1))$, $r \geqslant 3$		Z_2
$\bar{D}^T(Sp(2r))$, $r \geqslant 3$		Z_2
$\bar{D}^T(F_4)$		1
$\bar{D}^T(G_2)$		1
$GD(H_r)$		1
GD(BD)		1

Table IV: Direct Reduction of TM Equations for Simply Laced D(G)

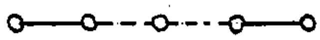
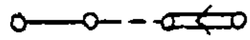
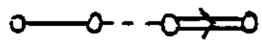
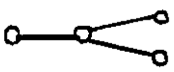
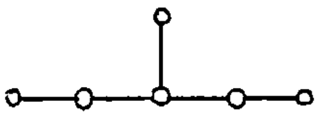
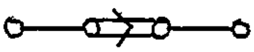
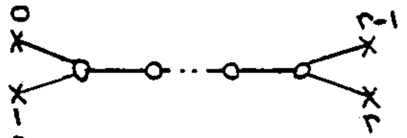
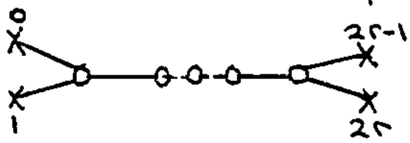
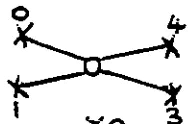
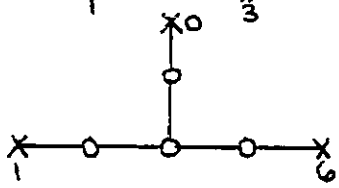
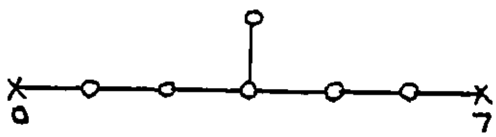
G	D(G)	D*(G)	
SU(2r+2), $r \geqslant 1$			= D(Sp(2r+2))
SO(2r), $r \geqslant 4$			= D(SO(2r-1))
SO(8)			= D(G_2)
E_6			= D(F_4)

Table V: Direct Reductions of TL Equations for Simply Laced D(G) Omitting Those Based on Γ(G)

G	$\tilde{D}(G)$	Symmetry used (cycle notation)	D*(G)
SO(2r), r ⩾ 4		(r-1,r)(0,1)	$D^T(Sp(2r-4))$
SO(4r), r ⩾ 3		{ (0,2r-1)(1,2r) { (0,2r-1,1,r)	$D^T(SO(2r+1))$ GD(H_r)
SO(8)		(0,1,3,4)	GD(BD)
E_6		(0,1,6)	$D^T(G_2)$
E_7		(0,7)	$D^T(F_4)$

CHAPTER III

THE ALGEBRAIC STRUCTURE OF TODA SYSTEMS

3.1 Introduction

For a variety of reasons, both mathematical and physical, there is much current interest in dynamical systems which are integrable classically, or even better, quantum mechanically.[1,2] The systems usually considered have a number of degrees of freedom which is either finite (so that the system is called one-dimensional, with time the dimension) or infinite with extra degrees of freedom counted by a single space parameter (so that the system is called two dimensional with time and space the dimensions). A particle physicist is interested in relativistically invariant two dimensional systems and even more in four dimensional systems. There already exist many interesting relationships[3,4] between self-dual gauge field configurations describing instantons or monopoles and integrable systems in one or two dimensions.

Integrable systems exhibit much interesting structure, for example, unexpected conservation laws (or symmetries) which restrict the phase space through which the classical system may evolve. In the two dimensional systems there occur other objects with a simple time dependence which relate to the Bethe Ansatz valid in some quantum statistical systems. It is of great importance to evaluate the Poisson or quantum commutator bracket between these objects.

Underlying the integrability is a Lax pair[5] or zero curvature[6] condition involving a second type of commutator structure involving what we shall call Lie brackets in order to distinguish them from the Poisson or quantum brackets previously mentioned. It seems that the proper understanding of the theory involves the interrelationship between these two sets of brackets, the Poisson or quantum bracket and the Lie bracket. Major progress in this direction has been made by the Leningrad school[1,2,7] in the course of developing the "quantum inverse scattering method' using ideas from models in statistical physics. In specific two dimensional integrable systems a crucial role is played by an operator $\mathbb{P}$ which relates the two bracket structures.

Our aim in this chapter is to construct and understand $\mathbb{P}$ for two (infinite) families of theory of interest for both physical and mathematical reasons. The best understood Lie bracket structures correspond to the simple Lie algebras [8] (classified by Cartan) and the affine Euclidean Kac-Moody algebras [9] and these are the ones relevant to the two families of theory.

The corresponding equations are known as the Toda equations[10] and couple coordinates ρ_a in a nonlinear way.

$$\frac{d^2}{dt^2}\rho_a = -\sum_{b=1}^{R} C_{ab}\, e^{\rho_b} \tag{3.1.1}$$

The R x R matrix C is a special sort of matrix with integer entries, called a Cartan matrix.[8] These matrices give the clue to the Lie algebraic structure mentioned above since they encode the structure of that algebra in a way to be explained.

When C is a Cartan matrix associated with simple Lie algebra equations (3.1.1) are called "Toda molecule" equations[11] whereas when C is a Cartan matrix associated with an affine Euclidean Kac-Moody algebra,[9] the equations (3.1.1) are known as "Toda lattice" equations. The possible matrices C are known from the classification theory of these algebras and can be denoted by Dynkin diagrams which are listed and explained above. Toda' original equation corresponded to the large N limit of the SU(N) algebra.[10]

We shall just consider the classical one dimensional systems (3.1.1). The form of $\mathbb{P}$ we find leads us to think it will also play a role in the quantum and two dimensional versions of (3.1.1). The two dimensional versions of(3.1.1) are relativistically invariant and therefore of interest to particle physicists. The simplest (single component) member of these two families are respectively the Liouville equation and the Sinh-Gordon equation.

The algebraic structures mentioned are respectively finite and infinite dimensional but nevertheless both possess a finite dimensional space of mutually commuting generators H_i (called a "Cartan subalgebra"). The remaining generators can be arranged as step operators for roots, a

$$[H_i, E_a] = a_i E_a$$

The roots a span an R dimensional space and can be split into two subsets, the positive roots, and their negatives. The positive roots can always be expressed as sums of R "simple roots" whose scalar product specifies the Cartan matrix occurring in the Toda equations(3.1.1)[8,9]

$$C_{ab} = 2a\cdot b/b^2 , \qquad a,b \text{ simple roots} \tag{3.1.2}$$

The general theory guarantees that the complete algebra can be reconstructed (up to an isomorphism) from the Cartan matrix (or Dynkin diagram), which therefore contains all relevant information.

The Lie bracket structure enters in the formulation of the Lax pair equations (see (3.2.1) in section 2). This equation ensures that the matrix A of the Lax pair develops in time in such a way that the quantities $Tr(A^N)$ remain constant. The interplay between the two types of bracket enters when the Poisson bracket between different matrix elements of A is evaluated. When this can be expressed as a Lie bracket involving the operator $\mathbb{P}$ (see equation (3.2.3)) it follows that the conserved quantities $Tr(A^N)$ have vanishing mutual Poisson brackets. Any of these quantities may be considered as a new Hamiltonian without disturbing these properties since an appropriate Lax pair can be formed from A and $\mathbb{P}$ as explained in section 2.

In section 3, $\mathbb{P}$ is constructed for the Toda molecule equations in the form

$$\mathbb{P} = -\frac{1}{2} \sum_{\text{roots } a} (\text{sign} a)\, E_a \otimes E_{-a} \tag{3.1.3}$$

where E_a is the suitably normalized step operator for the root a and (sign a) is ±1 according as the root a is positive or negative. The sum extends over all the roots of the Lie algebra defined by the Cartan matrix occurring in (3.1.1).

In section 4, $\mathbb{P}$ is constructed for the Toda lattice equations and shown to have precisely the same form (3.1.3) providing the sum extends over the infinite system of roots of the affine Euclidean Kac-Moody algebra associated with the Cartan matrix in (3.1.1) and the step operators are those appropriate to the centre-free version of the algebra, called the loop algebra, and written as generators of the associated finite dimensional Lie algebra times powers of the spectral parameter as explained in section 4.

Equation (3.1.3) stating that $\mathbb{P}$ has a transparent uniform structure for the two families of equation considered constitutes our main result. The form (3.1.3) applies equally to the equations besides (3.1.1) based on Hamiltonians $\mathrm{Tr}(A^N)$, $N > 2$ instead of $\mathrm{Tr}(A^2)$. We think it striking that P depends just on the relevant root system. In deriving (3.1.3) we used in an essential way the general feature of root systems that a positive root minus a simple root cannot be a negative root. We therefore expect our construction to generalize to possible more general infinite algebraic structures of this kind which may underlie other dynamical systems. Simple examples are the Toda equations involving generalized Cartan matrices and corresponding to the remaining Euclidean Kac-Moody algebras,[12,13] i.e. the twisted ones.

Section 5 discusses the result mentioning that (3.1.3) should apply equally to the quantum case. Comparison is made with another sort of integrable dynamical system with a Lie group theoretical basis, namely that of a "free particle on a Lie group". These systems possess a structure like $\mathbb{P}$ but are not completely integrable.

3.2 Lax Pairs and the $\mathbb{P}$ Operator

Consider a classical dynamical system with a finite number of degrees of freedom (such as the Toda systems (3.1.1)) with a Lax Pair A, B depending on the canonical variables (q_a, p_a) such that

$$\frac{dA}{dt} = i[B,A] \tag{3.2.1}$$

whenever the Hamiltonian equations of motion hold. Then the quantities $Tr(A^N)$ are conserved in time. If A and B are generators of a simple Lie algebra we expect the number of independent constants of the motion of the form $Tr(A^N)$ to equal the rank of the algebra. The important question is to evaluate the Poisson bracket, (PB), of the different constants. If the PB of two quantities vanish they are said to be in involution. This allows a canonical transformation to coordinates where the $Tr(A^N)$ are canonical momenta and the system is then (formally) integrated providing the number of independent $Tr(A^N)$ equals the number of degrees of freedom.[14)]

The first stage is evidently to evaluate the PB of two distinct matrix elements of A, A_{ij} and $A_{k\ell}$ say. Then the defining property of $\mathbb{P}$ is that it obeys [2)]

$$\begin{aligned}\{A_{ij},A_{k\ell}\}_{PB} &= \sum_a \frac{\delta A_{ij}}{\delta q_a}\frac{\delta A_{k\ell}}{\delta p_a} - \frac{\delta A_{ij}}{\delta p_a}\frac{\delta A_{k\ell}}{\delta q_a} \\ &= \mathbb{P}_{ik;i'\ell} A_{i'j} + \mathbb{P}_{ik;jk'} A_{k'\ell} \\ &\quad - A_{ij'}\mathbb{P}_{j'k;j\ell} - A_{k\ell'}\mathbb{P}_{i\ell';j\ell}\end{aligned} \tag{3.2.2}$$

or in a more compact notation[2)]

$$\{A \underset{,}{\otimes} A\}_{PB} = [\mathbb{P}, A \otimes 1 + 1 \otimes A] \tag{3.2.3}$$

where products between matrices in the tensor product space (of the Lie algebra with itself) are defined so that

$$(A \otimes B)\ (C \otimes D) = AC \otimes BD \quad . \tag{3.2.4}$$

We shall talk of left and right entries in an obvious way. Because of the antisymmetry of the PB we expect that $\mathbb{P}$ will be antisymmetric with respect to interchange of the left and right entries and will write, corresponding to this,

$$\mathbb{P}^T = -\mathbb{P} \ . \tag{3.2.5}$$

Further, since A is hermitian (so the conserved quantities are real), we expect

$$\mathbb{P}^+ = -\mathbb{P} \tag{3.2.6}$$

when left and right entries are conjugated separately without changing positions. Our aim is to construct $\mathbb{P}$ satisfying (3.2.3), (3.2.5) and (3.2.6) but before doing so we wish to discuss the consequences of the existence of such a $\mathbb{P}$. From (3.2.3),

$$\begin{aligned}\{Tr(A^N), A\}_{PB} &= N\ Tr_L(A^{N-1} \otimes 1\ \{A \overset{\otimes}{,} A\}_{PB}) \\ &= N\ Tr_L(A^{N-1} \otimes 1\ [\mathbb{P}\ , A \otimes 1 + 1 \otimes A])\end{aligned}$$

where Tr_L indicates trace on the indices of the left entry. The first term vanishes while the second yields

$$= N[Tr_L(A^{N-1}\ \mathbb{P}\ ),\ A] \tag{3.2.7}$$

From this it immediately follows that

$$\{Tr(A^N),\ Tr(A^M)\}_{PB} = 0 \tag{3.2.8}$$

i.e., the constants of the motion $Tr(A^M)$ are all in involution. The Hamiltoni for the system is usually taken as $Tr(A^2)$, as is the case for the Toda systems

Then(3.2.7) tells us, putting $N = 2$, that

$$-\frac{dA}{dt} = \{H,A\}_{PB} = 2[\mathrm{Tr}_L(A\,\mathbb{P}),A]$$

from which we deduce that B, the second element of the Lax pair in(3.2.1) is just $2i\mathrm{Tr}_L(A\mathbb{P})$. The validity of(3.2.3) and hence(3.2.7) does not depend on any particular choice of Hamiltonian once A is expressed in canonical variables. Instead of $\mathrm{Tr}(A^2)$ we should choose $\mathrm{Tr}(A^N)$ as Hamiltonian (or even any linear combination of these quantities) and equation(3.2.7) would then imply

$$-\frac{dA}{dt} = \{H,A\}_{PB} = \{\mathrm{Tr}(A^N),A\}_{PB}$$
$$= N\,[\mathrm{Tr}_L(A^{N-1}\mathbb{P}),A]$$

Hence for the system with $H = \mathrm{Tr}(A^N)$ we have the Lax pair A and

$$B_N = iN\,\mathrm{Tr}_L(A^{N-1}\,\mathbb{P}) \tag{3.2.9}$$

The same quantities $\mathrm{Tr}(A^N)$ would be constants of the motion in involution. We see that $\mathbb{P}$ is more fundamental than B in the sense that it is independent of the dynamical coordinates and universal in the sense that it applies equally to all the dynamical systems based on linear combinations of the $\mathrm{Tr}(A^N)$. The goal of this paper is to construct $\mathbb{P}$ for the Toda systems(3.1.1) mentioned above and show it has a remarkably simple and transparent structure. We shall proceed in two steps dealing first with the TM equations and then using the result to obtain $\mathbb{P}$ for the TL equations.

3.3 Poisson brackets and the $\mathbb{P}$ operator for the Toda molecule equations

In this section we consider the Toda molecule equations. Thus C in (3.1.1) is the Cartan matrix K_{ab} for a simple Lie algebra. We start with the zero curvature condition for the two dimensional version of equation(3.1.1)[15] in the form presented by Leznov and Saveliev.[16] We then proceed in 3 steps: (i) construction of A (see eq. 2.1) in terms of canonical coordinates and momenta for the one dimensional case, (ii) evaluation of the Poisson bracket between the different matrix elements of A, and (iii) construction of $\mathbb{P}$ satisfying (3.2.3).

We refer the reader to the previous chapter for definitions of potential A_μ and other notation. According to that we take

$$A = A_x = \frac{1}{2}(A_u + A_v) = \sum_{a\varepsilon\Delta} \frac{1}{2}\dot{\phi}_a H_a + \frac{1}{2} e^{\frac{1}{2}K_{ab}\phi_b} (E_a + E_{-a}) \tag{3.3.1}$$

$$-iB = A_t = \frac{1}{2}(A_u - A_v) = \sum_{a\varepsilon\Delta} \frac{1}{2} e^{\frac{1}{2}K_{ab}\phi_b} (E_a - E_{-a}) \tag{3.3.2}$$

where H_a are the Cartan subalgebra generators, E_a the step operators for the simple roots α^a in the Chevalley basis and we have assumed ϕ is independent of x in order to obtain the one dimensional equations (3.1.1). The sums extend over the set of simple roots Δ. Then

$$\begin{aligned} \frac{dA}{dt} - i[B,A] &= \frac{1}{2} \sum_{a\varepsilon\Delta} H_a(\ddot{\phi}_a + e^{K_{ab}\phi_b}) \\ &= \frac{1}{2} \sum_{a\varepsilon\Delta} H_a K^{-1}_{ac} (\ddot{\rho}_c + K_{cb} e^{\rho_b}) \end{aligned} \tag{3.3.3}$$

using the commutators in the Chevalley basis:

$$\begin{aligned} [H_a, E_b] &= E_b K_{ba} && \text{(no sum over b)} \\ [E_a, E_{-b}] &= H_a \delta_{ab} && \alpha^a, \alpha^b \varepsilon \Delta \end{aligned} \tag{3.3.4}$$

and letting $\rho_a = K_{ab}\phi_b$. We have used the fact that the matrix K_{ab} is non-singular for any simple Lie algebra.

Thus A and B given by (3.3.1) and (3.3.2) constitute a Lax pair for the Toda molecule equations (3.1.1). Normalizing the Killing form by

$$\mathrm{Tr} E_a E_{-b} = \frac{2\delta_{ab}}{(\alpha^a)^2} \quad ; \quad \mathrm{Tr}(H_a H_b) = \frac{4\alpha^a \cdot \alpha^b}{(\alpha^a)^2 (\alpha^b)^2} \tag{3.3.5}$$

we find that the energy integral $Tr(A^2)$ has the kinetic term

$$K = \sum_{a,b} \dot{\phi}_a \dot{\phi}_b \frac{\alpha^a \cdot \alpha^b}{(\alpha^a)^2 (\alpha^b)^2}$$

so that the canonical momentum conjugate to ϕ_a is

$$\pi_a = \frac{\delta K}{\delta \dot{\phi} a} = \frac{1}{(\alpha^a)^2} K_{ab} \dot{\phi}_b \tag{3.3.6}$$

Since $H_a = \frac{2\alpha^a \cdot H}{(\alpha^a)^2}$ and α^a can be expanded in terms of fundamental weights λ^b as

$$\alpha^a = K_{ab} \lambda^b$$

which follows from the definition of λ^b;

$$\frac{2\lambda^a \cdot \alpha^b}{(\alpha^b)^2} = \delta^{ab}$$

we find

$$A(\phi,\pi) = \sum_{a\varepsilon\Delta} [\lambda_a \cdot H\pi_a + \frac{1}{2} e^{\frac{1}{2}K_{ab}\phi_b} (E_a + E_{-a})] \ . \tag{3.3.7}$$

This is the desired expression for A in terms of canonical variables and completes step (i). It is straightforward to evaluate the Poisson bracket(3.2.2

$$\{A \underset{,}{x} A\}_{PB} = \frac{1}{8} \sum_{a\varepsilon\Delta} (\alpha^a)^2 e^{\frac{1}{2}K_{ab}\phi_b} [(E_a + E_{-a}) \otimes H_a - H_a \otimes (E_a + E_{-a})] \tag{3.3.8}$$

Comparing with the defining equation for $\mathbb{P}$, 2.3) we see it is sufficient for $\mathbb{P}$ to satisfy

$$[\mathbb{P}, H_a \otimes 1 + 1 \otimes H_a] = 0 \tag{3.3.9}$$

$$[\mathbb{P}, E_a \otimes 1 + 1 \otimes E_a] = \frac{1}{4}(\alpha^a)^2 (E_a \otimes H_a - H_a \otimes E_a); \ \alpha^a \varepsilon\Delta \tag{3.3.10}$$

since if $\mathbb{P}$ is antihermitian in the sense of (3.2.6) it follows from (3.3.10) that

$$[\mathbb{P},\ E_{-a} \otimes 1 + 1 \otimes E_{-a}] = \frac{1}{4}(\alpha^a)^2 (E_{-a} \otimes H_a - H_a \otimes E_{-a}) \qquad (3.3.11)$$

The final stage, (iii) is therefore to construct $\mathbb{P}$ satisfying (3.3.9), (3.3.10), (3.2.5) and (3.2.6). The first step is to consider the Casimir like operator $\mathbb{C}$ defined by

$$\mathbb{C} = \sum_{i=1}^{\dim \underset{\sim}{g}} T_i \otimes T_i \qquad (3.3.12)$$

where the T_i constitute an orthonormal basis (with respect to the Killing form) for the generators of the compact Lie algebra $\underset{\sim}{g}$;

$$[T_i, T_j] = i\, f_{ijk}\, T_k\,, \quad f_{ijk} \text{ totally antisymmetric} \qquad (3.3.13)$$

$\mathbb{C}$ has the fundamental property that it commutes with any generator T in the sense that

$$[\mathbb{C},\ 1 \otimes T + T \otimes 1] = 0 \qquad (3.3.14)$$

as is easily seen using (3.3.13). It is useful to split $\mathbb{C}$ into three parts

$$\mathbb{C} = \mathbb{C}_+ + \mathbb{C}_- + \mathbb{C}_o \qquad (3.3.15)$$

where

$$\mathbb{C}_+ = \sum_{\alpha>o} (\alpha^2/2) E_\alpha \otimes E_{-\alpha} \qquad (3.3.16)$$

$$\mathbb{C}_- = \sum_{\alpha<o} (\alpha^2/2) E_\alpha \otimes E_{-\alpha} = \sum_{\alpha>o} (\alpha^2/2) E_{-\alpha} \otimes E_\alpha \qquad (3.3.17)$$

$$\mathbb{C}_o = \sum_{i=1}^{n} H_i \otimes H_i \qquad (3.3.18)$$

Here the E_α are the Chevalley step operators for all the roots α of the algebra , and H_i the Cartan subalgebra in the Cartan-Weyl basis. Together they obey

$$
\begin{aligned}
&[H_i,H_j] = 0 && i,j = 1 \ldots . r = \text{rank } \underset{\sim}{g} \\
&[H_i,E_\alpha] = \alpha_i E_\alpha && \\
&[E_\alpha,E_\beta] = N_{\alpha\beta}E_{\alpha+\beta} \quad \text{if } \alpha+\beta \text{ a root} && \\
&\qquad = \frac{2\alpha\cdot H}{\alpha^2} \quad \text{if } \alpha+\beta = 0 && \\
&\qquad = 0 \text{ if neither true.} &&
\end{aligned}
\tag{3.3.19}
$$

Note that $H_a \equiv 2\,\frac{\alpha^a\cdot H}{(\alpha^a)^2}$ and $E_a \equiv E_{\alpha^a}$ above.

The decomposition (3.3.15) is useful because as we shall show

$$\mathbb{P} = -\frac{1}{2}\,(\mathbb{C}_+ - \mathbb{C}_-) \tag{3.3.20}$$

This evidently satisfies the transposition and hermiticity properties (3.2.5) and (3.2.6). Since (3.3.9) is trivial to check it only remains to establish (3.3.10).

First note that

$$[\mathbb{C}_o, E_a \otimes 1 + 1 \otimes E_a] = \frac{(\alpha^a)^2}{2}\,(E_a \otimes H_a + H_a \otimes E_a) \tag{3.3.21}$$

while

$$[\mathbb{C}_+, E_a \otimes 1 + 1 \otimes E_a] = -\frac{(\alpha^a)^2}{2}\,(E_a \otimes H_a) + X \tag{3.3.22}$$

where X is a sum of terms proportional to $E_{\alpha'} \otimes E_{-\alpha}$ with α and α' both positive roots. This follows from (3.3.19) and the simple fact that a positive root minus a simple root can never be a negative root. Likewise

$$[\mathbb{C}_-, E_a \otimes 1 + 1 \otimes E_a] = -\frac{(\alpha^a)^2}{2}\;H_a \otimes E_a + X' \tag{3.3.23}$$

where X' is a sum of terms proportional to $E_{-\alpha'} \otimes E_{\alpha}$, with α and α' both positive roots. But (3.3.14) tells us that $\mathbb{C}_o + \mathbb{C}_+ + \mathbb{C}_-$ commutes with

$E_a \otimes 1 + 1 \otimes E_a$. Thus by (3.3.21), (3.3.22), (3.3.23), $X + X'$ vanishes. Since X and X' cannot cancel as their components are linearly independent they must vanish individually. Subtracting (3.3.23) from (3.3.22) we then get

$$[\mathbb{C}_+ - \mathbb{C}_-, E_a \otimes 1 + 1 \otimes E_a] = -\frac{(\alpha^a)^2}{2}(E_a \otimes H_a - H_a \otimes E_a)$$

It follows that $\mathbb{P}$ given by (3.3.20) has the desired properties (3.3.9), (3.3.1 This $\mathbb{P}$ is unique because if not, we would have an operator commuting with all the $1 \otimes T_i + T_i \otimes 1$. The only such operator is $\mathbb{C}$ (3.3.12) which is symmetric and hermitian in the sense of (3.2.5) and (3.2.6) and so cannot be added to $\mathbb{P}$ without destroying the properties (3.2.5) and (3.2.6).

If we rescale the step operators $\sqrt{\frac{\alpha^2}{2}}\, E_{\pm\alpha} \to E_{\pm\alpha}$ so they become part of an orthonormal Cartan Weyl basis instead of a Chevalley basis, $\mathbb{P}$ takes the form (3.1.3) given in the introduction.

Once $\mathbb{P}$ is constructed it follows from the previous section that the integrals $\mathrm{Tr}(A^N)$ are all in involution. Since the number of independent such integrals equals $r = \text{rank } \underset{\sim}{g}$, the number of dynamical coordinates, we see the Toda molecule system is completely integrable in the sense of Arnold.[14)] This was known for $\underset{\sim}{g} = SU(N)$[17)] but as far as we know this is the first explicit proof which works uniformly for all simple Lie algebras.

3.4 Poisson brackets and the $\mathbb{P}$ operator for the Toda Lattice equations

Here the matrix C in (3.1.1) is an extended Cartan matrix $\bar{K}_{ab}$ associated with any simple Lie algebra $\underset{\sim}{g}$ by adjoining $\alpha^o = -\psi$ (the negative of the highes root ψ) to the set of simple roots Δ to obtain $\bar{\Delta} = \{\alpha^o, \Delta\}$ and defining $\bar{K}_{\bar{a}\bar{b}}$ as in (3.1.2) where $\bar{a}$ and $\bar{b}$ run over 0, 1.....r. By the properties of roots and coroots of $\underset{\sim}{g}$,

$$\psi = \sum_{a=1}^{n} N_a \alpha^a \qquad \psi/\psi^2 = \sum_{a=1}^{n} M_a \alpha^a/(\alpha^a)^2 \qquad (3.4.1)$$

with N_a, M_a integers. Hence $\bar{K}_{\bar{a}\bar{b}}$ is a singular matrix of rank r with null vectors $N_{\bar{a}}$, $M_{\bar{a}}$,

$$N_{\bar{a}}\, K_{\bar{a}\bar{b}} = \bar{K}_{\bar{a}\bar{b}}\, M_{\bar{b}} = 0 \tag{3.4.2}$$

where we have defined $N_o = M_o = 1$ and let barred indices run over $\bar{\Delta}$ rather than Δ.

From (3.1.1) and (3.4.2), we see that

$$\frac{d^2}{dt^2}\,(\Sigma\, N_{\bar{a}}\rho_{\bar{a}}) = 0$$

and in fact we choose, following [12]

$$\Sigma\, N_{\bar{a}}\rho_{\bar{a}} = 0 \tag{3.4.3}$$

as an extra constraint.

The variables ρ_a relate best to the points of the Dynkin diagram which displays the symmetry of the equations [12] but in order to evaluate Poisson brackets we need an independent set of r variables. A convenient choice is $\phi_a = \{\phi_1 \ldots . \phi_r\}$; $\phi_o = 0$, defined by

$$\rho_{\bar{a}} = \bar{K}_{\bar{a}\bar{b}}\, \phi_{\bar{b}} = \bar{K}_{\bar{a}b}\, \phi_b \tag{3.4.4}$$

Then (3.1.1) reads, using (3.4.2)

$$\ddot{\phi}_a = -e^{K_{ab}\phi_b} + M_a\, e^{\bar{K}_{oc}\phi_c} \tag{3.4.5}$$

If, following [12], we define

$$A = A_x = \frac{1}{2}\,(A_u + A_v) = \frac{1}{2} \sum_{a\varepsilon\Delta} \dot{\phi}_a\, H_a + \frac{1}{2} \sum_{\bar{a}\varepsilon\bar{\Delta}} e^{\frac{1}{2}\bar{K}_{\bar{a}b}\phi_b}(E_{\bar{a}} + E_{-\bar{a}})$$

$$-iB = A_t = \frac{1}{2}(A_u - A_v) = \frac{1}{2}\sum_a e^{\frac{1}{2}K_{\bar{a}b}\phi_b}(E_{\bar{a}} - E_{-\bar{a}}) \tag{3.4.6}$$

and using the commutators (3.3.4), (3.3.19) and (3.4.1) we find

$$\frac{dA}{dt} - i[B,A] = \frac{1}{2}\sum_{a\varepsilon\Delta} H_a(\ddot{\phi}_a + e^{K_{ab}\phi_b} - M_a\, e^{\bar{K}_{oc}\phi_c}) \qquad (3.4.7)$$

thereby establishing that A, B in (3.4.6) constitute a Lax pair for the Toda lattice equations. The energy integral TrA^2 has the same kinetic term as before. Hence the canonical momentum conjugate to ϕ_a is again given by (3.3.6) and A in terms of canonical variables is

$$A(\phi,\pi) = \sum_{a\varepsilon\Delta} \lambda_a \cdot H\pi_a + \frac{1}{2}\sum_{a\varepsilon\Delta}{}_{-}\, e^{\frac{1}{2}K_{\bar{a}b}\phi_b}(E_{\bar{a}} + E_{-\bar{a}}) \qquad (3.4.8)$$

Now the commutators used in (3.4.7),

$$[E_o, E_{-o}] = H_o = -\Sigma\, M_a H_a ; [H_a, E_{\pm o}] = \pm E_{\pm o}\, \bar{K}_{oa}$$

can be satisfied by taking

$$E_o = \lambda E_{-\psi} \quad , \qquad E_{-o} = \frac{1}{\lambda} E_{\psi} \qquad (3.4.9)$$

where λ is a parameter which we shall take as real. Since the Lax pairs defined by (3.4.6) for different values of λ are gauge inequivalent[12] we now have an infinite family of Lax pairs for the Toda lattice equations labelled by λ. This phenomenon does not occur for the Toda molecule equations. Thus, a "spectral parameter" λ enters the Toda lattice theories.

The quantities $E_{\pm a}$ are all generators of the Lie algebra $\underset{\sim}{g}$. The quantities $\lambda^n T_i$, n = 0, ±1, ±2, ..., T_i as in (3.3.13) can be regarded as generators of an affine Kac-Moody algebra $\underset{\sim}{g} \otimes \mathcal{C}(\lambda,\lambda^{-1})$ if different powers of λ are linearly independent.[9] The step operators corresponding to the simple roots are E_o, E_1 .... E_r and the Cartan matrix defined by the simple roots is $\bar{K}_{\bar{a}\bar{b}}$ as stated in section 1. This important observation is due to Leznov and Saveliev[16], Drinfeld and Sokolov[13], and G. Wilson[18], and will be important in what follows. Thus the Lax operator A for the Toda molecule and lattice systems both involve step operators for the simple roots, in the former case for a simple Lie algebra $\underset{\sim}{g}$ and in the latter case for a Kac-Moody algebra $\underset{\sim}{g} \otimes \mathcal{C}(\lambda,\lambda^{-1})$.

When we evaluate the Poisson brackets between different matrix elements of A we shall want to use λ as an extra matrix element label, and so will take λ as the spectral parameter in the left entry and μ as the distinct spectral parameter in the right entry.

Then the $\mathbb{P}$ operator will be denoted $\bar{\mathbb{P}}(\lambda,\mu)$ and will satisfy

$$\{A(\lambda) \underset{,}{\otimes} A(\mu)\} = [\bar{\mathbb{P}}(\lambda,\mu),\ A(\lambda) \otimes 1 + 1 \otimes A(\mu)]$$

Using (4.8) we easily evaluate the left hand side

$$= \frac{1}{8}\Big(\sum_{a\varepsilon\Delta} (\alpha^a)^2\ e^{\frac{1}{2}K_{ab}\phi_b}[(E_a + E_{-a}) \otimes H_a - H_a \otimes (E_a + E_{-a})]$$

$$-(\psi)^2\ e^{\frac{1}{2}K_{ab}\phi_b}[(\lambda E_{-\psi} + \frac{1}{\lambda} E_\psi) \otimes H_\psi - H_\psi \otimes (\mu E_{-\psi} + \frac{1}{\mu} E_\psi)]\Big)$$

Thus we require $\bar{\mathbb{P}}(\lambda,\mu)$ to satisfy the previous equations(3.3.9,(3.3.10) and in addition

$$[\bar{\mathbb{P}}(\lambda,\mu),\ \lambda\ E_{-\psi} \otimes 1 + 1 \otimes \mu E_{-\psi}]$$
$$= -\frac{1}{4}\ \psi^2\ (\lambda E_{-\psi} \otimes H_\psi - H_\psi \otimes \mu E_{-\psi}) \qquad (3.4.10)$$

We also expect the transposition and hermiticity properties(3.2.5),(3.2.6)

$$\bar{\mathbb{P}}(\lambda,\mu)^T = -\ \bar{\mathbb{P}}(\mu,\lambda) \qquad (3.4.11)$$

$$\bar{\mathbb{P}}(\lambda,\mu)^+ = -\ \bar{\mathbb{P}}(\frac{1}{\lambda},\frac{1}{\mu}) \qquad (3.4.12)$$

Because of the uniqueness of $\mathbb{P}$(3.3.20) satisfying(3.3.9) and the comments at the end of the last section, $\bar{\mathbb{P}}$ must have the form

$$\bar{\mathbb{P}}(\lambda,\mu) = \mathbb{P} \qquad + F(\lambda,\mu)\ \mathbb{C} \qquad (3.4.13)$$

where $F(\lambda,\mu)$ is some function and $\mathbb{C}$ is defined in (3.3.12). It remains to construct $F(\lambda,\mu)$ so that (3.4.10) is satisfied and then check (3.4.11) and (3.4.12).

Now

$$[\mathbb{C}_+, E_{-\psi} \otimes 1] = \frac{\psi^2}{2} H_\psi \otimes E_{-\psi} + Y_1 \tag{3.4.14}$$

where the first term is the contribution of $\alpha = \psi$ in $\mathbb{C}_+$ (3.3.16) and Y_1 is a sum of terms proportional to $E_{-\alpha'} \times E_{-\alpha}$ with α and α' both positive roots (using the fact that if $\psi - \alpha$ is a root it must be positive). Further

$$[\mathbb{C}_+, 1 \otimes E_{-\psi}] = 0 \tag{3.4.15}$$

since $\psi + \alpha$ cannot be a root. Likewise

$$[\mathbb{C}_-, E_{-\psi} \otimes 1] = 0 \tag{3.4.16}$$

$$[\mathbb{C}_-, 1 \otimes E_{-\psi}] = (\psi^2/2) E_{-\psi} \otimes H_\psi + Y_2 \tag{3.4.17}$$

where Y_2, like Y_1 is a sum of terms proportional to $E_{-\alpha'} \times E_{-\alpha}$ with α and α' both positive roots. Hence by (3.4.14) to (3.4.17) and (3.3.21),

$$[\mathbb{C}, E_{-\psi} \otimes 1 + 1 \otimes E_{-\psi}] = Y_1 + Y_2$$

But by (3.3.14) this vanishes, so Y_1 and Y_2 cancel each other rather than individually vanishing. Using these facts we find for (3.4.13)

$$\begin{aligned}
&[\bar{\mathbb{P}}(\lambda,\mu), \lambda E_{-\psi} \otimes 1 + 1 \otimes \mu E_{-\psi}] \\
&\quad = \left(-\frac{(\lambda+\mu)}{2} + (\lambda-\mu)F(\lambda,\mu)\right) Y_1 \\
&\quad + \left(-\frac{\lambda}{2} + (\lambda-\mu)F(\lambda,\mu)\right) \frac{\psi^2}{2} H_\psi \otimes E_{-\psi} \\
&\quad - \left(-\frac{\mu}{2} + (\lambda-\mu)F(\lambda,\mu)\right) \frac{\psi^2}{2} E_{-\psi} \otimes H_\psi
\end{aligned}$$

and comparing with (3.4.10) we see that the coefficient of Y_1 should vanish, so

$$F(\lambda,\mu) = \frac{1}{2}\left(\frac{\lambda+\mu}{\lambda-\mu}\right) = - F(\mu,\lambda) = - F\left(\frac{1}{\lambda},\frac{1}{\mu}\right)$$

With this choice (3.4.10), (3.4.11) and (3.4.12) are all satisfied and

$$\overline{\mathbb{P}}(\lambda,\mu) = \mathbb{P} - \frac{1}{2}\frac{\mu+\lambda}{\mu-\lambda}\mathbb{C} = -\frac{1}{2}(\mathbb{C}_+ - \mathbb{C}_- + \frac{\mu+\lambda}{\mu-\lambda}\mathbb{C})$$

$$= -\frac{\mu}{\mu-\lambda}\mathbb{C}_+ - \frac{\lambda}{\mu-\lambda}\mathbb{C}_- - \frac{1}{2}\frac{\mu+\lambda}{\mu-\lambda}\mathbb{C}_o$$

This form is indeed reminiscent of that found in studies of other dynamica systems, but seems to us to be simpler. Now comes the main point of this chapter in terms of interpreting $\overline{\mathbb{P}}(\lambda,\mu)$; since λ and μ are formal parameters we may write

$$\frac{\mu+\lambda}{\mu-\lambda} = \frac{\mu}{\mu-\lambda} - \frac{\lambda}{\lambda-\mu} = \frac{1}{1-\lambda/\mu} - \frac{1}{1-\mu/\lambda}$$

$$= \sum_{n=1}^{\infty} (\lambda/\mu)^n - (\mu/\lambda)^n$$

Hence

$$-2\,\overline{\mathbb{P}} = \sum_{\alpha>o} \frac{\alpha^2}{2} E_\alpha \otimes E_{-\alpha} + \sum_{n=1}^{\infty} \sum_{\text{all }\alpha} \frac{\alpha^2}{2} \lambda^n E_\alpha \otimes \mu^{-n} E_{-\alpha}$$

$$+ \sum_{n=1}^{\infty} \sum_{i=1}^{r} \lambda^n H_i \otimes \mu^{-n} H_i$$

$$-\Big(\sum_{\alpha>o} \frac{\alpha^2}{2} E_{-\alpha} \otimes E_\alpha + \sum_{n=1}^{\infty} \sum_{\text{all }\alpha} \lambda^{-n} E_\alpha \otimes \mu^n E_{-\alpha}$$

$$+ \sum_{n=1}^{\infty} \sum_{i=1}^{r} \lambda^{-n} H_i \otimes \mu^n H_i\Big)$$

Now the step operators for the positive roots of the affine Kac-Moody algebra $\underset{\sim}{g} \otimes \mathcal{C}(\lambda,\lambda^{-1})$ are precisely [9]

E_α, $\alpha > o$; $\lambda^n E_\alpha$ $n \geq 1$, α any root (positive or negative);

$$\lambda^n H_i,\ n \geq 1 \quad i = 1 \ldots\ldots r \ .$$

if the step operators for the simple roots are $\lambda E_{-\psi}$, $E_1 \ldots\ldots E_r$. Thus if we rescale the E_α as before, $\sqrt{\frac{\alpha^2}{2}}\, E_\alpha \rightarrow E_\alpha$ we find that $\bar{\mathbb{P}}(\lambda,\mu)$ is in the form(3.1.3) precisely. Thus just as for the Toda molecule case A involved step operators for the simple roots and $\mathbb{P}$ step operators for all the roots of a simple Lie algebra $\underset{\sim}{g}$, so for the Toda lattice case A involves the step operators for simple roots and $\bar{\mathbb{P}}$ the step operators for all the roots of an affine Kac-Moody algebra $\underset{\sim}{g} \otimes \mathcal{C}(\lambda,\lambda^{-1})$. This is our main result.

3.5 Discussion

We want to discuss the sense in which the $\mathbb{P}$ operator we have constructed for two families of Toda systems is simple and universal. This simplicity and universality suggests that $\mathbb{P}$ may have an analogue in other dynamical systems based on bigger algebraic structures.

As explained the two families of Toda systems were based on the systems of simple roots for the simple Lie algebra $\underset{\sim}{g}$ and the affine Kac-Moody algebras $\underset{\sim}{g} \otimes \mathcal{C}(\lambda,\lambda^{-1})$. $\mathbb{P}$ was constructed in terms of the step operators for the complete root systems in a uniform way. Unlike the second element of the Lax pair, B, it applied equally to the many different dynamical systems based on the given algebra.

Our argument in constructing $\mathbb{P}$ was based on the fundamental geometric fact that the root system concerned could be split into positive and negative parts and possessed a basis of simple roots. Since this is true also of the remaining Euclidean Kac-Moody algebras based on the generalized Dynkin diagram (or "twisted loop algebras") we expect our own result to apply equally to the integrable systems associated with them [13] (which are subsystems of those discussed here based on extended diagrams [12]) but will not check it here. Our expectation is that there are yet more general systems of this nature associated with bigger algebraic structures. The two dimensional Toda equations may furnish an example.

Unlike the Lax operators A and B, $\mathbb{P}$ is independent of the dynamical coordinates and also of the choice of independent dynamical variables. Thus it is very likely that it will pass over into the quantum theory unchanged. Indeed the fundamental equation(3.2.3) holds for the quantum Toda systems when the Poisson bracket is replaced by a quantum commutator bracket and should presumably provide the basis for an integrable quantum theory. We hope to investigate this possibility in the future.

The statement that the Lax operators form a linear combination of Cartan subalgebra generators and step operators for Δ or $\overline{\Delta}$ is invariant with respect to gauge transformations generated by the Cartan subalgebra of $\underset{\sim}{g}$. $\mathbb{P}$ is also invariant to such transformations applied equally to its left and right entries Since the spectral parameter takes distinct values in the left and right entrie these gauge transformations must not depend on the spectral parameter.

This means that for other choices of A related to the Leznov Saveliev [16] choice adopted above by a spectral parameter-dependent gauge transformation, $\mathbb{P}$ will assume a different and in fact more complicated form. This seems to be the reason our simple form has not previously been discovered. The point of the Leznov Saveliev definition is that it associates the spectral parameter with the natural grading of the Kac-Moody algebra.[9]

We suspect that the crucial feature of $\mathbb{P}$ is that it plays the role of structure constants when the fundamental Poisson bracket equation(3.2.3) is regarded as imposing a second Lie algebraic structure on A. Thus A is a sum of products of two generators taken from the Lie algebra of Poisson brackets (3.2.3) and the Lie algebra(3.3.13).

We have claimed that P is more fundamental than B. To illustrate let us construct B from A and $\mathbb{P}$ when $H = \mathrm{Tr}(A^2)$. For the Toda molecule case, it is easy to verify using (3.3.5) that (3.2.9) and (3.3.1) yield (3.3.2). In the Toda lattice case equation (3.2.9) must be modified on account of the spectral parameter dependence. Since $\mathrm{Tr}A(\lambda)^2$ is λ independent (3.2.9) becomes

$$- \frac{dA(\mu)}{dt} = \{H, A(\lambda)\} = \{\mathrm{Tr}A(\lambda)^2 \underset{,}{\otimes} A(\mu)\}$$

$$= 2[\mathrm{Tr}_L(\overline{\mathbb{P}}(\lambda,\mu)A(\lambda)), A(\mu)]$$

$\overline{\mathbb{P}}(\lambda,\mu)$ is singular when $\mu = \lambda$ yet the left hand side is regular and in fact completely independent of λ. This apparent paradox is resolved by calculating

$$-2\mathrm{Tr}_L(\overline{\mathbb{P}}(\lambda,\mu)A(\lambda)) = \frac{\mu+\lambda}{\mu-\lambda} A(\mu) + iB(\mu)$$

which indeed yields the correct Lax equations as the λ dependent term cancels out of the commutator.

It should be noted that we never had to mention a specific representation of the simple Lie algebra concerned. In fact our expression (3.1.3) for $\mathbb{P}$ is valid in any representation. Throughout we made implicit use of the fact that the trace of a quadratic expression of generators is independent of the representation except for a common scale factor which can be absorbed in the normalization of the trace. This is not true for traces of higher powers of generators. This means that when, as in section 2, a higher order Hamiltonian is expressed as a linear combination of $\mathrm{Tr}(A^N)$, then the coefficients will depend on the representation considered. That is why it is important to have the freedom of taking their coefficients arbitrary as we mentioned.

It is instructive to consider another favorite dynamical system associated with Lie groups; that of a free particle moving on a simple Lie group. From the work of Pohlmeyer and others[19] the Lax pair is simply

$$A = -ig^{-1}\frac{dg}{dt} = A_i T_i \quad , \; B = 0$$

Clearly there are dim $\underset{\sim}{g}$ coordinates and the same number of conserved quantities A_i. The canonical formalism implies that

$$\{A_i, A_j\}_{PB} = i\, f_{ijk} A_k \tag{3.5.1}$$

so the number of conserved quantities in involution only equals rank $\underset{\sim}{g}$ and the theory is not completely integrable in the sense of Arnold[14] even though it is formally integrable. In the notation of the present paper (3.5.1) can be written

$$\{A \otimes A\}_{PB} = \frac{1}{2}[\mathbb{C},\ A\otimes 1 - 1 \otimes A]$$

from which it follows again that the $\mathrm{Tr}(A^N)$ are in involution but only rank $\underset{\sim}{g}$ are independent from the theory of Casimir invariants.

More generally one might expect

$$\{A \underset{,}{\otimes} \dot{A}\}_{PB} = [\,\mathbb{P},\ A\otimes 1] - [\,\mathbb{P}^T,\ 1\otimes A]$$

Our two examples seem to constitute two extremes in the sense that for the Toda systems $\mathbb{P}$ is antisymmetric while for the particle on a group it is symmetric.

REFERENCES

INTRODUCTION

(1) C.N. Yang and R.L. Mills, Phys. Rev. 96 (1954) 1605

(2) K. Symanzik, unpublished

(3) M.F. Atiyah, N.J. Hitchin, V.G. Drinfeld and Yu.I. Manin. Physics Lett 65A (1978) 185

(4) P.W. Higgs, Phys. Rev. Lett. 12, (1964) 132
Phys. Rev. Lett. 13, (1964) 508
Phys. Rev. 145 (1966) 1156
T.W.B. Kibble, Phys. Rev. 155 (1967) 1557
F. Englert and R. Brout, Phys. Rev. Lett. 13 (1964) 321

(5) T.W.B. Kibble, Acta Physica Polonica B13 (1982) 723

(6) M. Toda, "Theory of Non-Linear Lattices", Springer series in Solid State Physics, 20 (1981)

Chapter I

(1) H.B. Nielsen and P. Olesen, Nucl. Phys. B61 (1973) 45

(2) T.W.B. Kibble, J. Phys. A9 (1976) 1387

(3) T.W.B. Kibble, G. Lazarides and Q. Shafi, Phys. Letters, 113B (1982) 237.

(4) P. Goddard and D. Olive, Nucl . Phys. B191 (1981) 511

(5) G. Racah in "Group Theoretical Concepts and Methods in Elementary Particle Physics" p1-36, Gordon and Breach, New York (1964)

(6) J.E. Humphreys, Introduction to Lie algebras and representation theory, Springer, Berlin (1972)

(7) E. Corrigan and D. Olive, Nucl. Phys. B110 (1976) 237

(8) P. Goddard, J. Nuyts and D. Olive, Nucl. Phys. B125 (1977) 1

(9) D. Olive "Relations between Grand Unified and Monopole Theories" ICTP/81/82-6

D. Olive and P. West: "The N=4 Supersymmetric E_8 gauge theory and coset space dimensional reduction" ICTP/81/82-25

(10) For a review see P.J.A. Peebles , The Large Scale structure of Space-Time (Princeton U.P.) 1980.

(11) Ya. B. Zeldovich, Mon. Not. Rot. Astron. Soc. 192 (1980) 663.

(12) A. Vilenkin, Phys. Rev. Lett. 46 (1981) 17

(13) A. Vilenkin, Phys. Rev. D24 (1981) 2082

(14) P. Goddard, J. Goldstone, C. Rebbi and C.B. Thorn, Nucl. Phys B56 (1973) 109.

(15) E.H. Brandt, J.R. Clem and D.G. Walmsley, J.L.T.P. (1979) 43

(16) A.Vilenkin, Phys. Rev. D23 (1981) 17

(17) R. Giovanelli and M.P. Haynes, Ast. J. 87 (1982) 1355.

(18) Ya. B. Zel'dovich, Mon. Not. Roy. Astron. Soc. 160 (1972)

(19) W.H. Press andE.T. Vishniac, Ap. J. 236 (1980) 323.

(20) A.E. Everett, Phys. Rev. D24 (1981) 858

(21) D.V. Nanopoulos and S. Weinberg, Phys. Rev. D20 (1979) 2484

(22) G. Steigman, Ann. Rev. Ast. and Astro. 14 (1976) 339

Chapter II

1. M. Toda, Physics Reports 18(1975) 1, "Theory of Non-linear Lattices" Springer series in Solid State Physics, 20(1981), Berlin.

2. A.N. Leznov and M.V. Saveliev, Lett in Math. Phys. 3(1979) 489, Comm. Math. Phys. 74(1980) 111.

3. A.N. Leznov and M.V. Saveliev, Phys. Lett 79B(1978) 294, Lett. in Math. Phys. 3(1979) 207.

4. N. Ganoulis , P. Goddard and D. Olive, Nucl. Phys. B205 [FS5] (1982) 601.

5. E. Witten, Phys. Rev. Lett. 38(1977) 121.

6. A.N. Leznov and M.V. Saveliev, "Cylindrically symmetric instantons for arbitrary compact gauge group" Serpukhov Preprint 78-177, Phys. Lett. 83B(1979) 314.

7. S. Berman, R. Moody and M. Wonenburger, Indiana Univ. Math. Journal, 21(1972) 1091.

8. R.K. Dodd and R.K. Bullough, Proc. Roy. Soc. Lond. A352(1977) 481.

9. P. Goddard and D. Olive, Nucl. Phys. B191(1981) 528.

10. G. Racah: "Lectures on Lie groups" CERN yellow report 61-8, and in "Group Theoretical concepts and methods in Elementary Particle Physics" p1-36, Gordon and Breach, New York (1964).

11. J.E. Humphreys: Introduction to Lie algebras and representation theory (Springer, Berlin 1972).

12. B.G. Wybourne: "Classical groups for physicists", Wiley (1974).

13. O.I. Bogoyavlensky, Comm. Math. Phys. 51(1976) 201.

14. B. Kostant, Adv. Math. 34(1979) 195.

15. M.A. Olshanetsky and A.M. Perelomov, Invent. Math. 54(1979) 261.
A.V. Mikhailov, M.A. Olshanetsky and A.M. Perelomov, Comm. Math. Phys. 79(1981) 473.

16. G. Wilson, Ergod. Th. and Dynam. Sys. 1(1981) 361.

17. S.A. Bulgadaev, Phys. Lett. 96B(1980) 151.

18. V.E. Zakharov and A.V. Mikhailov, Zh. Eksp. Teor. Fiz. 74(1978) 1953; Transl. Sov. Phys. JETP 47(6) (1978) 1017.

19. A.P. Fordy and J. Gibbons, Comm. Math. Phys. 77(1980) 21.

20. A.V. Mikhailov, CERN preprint TH. 3194(1981).

21. J.R. Faulkner, J. Algebra 62(1980) 384.

22. C.N. Yang, Phys. Rev. Lett. 38 (1977) 1377

23. L. Dolan, Phys. Rev. Lett. 47, (1981) 1371

Chapter III

(1) P.P. Kulish and E.K. Sklyanin: "Quantum Spectral Transform Method. Recent Developments", p. 61 in "Integrable Quantum Field Theories", edited J. Hietarinta and C. Montonen; Lecture Notes in Physics, vol. 151 (Springer, 1982).

(2) L. Faddeev: "Integrable Models in 1+1 Dimensional Quantum Field Theory", to appear in proceedings of the "Ecole d'Eté de Physique Théorique", Les Houches 1982 (North Holland to be published).

(3) See articles in "Monopoles in Quantum Field Theory", edited N.S. Craigie, P. Goddard and W. Nahm (World Scientific, Singapore 1982).

(4) D. Olive: "Current Topics in Elementary Particle Physics", Proc. Bad Honnef Summer Inst. ed K.H. Mütter and K. Schilling (Plenum, 1981), p. 199.

D. Olive: "Gauge Theories: Fundamental Interactions and Rigorous Results"; ed. P. Dita, V. Georgescu and R. Purice (Birkhäuser, Boston 1982), p. 33.

N. Ganoulis, P. Goddard and D. Olive: Nucl. Phys. B205 [FS5] (1982) 601.

(5) P.D. Lax: Comm. Proc. Appl. Math. 21, 467 (1968).

(6) V.E. Zakharov and A.B. Shabat. Funct. Anal. Appl. 13 (3) (1979); 13; English Transation: 166.

(7) E.K. Sklyanin: Doklady AN. SSSR, 244, 1337 (1979): English Translation: Sov. Phys. Dokl. 24 (1979) 107.

(8) G. Racah: "Lecture on Lie Groups" CERN yellow report 61-8 and in "Group Theoretic Concepts and Methods in Elementary Particle Physics", p. 1, (Gordon and Breach, 1964).

J.E. Humphreys: "Introduction to Lie algebra and representation theory" (Springer, 1972).

H. Georgi: "Lie Algebras in Particle Physics" (Benjamin 1982).

(9) V. Kac: Izv. Akad. Nauk. SSSR Ser. Mat. 32 (1968), 1323; English translation, Math. USSR Izvestija 2 (1968), 1271.

R.V. Moody: J. Algebra 10 (1968), 211; Canad. J. Math. 21 (1909), 1432.

for a concise and clear review see:

I.G. MacDonald: "Affine Lie Algebras and Modular Forms", Seminaire BOURBAKI 577-01 in Springer Lecture Notes in Math. 901 (1981), 258.

(10) M. Toda: "Theory of Nonlinear Lattices" Springer Series in Solid State Sciences 20 (1981).

(11) J. Moser: Springer Lecture Notes in Physics 38, p. 467; Adv. Math. 16, 197 (1975).

(12) D. Olive and N. Turok: The Symmetries of Dynkin Diagrams and the Reduction of Toda Field Equations", Nucl. Phys. B215 [FS7] (1983), 470.

(13) V.G. Drinfeld and V. Sokolov: Dokl. Akad. Nauk SSSR 258 (1981), 11; English translation: Sov. Math. Dokl. 23, 457 (1981).

(14) V.I. Arnold: "Mathematical Methods of Classical Mechanics" (Springer, 197

(15) O.I. Bogoyavlensky: Comm. Math. Phys. 51 (1976) 201.

S.A. Bulgadaev; Phys. Lett. 96B (1980) 151.

M.A. Olshanetsky and A.M. Perelomov, Invent. Math. 54 (1979), 261.

A.V. Mikhailov, M. A. Olshanetsky and A.M. Perelomov, Comm. Math. Phys. 79 (1981), 473.

(16) A.M. Leznov and M.V. Saveliev: Lett. Math. Phys. 3 (1979) 489; Comm. Math. Phys. 74 (1980), 111.

(17) S.V. Manakov; Sov. Phys. JETP 40 (1974) 269.

S. Wojciechowski: Lett. Nuovo Cim. 18 (1977), 103.

M. Adler; Invent. Math. 50 (1979), 219.

(18) G. Wilson; Ergod. Th. and Dynam. Sys. 1 (1981) 361.

(19) K. Pohlmeyer: Commun. Math. Phys. 46 (1976) 207.

E. Brezin, C. Itzykson, J. Zinn-Justin, J-B. Zuber, Phys. Rev. Lett. 82B (1979), 442.

E. Eichenherr and M. Forger, Nucl. Phys. B155 (1979), 381.